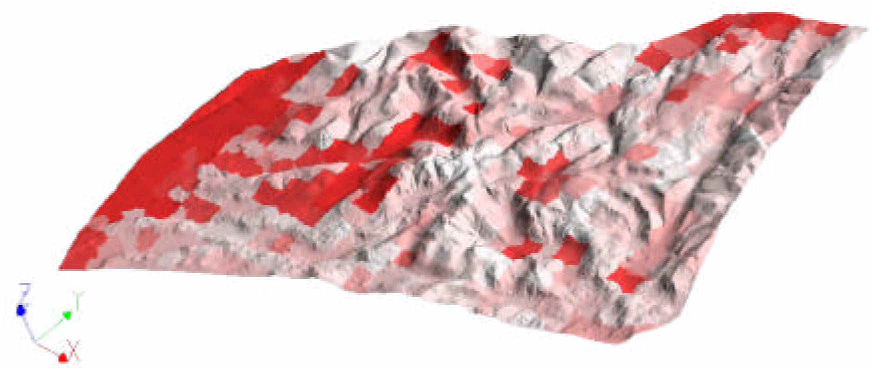
British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Confidence in geological interpretation

A methodology for evaluating uncertainty in common two and three-dimensional representations of subsurface geology.

Digital Geoscience Spatial Model Programme

Internal Report IR/04/164



BRITISH GEOLOGICAL SURVEY

INTERNAL REPORT IR/04/164

Confidence in geological interpretation

A methodology for evaluating uncertainty in common two and three-dimensional representations of subsurface geology.

S. Clarke

Key words

Geological modelling; uncertainty; confidence.

Front cover

Confidence in the Bedrock geological interpretation for the Lake District, represented on present day DTM

Bibliographical reference

CLARKE, S.M. 2004. Confidence in geological interpretation. A methodology for evaluating uncertainty in common two and three-dimensional representations of subsurface geology. *British Geological Survey Internal Report*, IR/04/164. 33pp.

Keyworth, Nottingham British Geological Survey 2004

BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham and Edinburgh; see contact details below or shop online at www.thebgs.co.uk

The London Information Office maintains a reference collection of BGS publications including maps for consultation.

The Survey publishes an annual catalogue of its maps and other publications; this catalogue is available from any of the BGS Sales Desks.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as its basic research projects. It also undertakes programmes of British technical aid in geology in developing countries as arranged by the Department for International Development and other agencies.

The British Geological Survey is a component body of the Natural Environment Research Council.

Keyworth, Nottingham NG12 5GG

☎ 0115-936 3241 Fax 0115-936 3488
e-mail: sales@bgs.ac.uk
www.bgs.ac.uk
Shop online at: www.thebgs.co.uk

Murchison House, West Mains Road, Edinburgh EH9 3LA

☎ 0131-667 1000 Fax 0131-668 2683
e-mail: scotsales@bgs.ac.uk

London Information Office at the Natural History Museum (Earth Galleries), Exhibition Road, South Kensington, London SW7 2DE

☎ 020-7589 4090 Fax 020-7584 8270
☎ 020-7942 5344/45 email: bgslondon@bgs.ac.uk

Forde House, Park Five Business Centre, Harrier Way, Sowton, Exeter, Devon EX2 7HU

☎ 01392-445271 Fax 01392-445371

Geological Survey of Northern Ireland, 20 College Gardens, Belfast BT9 6BS

☎ 028-9066 6595 Fax 028-9066 2835

Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB

☎ 01491-838800 Fax 01491-692345

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1EU

☎ 01793-411500 Fax 01793-411501
www.nerc.ac.uk

Foreword

This report is the published product of a study by the British Geological Survey (BGS) into the possibilities of modelling confidence in geological interpretations. It is the result of work carried out under the umbrella of the Lake District DGSM project to develop a method of assessing geological confidence in standard interpretations in order that they can be used in the construction of three-dimensional geological models.

Acknowledgements

This work is the result of lengthy discussions with a number of BGS staff in addition to those specifically working on the Lake District DGSM project. In particular I would like to thank:

Dr D. Millward (BGS) (Project Leader) for valuable discussion on the confidence assessment method and who single-handedly performed the tedious evaluation of confidence in all the geological maps and cross-sections of the Lake District.

Dr M. Cave (BGS) who contributed extensively to discussions on uncertainty.

Dr M. Lee (BGS) who made invaluable contributions to the confidence analysis of the geophysical interpretation of the Lake District batholith.

Dr G. Leslie (BGS) for valuable comments on the confidence assessment method during its early stages of development

C. Vye (BGS) for comments, contributions and application of confidence assessment to the ongoing modelling work in the Lake District.

Dr G. Tanner (Glasgow University) who is currently applying the method to his geological maps of the Glen Lochy area.

Contents

FIGURES

Summary

The Lake District DGSM project (Clarke, 2004) represented a first attempt by the British Geological Survey to construct a three-dimensional model from two-dimensional, and often poorly constrained, subsurface interpretations. Over the course of this project it became clear that success depended on objective evaluations of confidence in the input interpretations in order to resolve the inevitable conflicts between them.

Confidence in modelling has been considered by other BGS workers (Cave & Wood, 2002) but they have concentrated on uncertainty in the aspects of numerical modelling. Assessment of the uncertainty in the geological aspects of interpretations has not been considered.

This report describes a method of evaluating confidence in the geological aspects of two-dimensional interpretations of the subsurface that was developed out of necessity to resolve problems in the Lake District DGSM. This report gives a generic overview of the method but uses examples taken from the Lake District DGSM to demonstrate it.

The method is primarily aimed at evaluating confidence in standard two-dimensional representations of surface and subsurface geology (maps, cross-sections, contoured horizon plans etc.) so that their value to three-dimensional modelling can be assessed. However, the technique can be taken further to evaluate confidence in three-dimensional models and a method for doing so is described.

The assessment method is still in the embryonic stages of development. To date it has been applied to the Lake District DGSM with success and is currently being used to resolve similar problems in the Glen Lochy DGSM. However, with the exception of these projects it remains largely untested.

1 Introduction

A basic principle behind the science of geology is the development of three-dimensional models of the subsurface from the application of geological theories to a limited set of observations. A *model* is required because it is not possible to observe (and therefore know) everything about the geology of the subsurface.

To convey three-dimensional geological models to colleagues, geologists commonly employ a wide-ranging set of standard, two-dimensional methods including; geological maps of the bedrock and superficial geology, detailed cross-sections of the subsurface and contoured horizon plans derived from measured or remotely sensed data. Commonly, two-dimensional representations of geological models are augmented by full, three-dimensional computer visualisations. These visualisations may in themselves be first-order derivatives of geological observations or they may rely, in whole or part, on the combination of two-dimensional interpretations into a three-dimensional visualisation.

Two-dimensional geological interpretations are produced using a variety of geological techniques, are based on a specific set of geological theories and use, as input, different types of geological observation. Assumptions are inherent in each interpretative method and these limit the interpretation. Geologists implicitly understand the limitations of different two-dimensional interpretations and, to a certain extent, make expert judgments of the *confidence* that can be placed in them. Clients and members of the public may not always appreciate the same limitations. In three-dimensional computer visualisations of models, limitations and implicit confidence in interpretation are much harder to appreciate, even for the trained geologist. Moreover, interpretative confidence in three-dimensional models may vary greatly and stem from many more sources than is the case for two-dimensional interpretations.

If three-dimensional models are constructed from two-dimensional geological interpretations (three-dimensional interpretative modelling – see: Clarke, 2004), problems of confidence are compounded. Conflicts in the interpretation of the position of a given subsurface horizon or the interpretation of a geological property (e.g. lithostratigraphy) between different two-dimensional interpretations need to be resolved in order to produce a coherent, three-dimensional model. In the absence of directly observed, subsurface data (such as well-log data) this can only be achieved if confidence is assessed for the two-dimensional interpretations and used as a guide to the *likelihood of accuracy*.

Problems such as these quickly became apparent in the Lake District DGSM (Clarke, 2004). This DGSM represented a first attempt by the British Geological Survey to construct a representative, three-dimensional model of the subsurface in an area of the UK for which there are few, directly observable, subsurface data. The bulk of this model is based on the combination of various two-dimensional interpretations and assessments of confidence in these were key to the successful development of a three-dimensional model.

This report details a methodology, originally developed for the Lake District DGSM, that allows confidence in two-dimensional geological interpretations to be assessed. Examples of confidence assessment in standard geological interpretations (geological maps, cross-sections and contoured horizon plots) are given, taken from the Lake District DGSM. The principles of confidence are taken further to develop a method that captures and propagates confidence spatially to the three-dimensional model as a combination of supporting and refuting evidence.

1.1 AIMS OF THE PROJECT

The method of evaluating confidence reviewed here is a generic approach that can be applied to many different two-dimensional interpretations. It is intended to evaluate *geological* confidence not *measurement* or *computational* confidence (see Section 2.1). And to this end it is complimentary to many additional studies into modelling confidence by the BGS (Cave and Wood, 2002).

The specific aims of this report are:

- To describe a method of evaluating *geological confidence* in standard forms of two-dimensional representation of subsurface geology (maps, cross-sections and contoured horizon plans).
- To develop the method that allows relative confidences from different two-dimensional interpretations to be combined to support (or refute) a particular interpretation.
- To demonstrate the application of the technique to various examples of two-dimensional interpretations (from the Lake District DGSM) and to describe application of the method to three-dimensional interpretative modelling.

1.2 BACKGROUND READING

This report builds on, and is complementary to, the other work of other researchers both within and outside of the BGS. Additional texts that should be consulted in parallel with this report are:

- Cave, M. R. and Wood, B. 2002. Approaches to the measurement of uncertainty in geoscience data modelling. BGS Internal Report, IR/02/068.
- Funtowicz, S.O. and Ravetz, J.R. 1990. Uncertainty and quality in science for policy. Dordrecht: Kluwer
- Clarke, S.M. 2004 The Lake District DGSM An overview of the model and best practice guidelines. BGS Internal Report IR/04/00
- Shrader-Frechette, K S. 1993. Burying uncertainty. Risk and the case against geological disposal of nuclear waste. University of California Press.

2 Geological uncertainty

Two-dimensional interpretations (maps, cross-sections etc.) are the result of detailed and often lengthy processes in which the geologist collects raw field-data and observations, analyses (plots) and interprets those data based on geological theories, interpolates between or extrapolates beyond the data to 'complete the picture' and extends the information gleaned from so doing to other interpretations or other parts of the same interpretation.

At each stage of the process errors and/or uncertainty are introduced. In no particular order, these may include:

- errors in raw data measurement or spatial positioning
- errors and uncertainty resulting from the amount (or lack) of raw data
- errors related to scale or projection of data
- uncertainty in the geological theories employed in the interpretation
- uncertainty in the accepted interpretation of the regional geology or geological history of the area under study
- uncertainty with interpolative or extrapolative techniques
- uncertainty in expert knowledge (the limit of knowledge, training or experience of the scientist making the interpretation)
- uncertainty in the quality of the data, the method by which they were collected or the manner in which they have been processed
- uncertainty resulting from data conflict

It is important to appreciate that *error* and *uncertainty* are not the same thing. Error is defined as *the difference between an individual result and the true value of the measurand* (ISO, 2000). As such, errors can be quantified numerically by an error margin dependent on the measuring method and measurand. The error margin provides a range of values such that the true value of the measurand definitely lies within the range, or has a probability of lying within the range. Errors may form part of the uncertainty in an interpretation. By contrast, uncertainty can be defined as '*that which we do not know*' (Funtowicz and Ravetz, 1990) and arises from a combination of error with interpretation, interpolation and extrapolation of measured data under the influence of informed judgement. It cannot be directly quantified in the same manner. The degree of uncertainty in interpretation may vary spatially.

2.1 IDENTIFYING UNCERTAINTY

The first stage to evaluating uncertainty in a geological interpretation is to identify the sources and their relationships (Cave and Wood, 2002). One of the most successful methods of determining all sources of uncertainty within a given interpretation is to use a *'cause and effect'* (or fishbone) diagram (Kindlarski, 1984). Cave and Wood (2002) gave a detailed description of the construction of a fishbone diagram that is not repeated here, other than to note that the diagram identifies a hierarchical set of processes that contribute to overall uncertainty in the geological interpretation. By approaching uncertainty in this way the fundamental sources, their relationships and compound effects soon become clear.

Ultimately, sources of uncertainty, as identified by the fishbone diagram, can be characterised into three, broad groups as follows:

1. **Quantifiable errors.** Many sources of uncertainty in any two-dimensional interpretation are related to measurement or data errors (Ferson et al., 1999). These uncertainties may include positional accuracy errors, scale problems, and random and systematic errors resulting from the characteristics of the measured entity and the measuring method.

2. **Computational (or 'fitting') uncertainty.** Sources of uncertainty commonly stem from the specific mathematical interpolation or extrapolation algorithm used to interpret the data. The fitting of curves through sampled data in two dimensions and the interpolation of surfaces to data in three dimensions have specific associated uncertainties resulting from the mathematical model used to produce the interpolation.

3. **Geological interpretation uncertainty.** The remaining sources of uncertainty can be grouped together under this title. The category includes all sources of uncertainty that stem from the *geological* aspects of the interpretation, including data interpretation, the geological theories applied to the interpolations, the geological knowledge of the area, and the expertise of the interpreter.

Quantifiable errors (category 1) can be evaluated using error margins, accuracy statements and statistical analysis (e.g. Isaaks and Srivastava, 1989; Day and Underwood, 1991). The effects of computational uncertainty (category 2) can be evaluated using statistical re-sampling techniques (Efron, 1979; McKay et al., 1979; Iman and Helton, 1988; 1991; Efron & Tibshirani, 1993; Young, 1994; Davison and Hinkley, 1997; Meinrath et al., 2000; Wehrens et al., 2000).

Geological interpretation uncertainty (category 3) is more difficult to evaluate mathematically, because it often includes a degree of subjective (expert) judgement within the modelling process. It is the assessment of *geological interpretation uncertainty* that this report addresses.

3 Modelling geological interpretation uncertainty

Sources of uncertainty stemming from aspects of geological interpretation can be divided into two sets:

1) Uncertainty that is related to the geological properties of the raw data in some way, such as the amount of data, their distribution, density etc.
2) Uncertainty that is related to the interpretation of those data, such as the theories employed, the method used, expert judgement, the possibility of other equally valid solutions etc.

It is this distinction between uncertainty connected with the ***quantity*** of data and that associated with the ***quality*** of interpretation that forms the basis of the assessment method introduced here.

The method assesses *geological confidence*, which can be defined as *'the degree to which we consider that an interpretation is likely to be totally correct'*. To this end, confidence may be considered as the negative (or inverse) of uncertainty and it follows that total confidence implies zero uncertainty and vice versa. In modelling terms, we describe (and evaluate) *geological confidence* rather than *geological uncertainty* simply because of the negative connotations associated with the word uncertainty. In practice, it is an understanding and description of the uncertainties that are fundamental to the evaluation of confidence.

For a two-dimensional interpretation, confidence is assessed based on the *quantity* of raw data with which the interpretation was formulated and the *quality* of the methods and theories used in that formulation.

3.1 QUANTITY, QUALITY & CONFIDENCE

The geological confidence assessment method described here elicits point scores under the categories of *data quantity* (Section 3.1.2) and *interpretation quality* (Section 3.1.4). These scores are based on a five-point, linguistic scale from *'very-high'* to *'very-low'*. This linguistic scale is equated to a normalised, numerical scale for the practical purposes of modelling (Figure 3.1). Each linguistic scale term is represented numerically by the median value of its range and borderline cases between linguistic categories can be represented by values between those median values. In most cases this resolution has proved adequate; greater resolution can be achieved by subdividing linguistic categories but this practice can significantly increase both evaluation times and the (somewhat unavoidable) subjective input of the evaluator.

Finally, quantity and quality scores are mathematically combined (Section 3.1.3) to produce a confidence score.

To accommodate spatial variation in confidence across the extent of an interpretation, quantity and quality (and therefore confidence) scores are elicited for different, defined areas of the interpretation. In the most subjective case, these areas can be of arbitrary shape, defined by perceived changes in confidence, to produce areas with gross quantity, quality and confidence scores. In the most objective case, the interpretation is divided up into regular blocks of defined area and quantity/quality scores are elicited for each block.

Figure 3.1. The linguistic terms and their numerical equivalents used for modelling. Terms shown in italic are borderline cases between the five main categories.

Linguistic Term	Numerical Value
Zero	0.0
Very-Low	0.1
Very-Low to Low	0.2
Low	0.3
Low to Moderate	0.4
Moderate	0.5
Moderate to High	0.6
High	0.7
High to Very-High	0.8
Very-High	0.9
Maximum	1.0

3.1.1 Data quantity

The quantity of data in a given area of a two-dimensional interpretation is of fundamental importance to the reliability (or confidence in) the final interpretation. In this assessment of confidence, the quantity of data is evaluated for each defined area of the interpretation using the five-point linguistic scale demonstrated in Figure 3.1.

For a specific interpretation, the amount of data per unit area that is represented by a data quantity score of *'very-high'* is taken to be that quantity of data that is considered *ideal* for making that interpretation. The ideal amount of data is defined as that level above which increasing amounts of data do not appreciably increase the reliability of the interpretation, usually as a result of sampling techniques used in the method of interpretation. Clearly, the amount of data defined as ideal is different in different cases and for different geological aspects of the interpretation. The remaining four quantity scores (*high, moderate, low and very-low*) form equal interval, linearly distributed categories over the range from the ideal data quantity down to zero.

For example, in the case of geological map, in which boundary position and bedrock geological interpretation are assessed for confidence (see Section 3.2.2), the ideal amount of data, per defined area, for making the bedrock interpretation would be 100% bedrock exposure[1] since this amount negates the need to interpret between data. For each defined area of the map, the percentage of bedrock exposure is evaluated and equated to the linear five-point scale and hence *'very-high'* represents 80% to 100% exposure and *'very-low'* represents 0% to 20% exposure.

In a geophysical interpretation of the subsurface, based on surface geophysical station readings, the ideal data quantity per defined area is that which is considered the best for generating the interpretation, and is dependent on the approach adopted. For example, if the method of interpretation relies on subsampling randomly distributed stations to a 1 km grid spacing, then a station spacing of 1 km in both the north and east directions may be considered ideal, as an increase in station density beyond this will have little impact on the interpretation. It follows that 25 or more station localities per quarter-sheet would represent a data quantity score of *'very-high'*. The fact that the station localities may not be evenly distributed over the quarter sheet and are interpolated to a grid for the purposes of interpretation can be taken into account at the quality stage (Section 3.1.2).

[1] Bedrock exposure is defined as the percentage of the area for which the geological bedrock is visible and not covered by superficial deposits, vegetation, soil or manmade features.

Examples of the assessment of data quantity using the concept of an ideal amount of data are given in Section 4.

3.1.2 Interpretation quality

Interpretation quality (and associated uncertainty) is dependent on a number of factors that can be grouped together under eight subjects (after Funtowicz and Ravetz, 1990; Cave and Wood, 2002; Bowden, 2003):

- data quality
- geological theories employed
- methods used in data collection and interpretation
- auditability (ability to trace interpretation back to original data)
- calibration (the extent to which the interpretation fits the available data)
- validation (the extent to which the model is a unique fit to the data)
- objectivity (bias)
- knowledge of the interpreter

With knowledge of the processes that contribute to interpretation uncertainty (Section 2.1), a quality score is elicited from a series of descriptive statements (Figure 3.2) under each of these subjects (Funtowicz and Ravetz, 1990; Bowden, 2003) using the same five-point linguistic scale used for the data quantity score.

For each subject, a quality score is determined by assessing which of the statements is the best fit to the truth in each case. The individual subject quality scores are then numerically averaged to determine an overall quality score (Funtowicz and Ravetz, 1990). This processes is repeated for each of the defined areas of the interpretation. This is not such a lengthy process as it may appear as many of the subject areas will have constant scores for a given interpretation irrespective of the defined area, and can be discounted from the assessment (although they must be included in the averaging process).

3.1.3 Confidence

An overall confidence score for each defined area is then calculated from the quantity and quality scores using the graphical relationship shown in Figure 3.3 (Bowden, 2003).

The quantity score is reduced to an amount defined by the intersection of the corresponding quality score and the appropriate ‘relationship curve’ shown on Figure 3.3, to give a resultant confidence score. These relationship curves are specifically designed to reduce the high quantity score resulting from a vast amount of data to a zero confidence score if, for example, the quality of that data is so low as to make the data useless, or some aspect of the methodology employed in the interpretation is not suitable for the data. A high confidence score relies on *both* a sufficient quantity of data and high quality of interpretative methods.

The shape of the relationship curves is not unequivocal and may be tailored to specific needs if necessary. The curves demonstrated in Figure 3.3 are non-linear to reflect the proportionally greater effect that interpretative quality is likely to have with decreasing amounts of data.

Score	DATA	THEORY	METHOD	AUDITABILITY	CALIBRATION	VALIDATION	OBJECTIVITY	EXPERT
1.0–0.8 VERY HIGH	Measured or field data collected specifically for making this interpretation.	Well-established and highly accepted geological theories were used in the interpretation.	Field survey or direct measure of data following appropriate *Best Practice* guide-lines. Iterative checking of interp. against new data.	Well documented, clear link between raw geological data and interpretation.	Interpretation correlates with all available raw geological data.	This interpretation is the only geologically sound interpretation of available data.	No discernible bias to preconceived geological understanding or interpretative style.	Interpreter is familiar with theories/methods and has the local geological knowledge required to make the interpretation.
0.8–0.6 HIGH	Historical, or field measured data collected as part of the general survey process, or for another purpose to this interpretation.	Accepted theories with a high peer consensus.	Majority field survey or direct measure with iterative checking. Some additions from desk compilation of interpretations.	Poorly documented but traceable link from interpretation to raw geological data.	Interpretation correlates with the majority of available geological data - anomalies can be explained and discounted.	Interpretation is the strongest solution to available data but a few, less likely interpretations also exist.	Influence of understanding of the regional geology but not at the expense of evidence from data.	Interpreter is familiar with theories, methods required but unfamiliar with aspects of the local geology.
0.6–0.4 MOD.	Computed or derived/calculated geological data including statistically generated values.	Accepted theory but poorly tested, lacking in examples or with a low peer consensus.	Desk compilation from derived/ historical data with field-checking, or from specifically collected data with no field checking.	Raw geological data traceable in part.	Interpretation correlates with most available data but some small, local, unexplained anomalies exist.	Interpretation is valid but a few, equally valid interpretations would fit the data just as well.	Moderate bias towards specific style of interpretation in data-poor areas but data evidence not ignored.	Interpreter is unfamiliar with aspects of the theory or method required to make the interpretation.
0.4–0.2 LOW	‘Standard’ values or approximated numbers.	Preliminary theory, poorly tested and un-validated.	Desk compilation from historical, inappropriate or insufficient data with no field check.	Weak, unclear or ambiguous link to the raw geological data. Original data on which the interpretation is based can not be fully identified.	Interpretation correlates with most data but geologically significant and/or large-scale, unexplained anomalies exist.	A significant number of plausible interpretations would fit the data.	Strong bias towards specific style of interpretation even the face of refuting geological data.	Interpreter is significantly lacking in understanding of theory or methods required to make the interpretation.
0.2–0 VERY LOW	Ball-park approximation.	Crude speculation.	No discernable rigour, best guess interpretation.	No link to the raw geological data, no recorded input to interpretation.	Correlation with the minority of data or no apparent correlation.	The interpretation is speculative.	Obvious bias toward specific interpretation and disregard of significant geological data.	Interpreter has little or no knowledge of the theories or method required to make the interpretation.

Figure 3.2. The quality of geological interpretations can be assessed by determining which of the statements under each category is most appropriate and equating it to the linguistic score at the lefthand side. An overall quality score can be determined by averaging individual scores under each of the 8 categories.

Figure 3.2

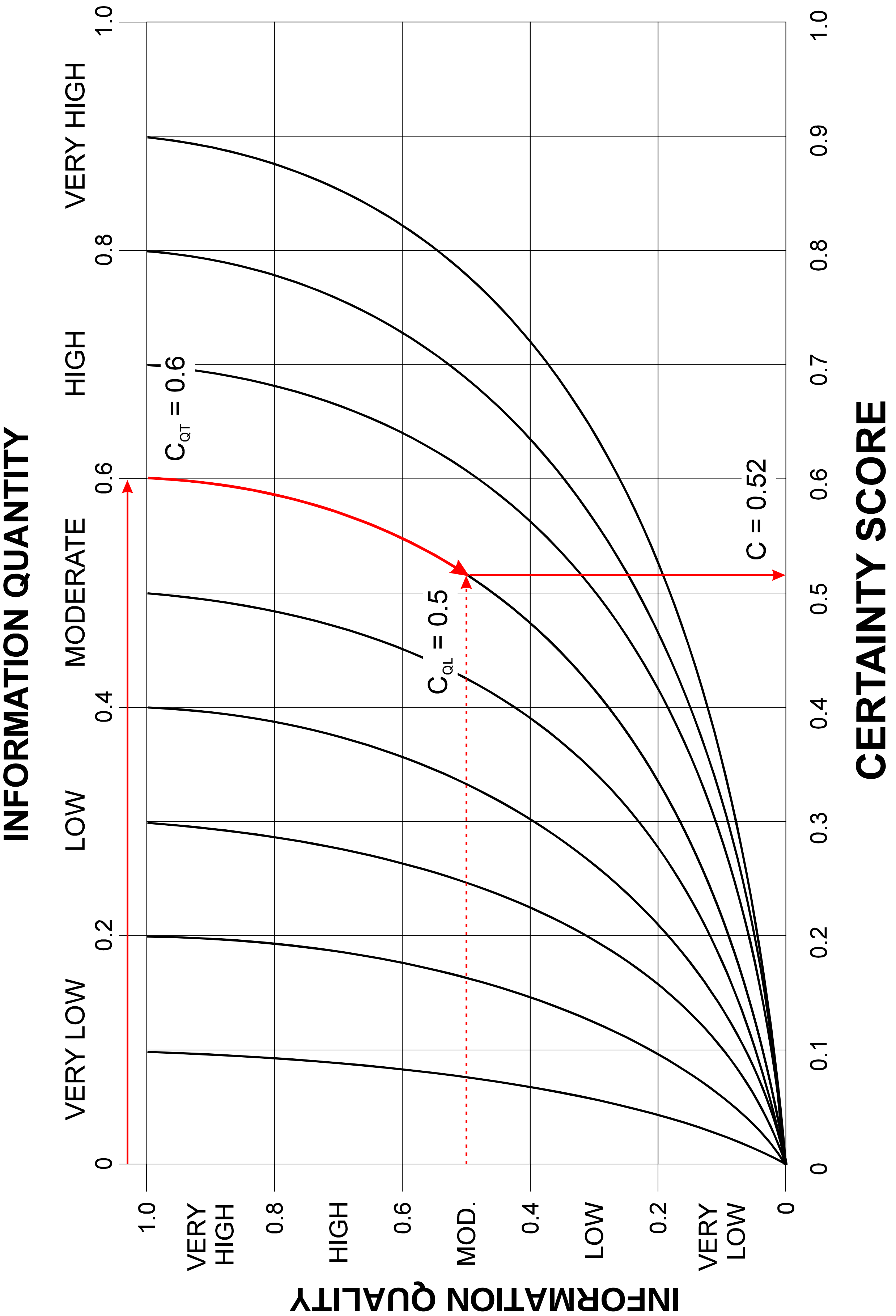


Figure 3.3. Confidence can be determined from a combination of the Data Quantity and Data Quality scores. The quantity score is reduced back along the appropriate curve by an amount equal to the quality score (after Bowden, 2003).

Figure 3.3

3.2 MODELLING CONFIDENCE –PRACTICAL ISSUES

The method of confidence assessment outlined in Section 3.1 above is based on theoretical and practical investigations of a number of workers in many fields (Funtowicz and Ravetz, 1990; Cave and Wood, 2002; Bowden, 2003; Clarke, 2004). The developed method has been applied to many different two-dimensional interpretations that form the input data for the Lake District DGSM (Clarke, 2004). Specific findings of some of these investigations are outlined in Section 4. However, as a result of this trial, it is clear that a number of generic issues must be considered for the practical implementation of the assessment method.

3.2.1 Spatial variation and sampling

It is important to appreciate that division of an interpretation into separate areas for the purposes of confidence evaluation can have sampling effects on the outcome. Whilst subdividing an area with constant quantity and quality (and therefore confidence) scores will have no effect on the resolution of that area, the same process will affect the confidence resolution at the boundary between areas of different confidence. The mathematical theories of sampling apply.

In theory, the most efficient method of subdividing and interpretation into sample areas would be to use a sampling area size that is inversely proportional to the local rate of change in confidence. In regions of constant confidence large sample area sizes are used and in regions of large variations in confidence small sample area sizes are used. In practice, varying sample area size as a function of confidence implies some preconceived knowledge of confidence variation in the interpretation.

The most practical method is to divide map-based interpretations into sample areas using the National Grid and cross-sections into sample areas of defined lateral distance and vertical thickness. For 1:50 000-scale interpretations, experiments in the Lake District (Clarke, 2004) have shown that the most appropriate sample area sizes are 1 km^2 areas for map-based interpretations and 0.5 km^2 (1 km wide by 0.5 km high) areas for cross-sections.

It is important to note that extra considerations in this respect are necessary if separate, two-dimensional interpretations are to be combined in three-dimensions (see sections 4.2.1 & 5.1).

3.2.2 Geological criteria

The discussion of geological interpretation confidence and its evaluation thus far has not considered the definition of the geological criteria for which confidence is assessed. In order to appreciate and assess confidence effectively it is important that the geological aspects of the interpretation that are of interest are determined and the assessment made accordingly. In most cases it is the *spatial position* of boundaries on the map or horizons in the subsurface that are of paramount interest (particularly for the construction of three-dimensional models from two-dimensional interpretations), and the confidence assessments can be made with respect to this. In other cases it may be the *lithological interpretation* that is of interest and assessments can be made accordingly. Confidence assessments can take into account more than one aspect of geology, but the broader the definition the more sources of uncertainty must be considered when evaluating both data quantity and data quality.

3.2.3 The assessor – history metadata

During the practical application of this assessment method to geological maps and sections used as input data for the Lake District DGSM, it became apparent that the only person with the knowledge required to make an objective assessment of confidence in a particular map or cross-section was the original author of that interpretation. For 1:50 000-scale interpretations (maps and sections) the complier of the original mapping may be best placed to make the assessment.

At 1:10 000-scale, this role falls to the survey geologist who made the original field interpretation. It is largely impossible for a geologist presented only with the interpretation and with no intimate knowledge of the area to make a valuable and reasonably objective confidence assessment of that interpretation.

3.2.4 Representing and interpreting the result

The mathematical analysis of confidence produces a numerical result between zero and one (Figure 3.3). Numerical scores are necessary for computer modelling purposes, but an assessment of confidence should not imply some form of quantification. Given that uncertainty (and therefore confidence) is dependent on 'that which we do not know' (Section 2) it is not possible, and misleading, to quantify it.

The numerical score can be used to produce a colour-shaded interpretation in which the colours represent levels of confidence. No scale should be associated with the colour shading beyond the definition of *very-high* and *very-low* confidence and the tonal variation between end-member colours should be continuous so as not to imply discrete values.

In practice, red-white colour scales have been found to be particularly useful (Cave and Wood, 2002; Clarke, 2004). By representing *very-high* confidence in white and *very-low* confidence in red, those areas of the interpretation about which we are uncertain can be scientifically highlighted without giving an overly negative impression of the reliability of the interpretation.

4 Assessing confidence – worked examples

This section presents some examples to show the application of the confidence assessment method outlined in Section 3 to a number of common, but different, types of standard two-dimensional geological interpretation. From these trials, it is possible to draw a number of important points for discussion. All examples are taken from work done for the Lake District DGSM (Clarke, 2004). In most cases, the confidence modelling work relied (in part or in whole) on assessments of data quantity and interpretation quality made by Dr D Millward (BGS), who was a major contributor to the latest geological survey of the Lake District (BGS 1996; 1999). Many of the conclusions drawn from the examples given are the result of lengthy discussions with him.

4.1 GEOLOGICAL BEDROCK MAP

The geological map (both bedrock and superficial) is perhaps the primary, two-dimensional representation of surface and subsurface geology. However, it is not a statement of fact, but an interpretation, at the time of survey, of a set of observations. The geologist uses these observations, combined with a geological background of the area and a selection of geological theories to complete the map between observations. Inevitably, the geological map is uncertain and the degree of uncertainty varies over its surface.

4.1.1 Geological interpretation confidence

In the example shown here, the 1:50 000-scale bedrock geological successions map of the north-western corner of the BVG outcrop within the Lake District (Figure 4.1) is assessed. This map is derived from 1:50 000-scale published geological bedrock maps (British Geological Survey, 1996; 1999) by combining related geological formations into 'successions' that comprise significant parts of the volcanic history (Millward, 2002). Therefore, the map's spatial resolution is that of the 50 000-scale published geological map.

It is assessed for confidence in the bedrock interpretation (position of boundaries/faults and lithostratigraphical interpretation). The assessment is performed for each 1 km^2 National Grid square.

4.1.1.1 DATA QUANTITY

For each square kilometre of the geological map, the quantity of data available for interpretation is defined as the percentage of bedrock exposure using the following scale:

Percentage exposure	Quantity score
80% to 100%	Very High
60% to 80%	High
40% to 60%	Moderate
20% to 40%	Low
0% to 20%	Very Low

Clearly, in a mountainous area such as the Lake District, there is a wide variation in exposure and therefore in quantity score, ranging from nearly full bedrock exposure and scores of '*very-high*' on mountain tops to almost zero bedrock exposure and scores of '*very-low*' in valleys, on

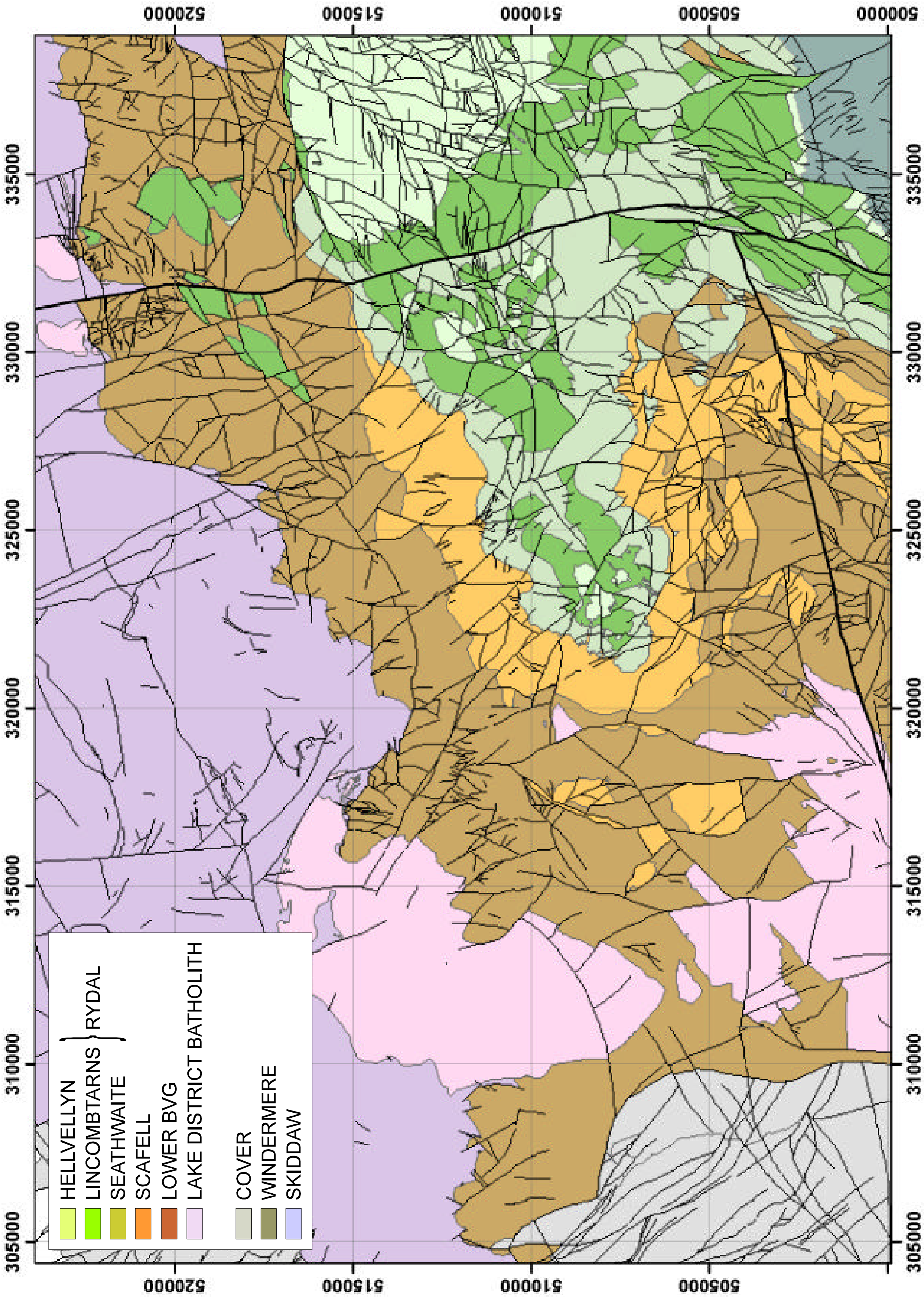


Figure 4.1. The bedrock geology of the Lake District.

coastal plains and in the urban centres. The quantity assessment for the Lake District geological bedrock map is shown in Figure 4.2a.

To a certain extent, the evaluation of data quantity for a geological bedrock map can be made with reference to corresponding superficial mapping and OS cultural datasets, and such datasets were used in this evaluation to partially automate the process. Expert knowledge input is also required in order to evaluate the impact of vegetation and small urban centres on exposure.

4.1.1.2 INTERPRETATION QUALITY

The quality of bedrock interpretation is assessed using the table shown in Figure 3.2. However, given the standard principles involved in making a geological map, some constant values (Section 3.1.2) apply for all assessed grid squares of the geological map:

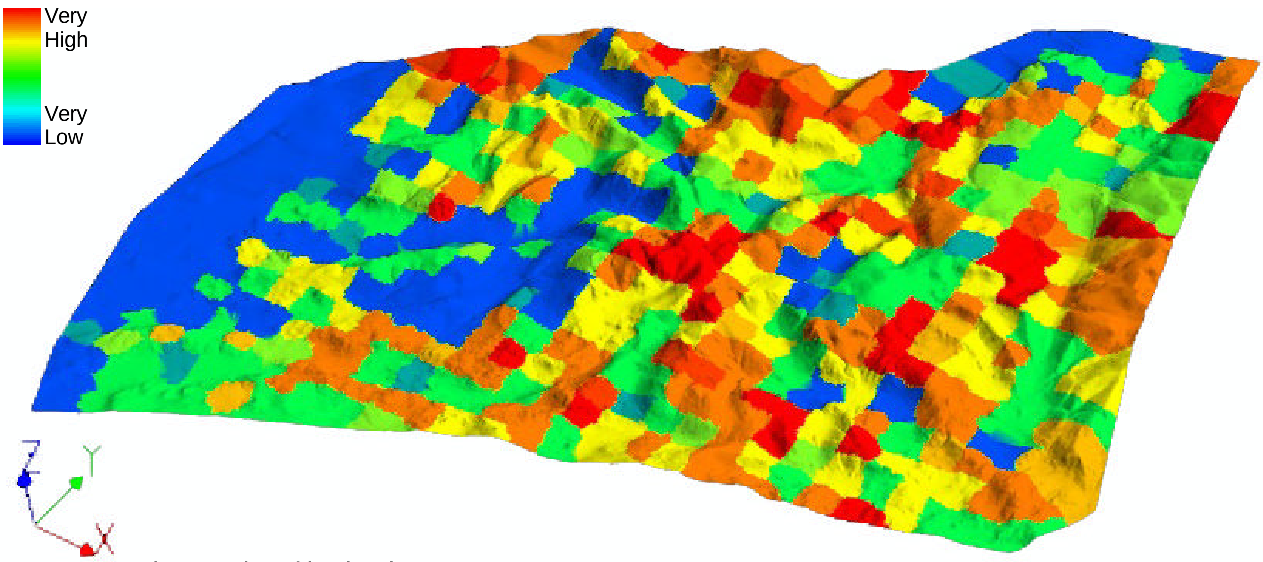
- **Data quality**
  Most of the data used in the interpretation are from first-hand observation and were collected specifically for the purpose of making a geological map. If follows that the *data quality* score in the quality assessment table will be *'very-high'* in most cases and *'high'* in cases where historical data have been used in the absence of field survey data.
- **Geological theory**
  The geological theories employed in the making of the Lake District geological map are well established and highly accepted. If follows that the *geological theory* score in the quality assessment table is *'very-high'* irrespective of the grid square.
- **Method**
  The Lake District geological map was constructed after extensive field survey, combined with iterative re-interpretation. It follows that the *method* score in the quality assessment table is *'very-high'* irrespective of the grid square.
- **Objectivity**
  The geological interpretation over the full extent of the map is based, to some extent, on a geological understanding of the area. It follows that the *objectivity* score in the quality assessment table is *'high'* irrespective of the grid square.
- **Expert knowledge**
  All geologists that worked in the Lake District are experts in fields of geology applicable to this area and therefore the *expert knowledge* score of the quality assessment table is *'very-high'* irrespective of the grid square.

The *auditability*, *calibration* and *validation* scores in the quality assessment table (Figure 3.2) will vary between grid squares depending on the fit of the interpretation to the original data and its exclusivity. It is variation in these parameters, when averaged with the others, that produces the variation in interpretation quality over the extent of the Lake District shown in Figure 4.2b.

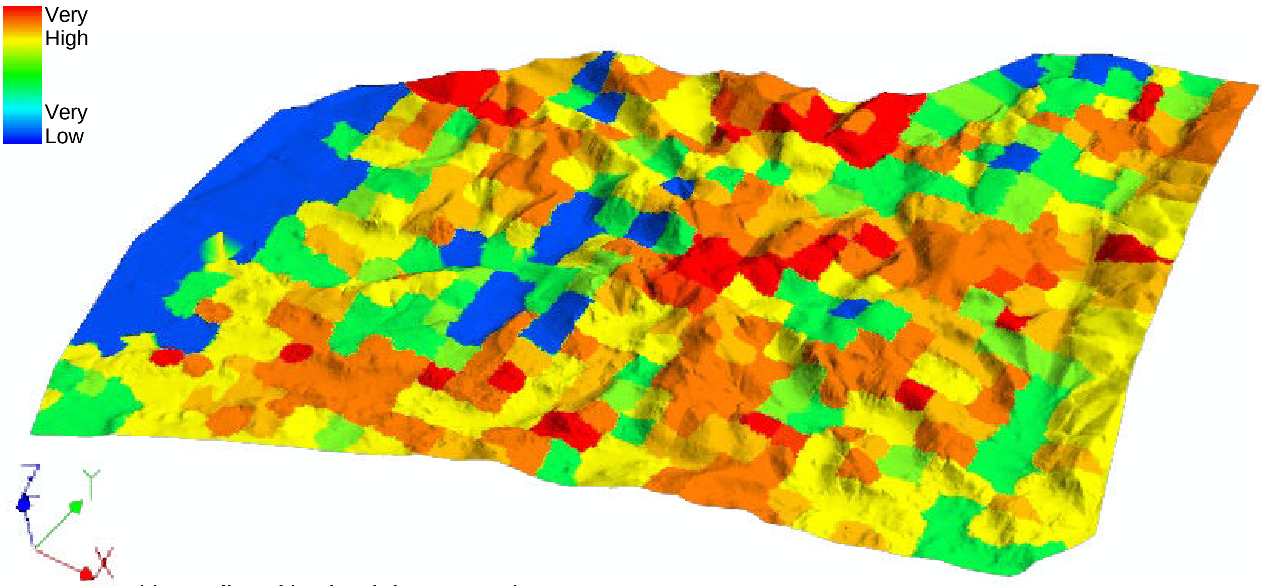
4.1.2 Confidence in geological maps - discussion

Quantity and quality scores for the bedrock geological interpretation are combined into a confidence plot (Figure 4.2c) using the graph shown in Figure 3.3. The result shows a wide spatial variation in confidence over the extent of the map. The high confidence interpretations are to be found on the exposed mountain tops, and low confidence interpretations are to be found in valleys, on the flat coastal plains and urban centres. This result is perhaps intuitive, but it can be specifically related to factors of both data quantity and interpretation quality that vary as a function of the relationship between geology and topography.

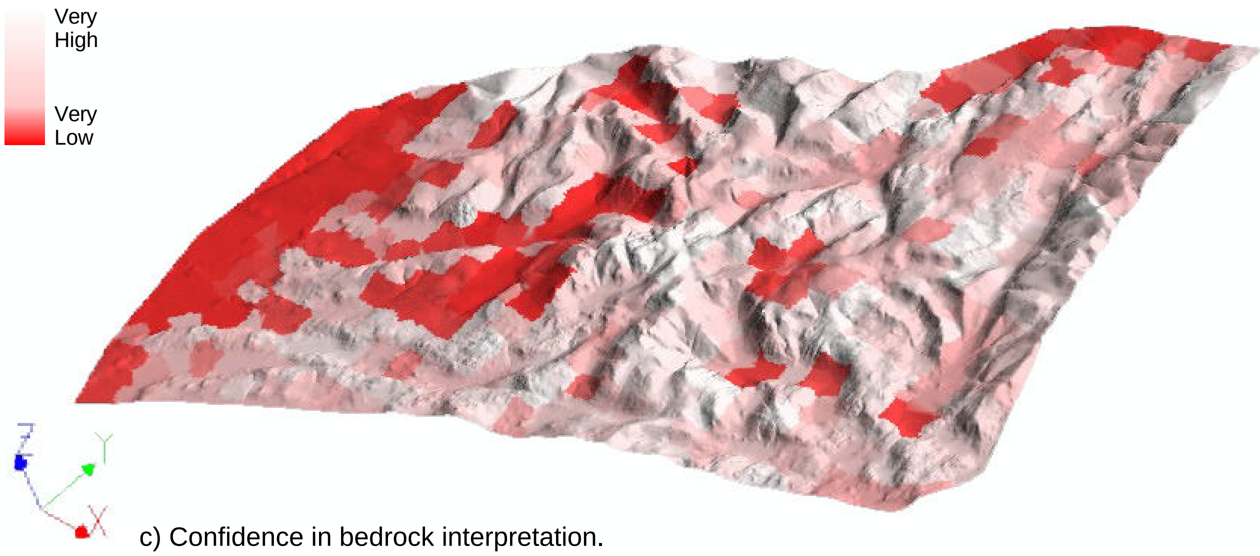
On the mountain tops, the lack of vegetation and superficial deposits clearly dramatically increases the quantity of raw data available to make a geological interpretation. In the valleys



a) Quantity of bedrock exposure.



b) Quality of bedrock interpretation.



c) Confidence in bedrock interpretation.

Figure 4.2. The confidence in the geological bedrock interpretation of the Lake District (draped onto the DTM). Confidence (c) is a combination of the quantity of bedrock exposure (a) and the quality of interpretation (b). Note how areas of high confidence coincide with maximum exposure and areas of low confidence coincide with areas of cover and urban settlement.

and on the coast plain, the extensive amounts of superficial deposits significantly reduce the quantity of raw data.

Confidence is affected by interpretation quality in addition to data quantity. The variation in interpretation quality over the extent of the map is largely a result of interaction of the geological structure with the topography. In the simplest case, deeply incised topography gives many more control points, and therefore limited scope for interpretation, compared with gently undulating terrain. However, the structure of the geology has a significant additional effect. Errors in the field measurement of dip angle for steeply dipping strata will have less of an effect on the projection of contacts along strike or down dip (through mountains) than shallowly dipping strata. Strata with dip magnitudes similar to that of the terrain surface can have a large spatial variation in projected interpreted boundaries for very small errors in measured dip angle. Folded strata further intersect the topography, thus reducing the scope for interpretation in unexposed areas. Faulted strata can constrain interpretation in areas where a fault is the only possible way of making the interpretation fit the data (validation), but faults may also decrease confidence of interpretation in poorly exposed terrain by providing a multitude of possible interpretations.

These effects of geology, and topography on map interpretation are captured in the assessment processes under the categories of *auditability, calibration,* and *validation* (and to some degree under the category of *expert judgement*). However, the *quality* score under these categories is implicitly related to the *quantity* of raw data and this is assessed independently. This explains the strong correlation between data quantity (Figure 4.2a) and confidence (Figure 4.2c). High levels of raw data lead to high confidence in interpretation. The correlation is not perfect and interpretative quality issues play a part (particularly around poorly exposed edges of well exposed ground), but a crude approximation to confidence in a bedrock geological map can be achieved by equating it to percentage bedrock exposure – a calculation that can be largely automated using GIS.

4.2 CROSS-SECTIONS

Geological cross-sections accompanying Geological Survey maps provide an interpretation of the third dimension. Cross-sections are based on the extrapolation of surface data to the subsurface in combination with geological theories, recognised local structural styles and data from other subsurface sources. Confidence in the subsurface interpretation can vary greatly over the extent of one section and between sections dependent on the confidence in the surface interpretation, the local structural style and the presence of additional constraints on interpretation in the subsurface. Some cross-sections are highly constrained whereas others are speculative and designed to represent regional relationships.

4.2.1 Geological interpretation confidence

In the example shown here, confidence in the bedrock interpretation (position of the horizons and their lithostratigraphical interpretation) is assessed for two 1:50 000-scale geological cross-sections of the Lake District. The sections are part of the cross-sections shown on BGS 1:50 000 sheets 029 (Keswick) and 038 (Ambleside).

In the Lake District DGSM project (Clarke, 2004), these sections, along with the geological map, formed part of the input interpretations to the three-dimensional model (Section 4.1). Given the three-dimensional relationship between these and other data, it was necessary to divide the cross-sections into evaluation areas (for the assessment of confidence) based on their intersection with the National Grid (rather than of equal units along their length), in order that confidence data from these and other sources fit together in three dimensions. The rationale behind this approach is given in Section 5 – Three-dimensional confidence. As the sections are not drawn in a north – south or east – west direction they are not parallel to the National Grid lines and therefore the spacing of the horizontal divisions on the section is not equidistant. The vertical divisions on both sections are at 500 m intervals relative to Ordnance Datum.

4.2.1.1 DATA QUANTITY

The quantity of raw data with which to make the cross-sectional interpretation is based on the surface bedrock exposure. The same scale as that used in the evaluation of the bedrock geological map is employed to assess the quantity of information available in each of the defined divisions along the length of the outcrop of the cross-section. Given that the same factors are being evaluated and the same grid spacing is used, the data quantity values for the evaluation areas along the line of section should be the same as those for the grid squares of the bedrock geological map through which the section passes, and may be derived directly from it.

In the subsurface, theoretically the quantity of data available for each division decreases rapidly with distance below the topographical cut. In practice, many surface data can be reliably projected into the shallow subsurface. For this reason, the ‘quantity’ of data with which to make an interpretation of the subsurface should be assessed based on the quantity of surface data that can be reliably projected into the subsurface given the local structural style and the cross-section orientation.

In addition to the quantity of surface data in the plane of the section, the quantity of data available at any given point in a cross-section must take into account subsurface data available from other sources, such as borehole or geophysical data.

4.2.1.2 INTERPRETATION QUALITY

The quality of the cross-sections is assessed using the assessment table shown in Figure 3.2. Given the standard techniques employed in constructing cross-sections for BGS 1:50 000-scale maps, the assessment categories of *geological theory, method, objectivity* and *expert knowledge* can be assumed constant over the extent of the cross sections, in a similar manner to the assessment of quality in the bedrock geological map.

The quality of data employed in the construction may well show variation over the extent of cross-section. The very-high quality field data used to construct surface relationships may be combined with many additional sources of data to interpret the subsurface. These data may have been collected for different purposes, at different times or by different organisations. In most cross-sections it will be inaccurate to assume that data quality is constant over the cross-sectional extent.

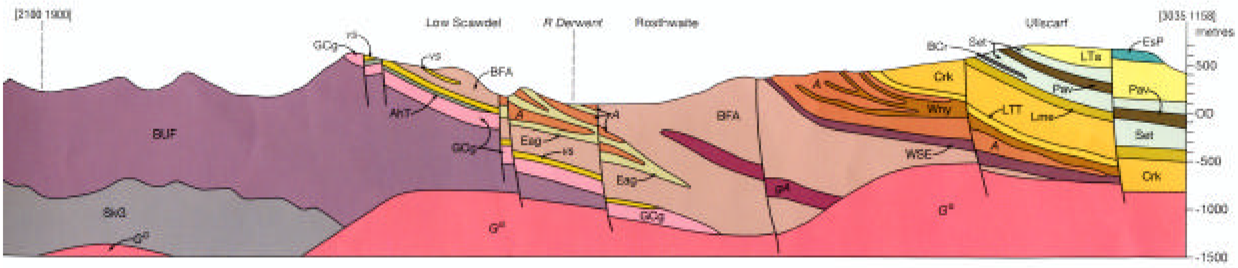
The auditability, calibration and validation scores in the quality assessment will vary over the extent of the cross-section dependent on the fit to the original data and the exclusivity of the interpretation. Commonly, a lack of data in the subsurface will result in low quality scores in these categories as the interpretation cannot be linked with original data and many alternative interpretations may exist.

4.2.2 Confidence in geological cross-sections – discussion

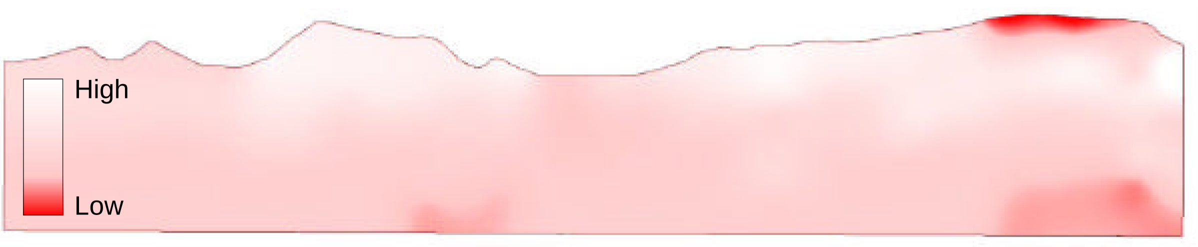
Confidence in geological cross-sections 029 and 038 is shown in Figure 4.3. These cross-sections are particularly pertinent to this investigation as they demonstrate some confidence relationships that may not be intuitively obvious (or correct).

Intuitively, and subject to the assumption that the only available data with which to make the cross-sectional interpretation are surface data, confidence in the interpretation of the position of boundaries within the subsurface must decrease rapidly below the surface (Figure 4.4a). The rate of decay is subjective and related to the geological style, but in general it is unlikely to be linear. An exponential decay curve is probably the most likely outcome. In practice, this simplistic

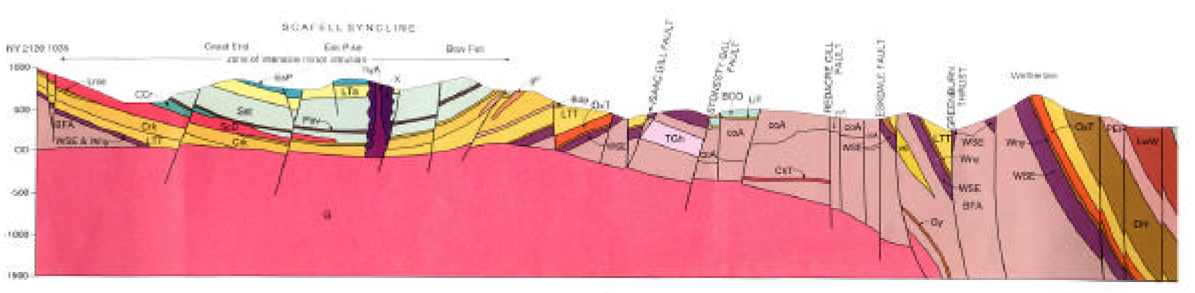
SECTION 029 (KESWICK)



SECTION 029 (KESWICK) - CONFIDENCE



SECTION 038 (AMBLESIDE)



SECTION 038 (AMBLESIDE) - CONFIDENCE

Figure 4.3. Confidence in the interpretative sections used in the Lake District DGSM (Scafell Caldera sub-area). 1:50000 Sections taken from BGS (1996;1999b).

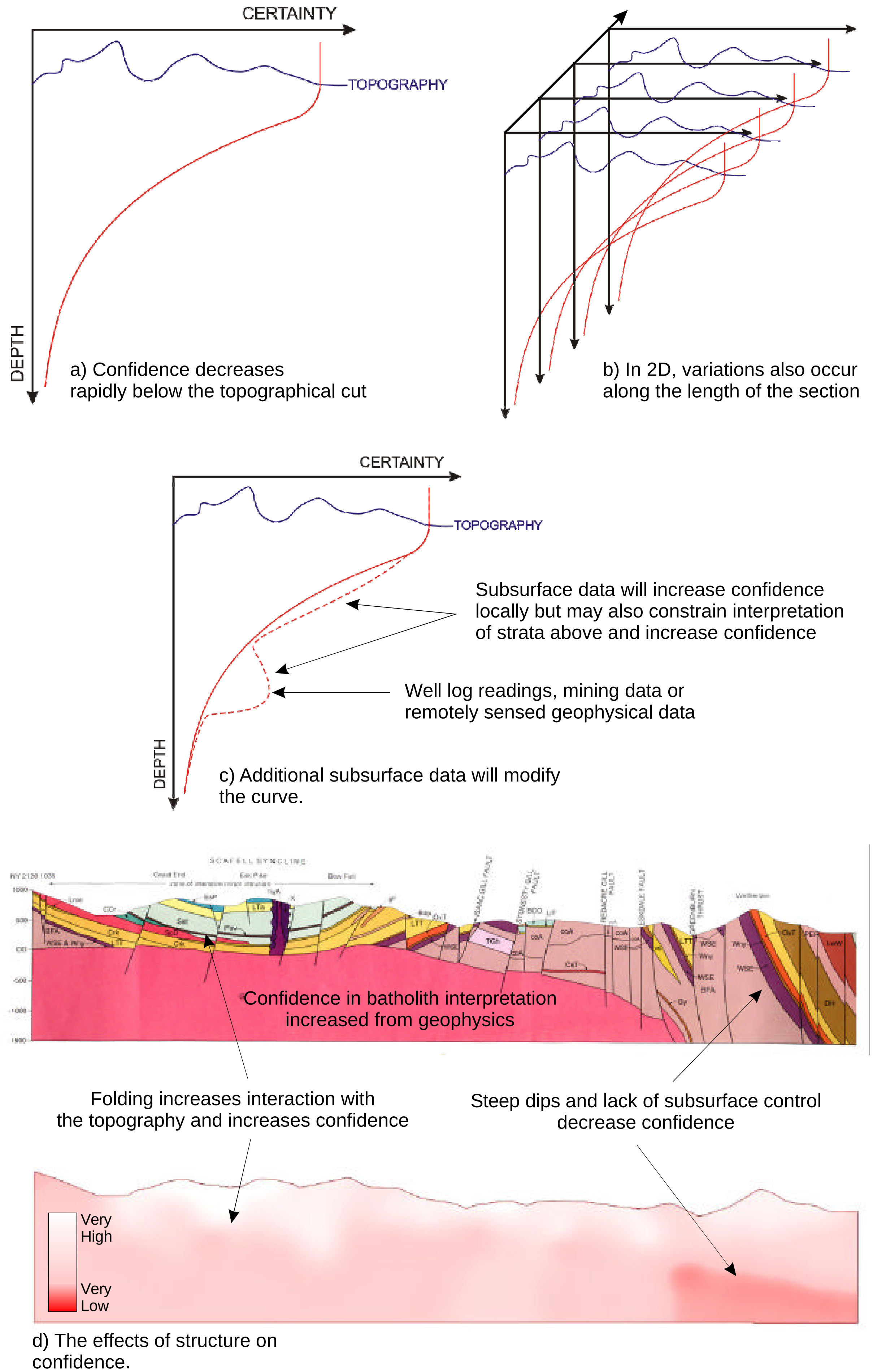


Figure 4.4. The theoretical effects of depth on confidence (a) aremodified by subsurface data (b) and depend on the local geological structure (d).

Figure 4.4

subsurface 'confidence curve' is modified by two important factors; topographical cut and subsurface data.

Confidence does not decay with depth below the topographical surface, but remains reasonably constant (or may even increase) for the vertical extent of the topographic cut, then decays below this level. This is simply a result of the extra information exposed in valleys (both along the line of, and adjacent to the section) that can be used to constrain subsurface interpretation of the hills. This effect is evident in the cross-sections from the Lake District (Figure 4.4).

Supplementary data in the subsurface may significantly increase confidence at given points. Such data are often related to a particular datum (e.g. a mine adit) or a particular horizon (such as geophysical investigations) resulting in a peak in the subsurface confidence curve for a given point (Figure 4.4a) and a band of higher confidence on the cross-section. Intuitively, these subsurface data should increase confidence over a very limited zone within the subsurface but, in practice, the constraining of a given subsurface horizon with a high degree of confidence often constrains possible interpretations above it (and in limited cases, below it) and thus increases confidence over a wider zone.

The effects of supplementary subsurface data can be seen in the Lake District cross-sections. Geophysical investigations (see Section 4.3) have constrained the position of the Lake District batholith and delimit its various components in the subsurface (Lee, 1989). Information from these studies was incorporated into the 1:50 000-scale geological cross-sections (British Geological Survey, 1996; 1999) by the geologist at the time of construction. The effect of this is to constrain the position and interpretation of the top batholith surface shown on the cross-sections and increase confidence in region of this horizon (Figure 4.4d). Below the top batholith surface little is known about the geology and little can be implied from the surface exposure or the geophysical investigations. Consequently, there is a zone of very low confidence below the top batholith horizon (Figure 4.4d). Furthermore, in areas of the cross-sections where the batholith surface is not present, the rate of decay of confidence below the topographical cut is increased due to the lack of constraint on interpretation of the overlying strata imposed by the presence of a constrained position for the batholith surface. These effects are visible on sections 029 and 038 Figure 4.3.

Given the discussion thus far, some form of decay in confidence between the base of the topographical cut and the top batholith surface in the Lake District cross-sections would be expected from Figure 4.4c. Figure 4.4d demonstrates that this is not always the case. Cross-sections 029 and 038 demonstrate a wide range of structural styles over a relatively short extent. Synclinal folds and fault are present, along with both steeply and shallowly dipping strata. The effects of structural style overprint the generic decrease in confidence with depth in the subsurface. These effects are clearly visible in figures 4.3 and 4.4d. Surface exposure of both limbs of the Scafell Syncline (cross-section 038) significantly constrains both the possible interpretations of this structure in the subsurface between exposed limbs and the possible interpretations of overlying strata. As a result, confidence in the region of the Scafell Syncline is significantly increased.

Similarly, a combination of faulting, variable dips, the position of the batholith surface and the topographical cut help to constrain subsurface interpretation and increase confidence over a significant proportion of cross-section 038 and parts of 029. Where strata are steeply dipping, subsurface interpretation is not so well constrained and confidence is reduced. This effect is compounded in regions where the batholith surface is not present.

The application of confidence modelling to the cross-section of the Lake District indicates that structural style and supplementary subsurface information can have a significant effect on the confidence of interpretation in the subsurface. Though generically, confidence will decrease rapidly below the topographical surface, in practice this trend is heavily modified (and in some case obliterated) by these effects.

4.3 CONTOURED HORIZON PLOTS (GEOPHYSCIAL INTERPRETATION)

In two-dimensions, it is commonplace for geologists to interpret the position of key subsurface horizons over a map area by constructing a contoured horizon plan. These plans often 'fill the gaps' between interpretations on cross-sections and/or outcrop data. The interpolation between these known points may be simply a mathematical relationship, it may involve some degree of expert knowledge input, or it may be derived from remotely sensed geophysical investigations. Clearly, there are a large number of possible permutations, combinations and interpolations that may have been incorporated in the resulting contoured horizon plan and therefore the history of the plan is of paramount importance in assessing confidence

4.3.1 Geological interpretation confidence

The example demonstrated here is a contoured interpretation of the position of the top batholith surface over the extent of the north-western Lake District (Figure 4.5). This interpretation is based on combined gravity and magnetic studies by Lee (1989). He used the geophysical data recorded from a large number of stations across the Lake District, to interpolate a number of two-dimensional model sections (Figure 4.5b) that constrain the subsurface structure. From these sections, he linearly interpolated a contoured 'depth to top batholith' plan (Lee, 1989, fig. 8.1, p. 114) and it is this plan that forms the basis of the contoured horizon plan of the top batholith shown here in Figure 4.5. The plan of Lee has been re-projected into elevation relative to OD and combined with outcrop limits of the top batholith (from the geological map) and the interpreted batholith horizons on the cross-sections of the Lake District (these lines were interpreted from the same geophysical investigations).

Confidence in the contoured horizon plan of top batholith is assessed for the position of the batholith surface, per square kilometre on a grid based on the BNG. This plan was originally used in the Lake District DGSM and the confidence assessment needs to be applied to the same grid pattern as other input data (such as the geological map) in order that confidences can be combined in three-dimensions (Section 5.1.2)

4.3.1.1 DATA QUANTITY

The contoured horizon plan is shown in Figure 4.5b with the location of the two-dimensional modelled geophysical sections (Lee, 1989) and cross-sections 029 and 038 (from: British Geological Survey 1996; 1999) superimposed.

For those grid-squares through which the outcrop line of the top batholith passes, data quantity scores are derived from the bedrock geological map, given that this is the interpretation on which the position of the top batholith within these grid squares is based. Similarly, for grid squares through which sections 029 and 038 pass, data quantity scores are derived from the appropriate grid square on these sections. This is possible because the cross-sections are divided up into blocks for confidence modelling based on the National Grid (Section 4.2).

The plan is linearly interpolated between the position of the top batholith interpretation on the cross-sections, its outcrop position and its interpreted position on the two-dimensional geophysical sections. It follows that the only other contributing data to the plan are the original base station readings used in the construction of the two-dimensional geophysical sections.

In order to make an assessment of data quantity employed in the construction of the geophysical sections it is necessary to equate the density of stations to the five-point linguistic scale. The extensive report of Lee (1989) states that the data from the stations were interpolated to data points with a 0.5 km spacing along the length of the geophysical sections in order to perform the modelling. After lengthy discussions with Lee, the judgement was made that, given this 0.5 km spaced interpolation, four (or more) base-stations within a square kilometre would be considered

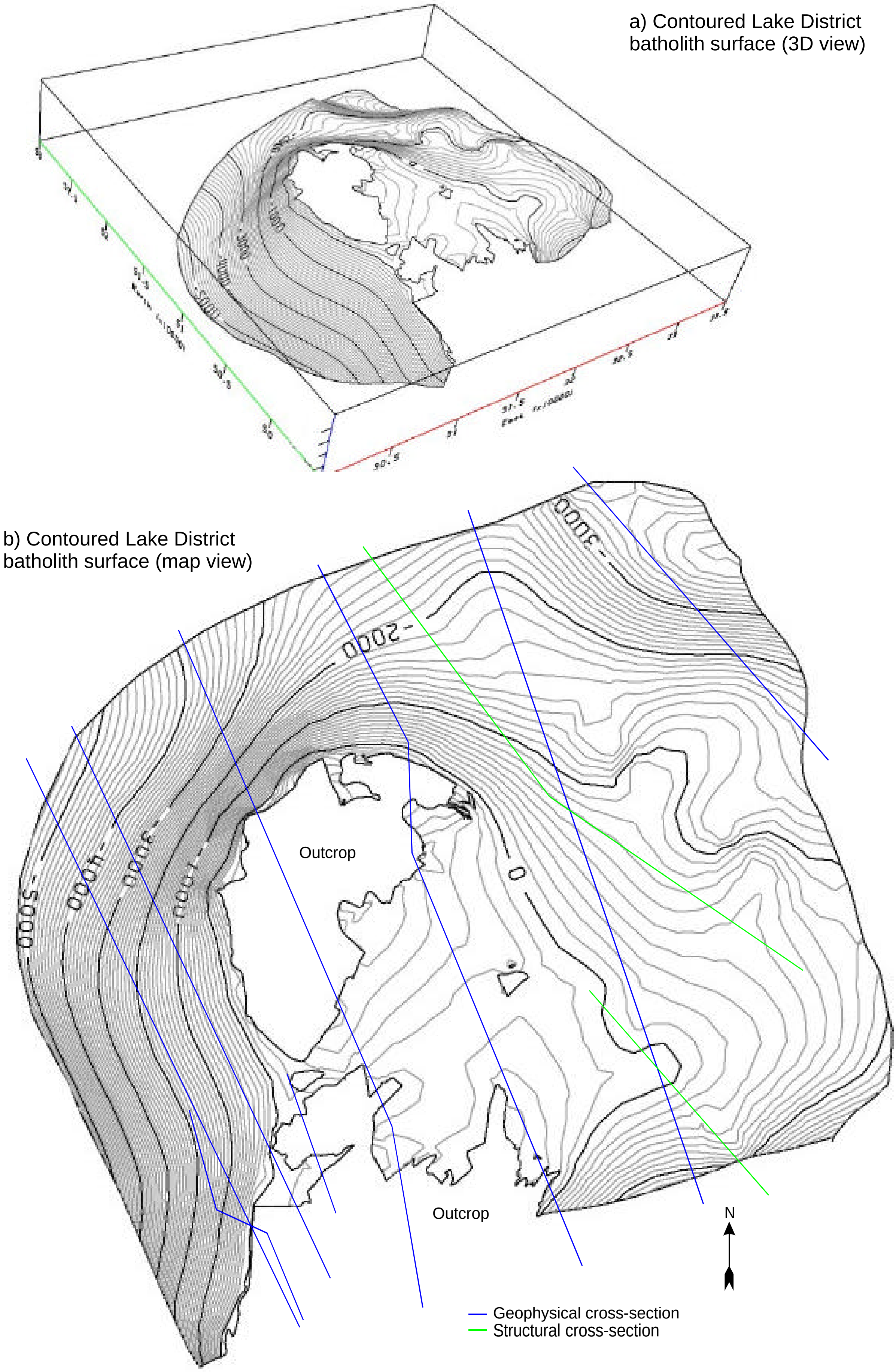


Figure 4.5. A contoured horizon plan of the elevation of top Lake District batholith surface. This plan is derived from the geophysical work of Lee (1989) combined with outcrop and cross-sectional data.

Figure 4.5

the ideal with which to make the interpolation. Four stations, evenly distributed over a square kilometre, gives a spacing of 0.5 km in the east and north directions. The remaining categories of the five-point linguistic scale are linearly distributed over the range of zero to four stations thus:

Stations	Quantity Score
4+	Very High
3	High
2	Moderate
1	Low
0	Very Low

4.3.1.2 INTERPRETATION QUALITY

In a similar manner to that for data quantity, the interpretation quality scores for those grid squares of the contoured horizon plan that intersect the outcrop or the Lake District cross-sections 029 and 038 can be derived from the interpretation quality assessments of those elements.

For the geophysical sections, interpretation quality is assessed using the table shown in Figure 3.2 with reference to the extensive discussion in Lee (1989) and controlling input from the author. Given that the geophysical data were recorded specifically for this investigation by standard geophysical methods, the quality score categories of *data, theory, method, objectivity, expert knowledge* and *auditability* can be assumed constant. The variation in interpretation quality of the top batholith on the geophysical sections is the result of variations in score under the categories of *calibration* and *validation*. These variations result largely from the significant number of other valid interpretations in some areas stemming from insufficient geophysical data, either gravity or magnetic, and therefore to a lack of validation of one interpretation.

4.3.2 Confidence in contoured horizon plans – discussion

Confidence is derived for each of the plan grid squares that intersect a geological or geophysical cross-section or the outcrop lines using the relationship shown in Figure 3.3. Given that the remainder of the plan that lies between geophysical sections, cross-sections and outcrop is mathematically interpolated and, ignoring the uncertainties involved in that mathematical interpolation algorithm (Cave and Wood, 2002), confidence can be interpolated to the remainder of the plan using the same algorithm. Confidence in the contoured horizon plan of the top Lake District batholith is shown in Figure 4.6.

Contoured horizon plans represent the interpreted position of a given horizon within the subsurface. They can be derived from the combination of many two-dimensional sources including both maps and sections. Confidences in these elements have to be combined with confidences in contributing data to the plan itself. When data derived by different means are to be combined in this manner, there are problems in the evaluation of *data quantity*. The five-point linguistic scale (and associated numerical scores) is a convenient method of evaluation, but the results for data derived by different means are only comparable if the *ideal* data quantity is well defined. This is often somewhat subjective.

In the example of the Lake District batholith contoured horizon plan (Figure 4.5), the data sources are surface observations and geophysical readings. It has been argued here (Section 4.1) that the *ideal* data quantity for making a bedrock geological map is 100% exposure per 1 km^2 given that this requires no interpolation. For geophysical interpretations based on virtual data points 0.5 km apart, interpreted from the available real data, the *ideal* is 4 (or more) readings per

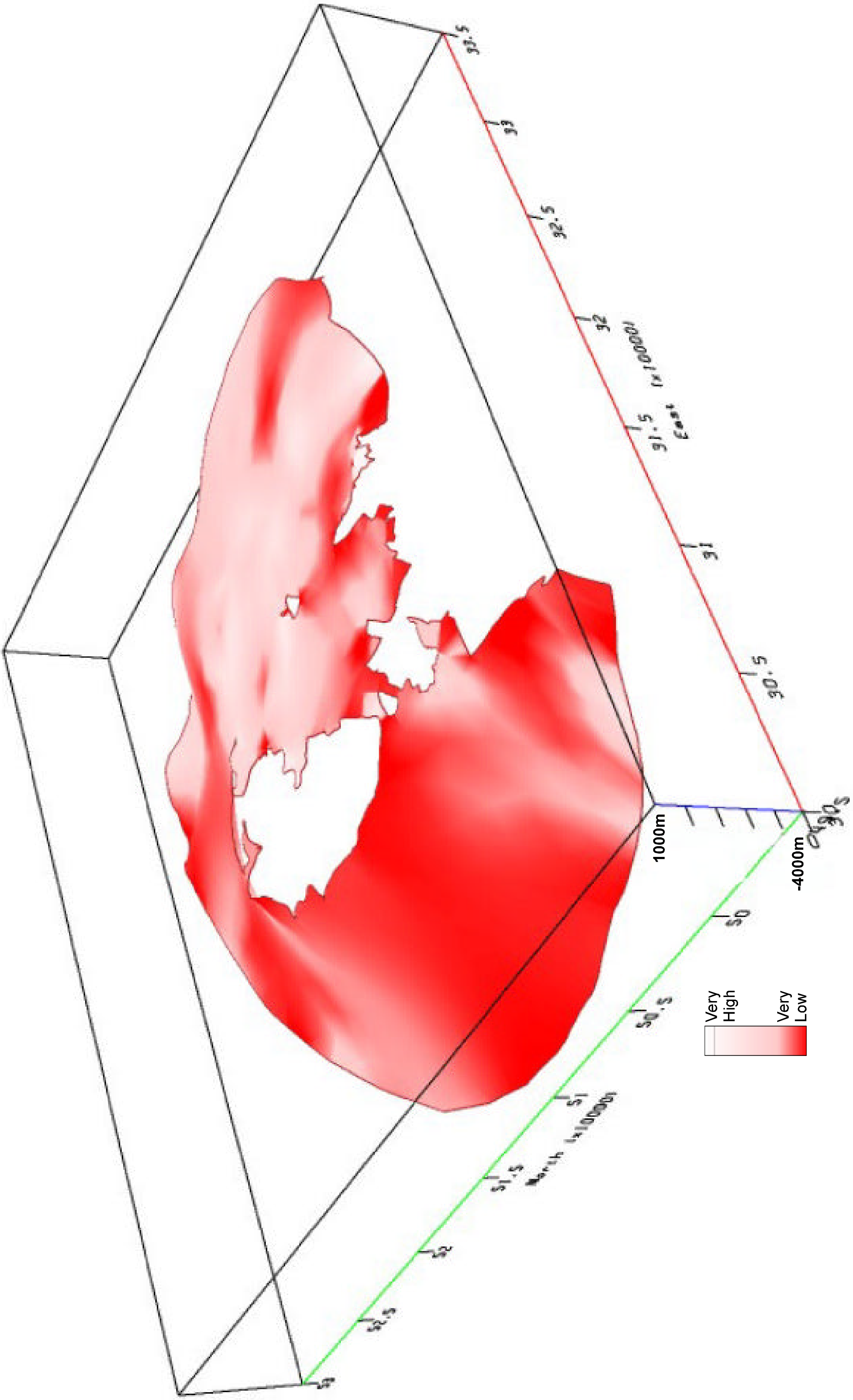


Figure 4.6. Confidence in the top Lake District batholith surface based on the gravity and magnetic modelling work of Lee (1989).

1 km^2, given that larger amounts of data will be effectively sub-sampled. The fact that these evaluations can be considered comparable is, admittedly, a moot point. Nevertheless, in the Lake District, the use of this approach to assess confidence in the position of the top of the Lake District batholith has produced a result that correlates strongly with the reported ‘feel’ of reliability in the result given by the original author (Lee, 1989).

5 Confidence in 3D modelling

The assessment of geological confidence in two-dimensional interpretations is potentially of great use in evaluating those interpretations and the detail they imply about the subsurface. Although it is a valuable tool when used in this manner, confidence modelling is infinitely more valuable as a tool to aid decision making in three-dimensional modelling.

One or more two-dimensional interpretations may be used as the input data to a three-dimensional model and, in this scenario, confidence assessments can be used to aid three-dimensional interpretation by resolving data conflicts and to propagate assessments of geological confidence to the resultant model.

5.1 IMPLEMENTING CONFIDENCE IN 3D

Two-dimensional confidence assessments are based on criteria that relate specifically to an individual interpretation (Section 3.2). The purpose of the assessment may be different between different interpretations, the defined areas on which the assessment is based may be different and the original raw data may be entirely different. To combine two-dimensional interpretations in three-dimensions these differences must be harmonised.

5.1.1 Assessment criteria

Section 3.2.2 stated the need for a set of criteria on which the confidence assessment is to be conducted. It is not possible to assess confidence in interpretation purely on the likelihood of the interpretation being right. It is necessary to consider which aspects are to be assessed.

If two-dimensional interpretations are to be combined in three-dimensions, it is clear that each interpretation must be assessed on the same geological criteria. It is not possible for interpretations assessed on different criteria to be combined, as the results are not comparable. For the purposes of three-dimensional modelling from two-dimensional interpretation, it is usually the spatial position of given horizons and/or the lithostratigraphical interpretation that are of importance to the modelling, and assessments of confidence in each input two-dimensional interpretation should be made on those criteria.

5.1.2 The confidence volume

Confidence assessments in two-dimensional interpretations are based on defined areas (Section 3.2.1). These areas may be of irregular or regular size and shape, and need not be the same in separate interpretations. However, if two-dimensional interpretations are to be combined within the same three-dimensional model it is important that the defined areas are of the same magnitude and have the same boundaries in all interpretations. In this way, the integration of the two-dimensional interpretations into a three-dimensional model defines a coherent ‘confidence volume’. When different areas are defined in different interpretations a coherent confidence volume is not achievable and confidence assessments cannot be combined in three-dimensions.

The most practical way to generate a coherent system is to assess all map-based data using a square kilometre grid (or multiples/divisions there of) based on the OS National Grid. In this way, the area of assessment is the same between interpretations and the boundaries of that area are in the same place. Cross-sectional interpretations should be divided along their length by vertical lines coincident with the National Grid lines that the cross-section intersects, and vertically by horizontal lines of a suitable spacing based on Ordnance Datum. Unless the section line is parallel to the National Grid (north-south or east-west), the horizontal divisions on the section will not be equidistant in the plane of the section but they will fit coherently into confidence blocks in the three-dimensional volume.

This approach has implications associated with mathematical sampling and it is important that a volume block size is selected such that is both practical and representative of the geological detail and confidence variation in all of the contributing interpretations. Large sample area sizes suitable for one interpretation may mask (average out) confidence variations in another interpretation. In practice, it has been found that a volume block size of 1 km by 1 km in the map plane and 500 m in the vertical plane is suitable for assessing confidence in three-dimensional models derived from standard 1:50 000-scale map-based interpretations and cross-sections (Clarke, 2004).

5.1.3 Relative confidences

If confidence assessments are always performed against the same scale of qualifying statements, confidence results from different interpretations are largely comparable (assuming the same assessment criteria are used – Section 5.1.1). Differences may occur in situations where different types of raw data are employed for different interpretations. In these cases, comparable confidence scores rely on the justification of what is considered the ideal quantity of data (Section 3.1.1).

In general there will be a density of data for each two-dimensional interpretation above which increased densities of data have no effect on the outcome due to sampling issues involved in the modelling process. If this level is taken as the ideal data quantity for the purposes of confidence modelling, then the confidence results for different interpretations can be considered broadly comparable.

For interpretation of field observations the ideal data quantity is clearly 100% exposure, but for geophysical investigations the ideal data quantity is less obvious and depends on the parameters of the modelling process (see the example in Section 4.3).

5.2 USING CONFIDENCE IN 3D

Combining two-dimensional interpretations into three-dimensions often results in areas of conflict between contributing interpretations (Clarke, 2004). Confidence assessment can be used to resolve these conflicts to produce a coherent, three-dimensional model and to record the fact that contributing interpretations conflicted in that model.

5.2.1 Evidence-based confidence assessment

Two or more two-dimensional interpretations may not agree on the position of a given horizon at a given point and, when modelling in three dimensions with two-dimensional interpretations, it is not acceptable to simply ‘average-out’ mismatches. Averaging is a purely mathematical technique that takes no account of the specific geological limitations inherent in different types of two-dimensional interpretations. Confidence assessments can be employed in order to determine the most likely position of a horizon at a given point.

For two interpretations that disagree on the position of a given horizon, the most realistic solution is simply the position of that horizon in the interpretation with the highest confidence. With three or more interpretations, the situation is not so clear-cut. Often a number of interpretations of varying confidence will agree on the position of a given horizon, whilst others (also with varying confidence) will refute this position and may agree on another; additional interpretations may agree with neither position. The selection of the best position for use in the three-dimensional model needs to be a factor of both the number of interpretations that support a given position and their relative confidences. This approach uses the principles of *evidence-*

based logic (Funtowicz and Ravetz, 1990; Bowden, 2003). Pieces of evidence, and their relative strengths, are used to support a particular outcome over others.

Given several interpretations of the subsurface, a number of scenarios exist. In a simple case of a number of two-dimensional interpretations of the position of a given horizon, all with exactly the same confidence, the mostly likely position for the given horizon is that position indicated by the largest number of interpretations; i.e. the statistically most likely position. Similarly, if the interpretations have different confidences and all of them indicate different unique positions for a given horizon then the mostly likely position is that indicated by the interpretation with the highest confidence.

However, these scenarios are end-member cases and in practice the interpretations have different confidences and some agree on one position but others agree on another. This scenario is complicated and the statistical effect of the number of interpretations that agree on a given position must be weighted by their relative confidences.

Work in resolving these issues in the Lake District DGSM in order to construct a coherent three-dimensional model from two-dimensional input interpretations (Clarke, 2004) has led to the development of the following methodology based on the principles of evidence-based logic. In the Lake District, this technique has proved sound but, as yet, it is largely untested in models of other areas or models that use less conventional two-dimensional interpretations.

5.2.1.1 Evidence based confidence evaluation

In situations where a number of two-dimensional interpretations suggest a number of possible positions for a given horizon within a given confidence volume block, the interpretations should be separated into groups of interpretations that support each other. Interpretations within a group should agree on the position of the given horizon. They need not agree exactly, but should agree within the scale resolution and accepted measurement error of the three-dimensional model. Overall confidences can then be derived for each group of interpretations based on the following argument.

On the normalised numerical scale of confidence, 1 (one) represents total, unequivocal, confidence in the interpretation. An interpretation with a confidence score of 1 defines the absolute and definite position of a boundary. 0 (zero) represents absolutely no confidence in the interpretation. An interpretation with a confidence score of 0 is a complete fabrication. Using this assumption and defining confidence (C) as the negative of uncertainty (U) it follows that:

$$C = 1 - U \qquad [5.1]$$

The confidence scores of interpretations that support each other can then be combined by selecting the interpretation with the highest confidence score and using the confidence scores of supporting interpretations to reduce the uncertainty in the selected interpretation. Given that confidence and uncertainty sum to 1 (Equation 5.1), the confidence scores of supporting interpretations reduce the uncertainty component of the selected interpretation by a factor equal to their respective confidence scores:

$$C_{TOTAL} = C_1 + (C_2(1-C_1)) + \ldots.. \qquad [5.2]$$

Where C_{TOTAL} is the combined confidence of two interpretations with individual confidences of C_1 and C_2 that support each other.

This process is iterative and additional supporting interpretations can be used to further reduce the uncertainty component of the group. Logically, supporting interpretations should be iteratively combined in order of decreasing confidence. In mathematical practice the order of combination is irrelevant.

Worked example:

A selected interpretation has confidence in a horizon position of 0.7 (*high*) and a second interpretation that supports this position has a confidence of 0.5 (*moderate*). A third interpretation also supports this position, but has a confidence of 0.2 (*low to very-low*).

Combined confidence in the first two interpretations is:

0.7 + (0.5*0.3) = 0.85.

The effect of the third interpretation is:

0.85 + (0.2*0.15) = 0.88

Total confidence in position indicated by all three interpretations is 0.88 (*very-high*)

In this way combined confidence scores can be elicited for each group of interpretations and therefore for each conflicting position of a given horizon. The combined confidence is a weighted statistic of the number of supporting interpretations and their relative confidences and can be used to determine which is the most suitable position. The process is repeated for each confidence volume block of the three-dimensional interpretation.

Theoretically, it is entirely possibly that, using this approach, a group with a number of low confidence interpretations may have a combined confidence that outweighs a single conflicting high confidence interpretation. In practice, if a number of interpretations agree on a position they should have individual confidences at least similar and probably greater than a smaller number of interpretations that refute that position. In other words, the group with the largest number of interpretations should contain those interpretations that have the highest confidence. If this is not the case it suggests that something fundamental to the interpretative process of one or more interpretations has been ignored in the confidence assessment.

Using rigorous assessments of confidence to determine the position of horizons results in a three-dimensional model that is a weighted best fit to the available data rather than a simple averaged best-fit model lacking in geological value.

5.2.1.2 Circular Support

The evidence-based approach to combining confidences from different interpretations inherently assumes that the interpretations were constructed independently. In many examples of two-dimensional geological interpretations, supporting evidence from additional sources is incorporated into the interpretation at the time of construction. Cross-sections 029 and 038 in the Lake District (British Geological Survey, 1996; 1999) are examples of this. The geophysically derived position of the top of the Lake District batholith was incorporated into the interpretation at the time of construction and used to constrain the remainder of the interpretation (Section 4.2).

It is not possible to use the position of the top batholith on cross-sections 029 and 038 to support the geophysical interpretation of the top batholith at these points within a combined three-dimensional model. In the evidence-based logic methodology, such a practice is termed *circular support* and artificially increases three-dimensional confidence.

Two solutions to this problem exist:

1) The effects of the additional data on the two-dimensional interpretation can be ignored in the confidence assessment of the interpretation itself and then factored into the three-dimensional assessment by following the evidenced-based logic approach in the normal way.
2) The effects of the additional data are considered in the confidence assessment of the two-dimensional interpretation but the interpretation is not used in the evidence based logic calculation of confidence in the three-dimensional model.

In practice, option 1 is difficult to achieve as it involves an assessment of how the two-dimensional interpretation would look in the absence of the additional data. Additionally, this approach produces a two-dimensional confidence assessment that is not representative of the interpretation as it stands and cannot be used independently to indicate confidence in that interpretation. Option 2 is the most effective approach to resolve this issue.

5.3 CONFIDENCE MODELLING IN 3D – AN EXAMPLE

In the Lake District DGSM project (Clarke, 2004) a number of two-dimensional interpretations of the subsurface geology were used to construct a structural framework in three dimensions. This framework was then combined with additional two-dimensional interpretations of the structure of the Borrowdale Volcanic Group (BVG) in order to construct a three-dimensional model of the bases of key lithostratigraphical surfaces within the BVG.

This process relied heavily on assessments of confidence in all input interpretations to resolve positional conflicts within the structural framework and in construction of the BVG surfaces.

Each input interpretation was assessed for confidence in positional and lithostratigraphical interpretation on a grid carefully designed to form a three-dimensional confidence volume with blocks 1 km by 1 km by 500 m. All map-based interpretations were assessed per square kilometre using the National Grid to define the grid square boundaries. Cross-sections were evaluated based on a grid of variable lateral spacing (to fit with the National Grid) and a vertical spacing of 500 m OD.

Comparisons of confidence between interpretations allowed a coherent structural framework to be developed. Many small, localised conflicts, particularly between geological cross-sections and the geophysical interpretation of the batholith surface, were resolved using the evidence-based approach outlined in Section 5.2. During this process it quickly became apparent that a number of generalisations could be made:

- At the topographical surface, the geological map has a higher confidence level than the geophysical interpretation of the batholith or the cross-sections: i.e. the outcrop position has priority at the topographical surface;
- Where the top surface of the batholith has not been incorporated in the interpreted cross-sections, the geophysical interpretation of the surface position has a higher confidence level and therefore has priority;
- For those faults that are not interpreted as vertical, the geometrical interpretations in the lines of cross-sections are based on both surface data and outcrop structure (unit thickness and fault-block dip etc) and therefore have a higher confidence than the fault surfaces contours interpolated from surface dip alone. In regions of conflict between fault surfaces and cross-sectional fault position, the cross-sections are given priority.

These generalisations are specific to this model and will not necessarily hold in other models of other areas. However, they are the result of detailed confidence assessment in three dimensions and a similar approach in other areas will yield a set of generalisations that can be applied to that specific model.

The framework confidence and confidence evaluations in other contributing interpretations were propagated to the resultant BVG interpretation. Confidence in the interpretation of the BVG surfaces can be reviewed using the digital model. A figure is not provided here, as it is not possible to capture the three-dimensional variations in confidence in one view.

In general, the interpretation of the base of the BVG surfaces in the Lake District DGSM shows a range of confidence related to the amount and type of input two-dimensional interpretation. Confidence is particularly high around outcrop and in the shallow subsurface where input interpretations with high confidence exist (maps and cross-sections). In the peripheral regions and at depth, the interpretation is less reliable due to both the lack of data, the low confidence in those data and the large number of conflicts between data. For a full discussion on the outcomes of three-dimensional modelling with two-dimensional interpretations in the Lake District, see Clarke (2004).

6 Conclusions

Uncertainty in geological interpretation is implicitly recognised by trained geologists but it is inherently difficult to evaluate objectively. The method presented here goes some way to producing a generic confidence evaluation scheme that can be used in a largely objective manner to produce a useful evaluation of confidence in geological interpretation.

6.1 TWO-DIMENSIONAL CONFIDENCE

In two dimensions, application of the assessment method to a number of standard geological interpretations (Section 4) has demonstrated its ability to highlight those areas of the interpretation in which confidence is high and the geological model is most likely to be correct, and where confidence is low and the model is less reliable. A geological understanding of the area allows the trained geologist to appreciate these confidences in an almost subliminal manner without dedicated confidence maps, but clients and members of the public cannot have the same appreciation. In these cases, objective confidence assessment can add value to BGS products.

Studies of confidence in two-dimensional interpretations from the Lake District have highlighted some of the relative strengths and weaknesses of particular interpretation styles. Geological maps show confidence relationships with outcrop quantity and local geological structure. Geological confidence decreases rapidly below the surface in cross-sectional interpretations although additional subsurface data can be employed in the construction leading to increased confidence locally, and perhaps regionally dependent on structural style and topographical relief. Given the methods employed in construction, the structure of the geology and its interaction with topography can significantly affect confidence in cross-sectional interpretation. Geophysically derived horizon plans usually involve some mathematical interpolation of remotely sensed data combined with surface observation and therefore have widely varying confidence.

6.2 THREE-DIMENSIONAL CONFIDENCE

In three dimensions, confidence assessments are a necessity to construct models from two-dimensional input interpretations. Only with assessments of relative confidence in the input interpretations is it possible to resolve conflicts in horizon position or interpretation.

Using an evidence-based logic approach it is possible to combine confidences in supporting and refuting interpretations to assess the most confident position of a subsurface horizon from all the available evidence. In so doing, the resultant three-dimensional model becomes a weighted best-fit interpolation to the input data rather than a mathematical averaged best-fit interpretation lacking in geological value.

Confidence in input, combined with the effects of supporting and refuting evidence can be carried forward to the three-dimensional model in order to express confidence in the result. The ability to manipulate and 'fly through' three-dimensional numerical models using modern computing platforms effectively masks the uncertainty within the interpretation and suggests a level of 'correctness' above and beyond that which would normally be assumed for a standard two-dimensional representations of geology. This is not only a perception of clients and the public, but also of trained geologists. Confidence assessments in three-dimension numerical models of geology are a necessity.

6.3 UNCERTAINTY IN CONFIDENCE

The science of geology is descriptive and interpretative. It relies on some degree of 'expert knowledge' and experience in making interpretations of available data. This degree of subjective input makes it difficult to objectively assess confidence in the geological aspect of the interpretation. The method demonstrated is as objective as possible, but it is not totally mathematical and does rely on some subjective input from the interpreter, particularly in the assessment of interpretation quality. The statements within the quality assessment table (Figure 3.2) are designed to be as tight and exclusive as possible but there is still room for some subjective interpretation by the assessor.

This introduces a dilemma because the subjective component of the assessment itself introduces uncertainty to the confidence assessment. *Should an assessment of confidence be undertaken in the confidence assessment?* Any measurement of any quantity cannot be without uncertainty (Funtowicz and Ravetz, 1990). Even wholly objective mathematical calculations based on measured data have uncertainties associated with the measurement of those data. For this reason it is not possible to derive a confidence measurement that is itself not without uncertainty. The measurement of confidence in confidence is impractical, recursive and cannot be taken into account in any method of confidence assessment (Funtowicz and Ravetz, 1990).

6.4 CONFIDENCE IN THE FUTURE

The method of confidence described here is designed to evaluate *geological interpretation confidence*. There are other sources of uncertainty in geological models (both two and three-dimensional) related to issues other than geological interpretation. Some of these can be assessed using a variety of mathematical and descriptive techniques, others cannot. Some have been studied and applied to geological interpretations by BGS workers (Cave and Wood, 2002; Bowden, 2003). It is hoped that this method can augment these techniques and help to assess overall confidence in geological models.

The work in the Lake District has shown that confidence assessments can only be undertaken objectively by the geologist who originally worked on making the interpretation. Unlike assessments of other forms of uncertainty (Section 2.1), confidence in geological interpretation is also harder, more time consuming, tedious and less reliable to do retrospectively. It follows that the most objective and reliable way to assess geological interpretation confidence in standard geological interpretations is for the survey geologist to make that assessment concurrently with the survey. In this way, confidence data are collected with the minimum of overhead.

An objective, consistent and methodical approach to evaluating geological confidence captures and records the interpretative aspects of science of geology. With this information, geological interpretations (in two or three dimensions) are more rigorous, scientifically supportable and, most importantly, much more useful. Confidence assessments allow users to make informed decisions as to the suitability of data or interpretations for the purpose, especially when the purpose is not the one the interpretation was originally made for. In the Lake District DGSM project (Clarke, 2004) it would not have been possible to derive a rigorous model of the BVG without assessments of confidence in the contributing interpretations. Fortunately, the original authors of those interpretations were available for consultation; had they not been, and in the absence of any recorded confidence assessment, the project would not have been successful.

References

Most of the references listed below are held in the Library of the British Geological Survey at Keyworth, Nottingham. Copies of the references may be purchased from the Library subject to the current copyright legislation.

BOWDEN, A. 2003. *Presentation and discussion of evidence support logic (ESL). A note to the British Geological Survey. Unpublished.*

BRITISH GEOLOGICAL SURVEY. 1996. Ambleside. England and Wales sheet 38. Solid Geology 1:50 000 (Keyworth, Nottingham: British Geological Survey.)

BRITISH GEOLOGICAL SURVEY. 1999. Keswick. England and Wales sheet 29. Solid Geology 1:50 000 (Keyworth, Nottingham: British Geological Survey.)

CAVE, M R, AND WOOD, B. 2002. Approaches to the measurement of uncertainty in Geoscience data modelling. *BGS Internal Report,* IR/02/068.

CLARKE, S.M. 2002. The Lake District DGSM. An overview of the model and best practice guidelines. *BGS Internal Report,* IR/04/000.

DAY, R.A. AND UNDERWOOD, A.L. 1991. *Quantitative analysis*. Prentice Hall, New Jersey

DAVIDSON, A.C. AND HINKLEY, D.V. 1997. *Bootstrap methods and their applications.* Cambridge University Press.

EFRON, B. 1979. Bootstrap methods: Another look at jackknife. *Annals of Statistics*, Vol. **7,** 1-26

EFRON, B. AND TIBSHIRANI, R.J. 1993. *An introduction to bootstrap*. Chapman and Hall, New York

FERSON, S., ROOT, W. AND KUHN, R. 1999. *RAMAS Risk calk: Risk assessment with uncertain numbers.* Applied Biomathematics. New York.

FUNTOWICZ, S.O. AND RAVETZ, J.R. 1990. *Uncertainty and Quality in Science for Policy*. Dordrecht: Kluwer

IMAN, R.L. AND HELTON, J.C. 1988. An investigation into uncertainty and sensitivity analysis techniques for computer models *Risk Analysis*, Vol. **8**(1), 71-90

IMAN, R.L. AND HELTON, J.C. 1991. The repeatability of uncertainty and sensitivity analyses for complex probabilistic risk assessments. *Risk Analysis,* Vol. **11**(4), 591-606

ISAAKS, E.H. AND SRIVASTAVA, R.M. 1989. *An introduction to applied geostatistics*. Oxford Univeristy Press

KINDLARSKI, E. 1984. Ishikawa diagrams fro problem-solving. *Quality Progress,* Vol. **17**(12), 26-30.

LEE, M.K. 1989. Upper crustal structure of the Lake District from modelling and image processing of potential field data. . *British Geological Survey Technical Report*, WK/89/1

MEINRATH, G., EKBERG, C., LANDGREN, A. AND LILJENZIN, J.O. 2000. Assessment of uncertainty in parameter evaluation and prediction. *Talanta,* Vol. **51**(2), 231-246.

MCKAY, M.D., CONOVER, W.J. AND BECKMAN, R.J. 1979. A comparison of three methods for selecting values of input variables in the analyis of output from computer code. *Technometrics,* Vol. **221**, 239-245

MILLWARD, D. 2002. Early Palaeozoic magmatism in the English Lake District. *Proceedings of the Yorkshire Geological Society,* Vol. **54**, 65-93.

SHRADER-FRECHETTE, K S. 1993. *Burying uncertainty. Risk and the case against geological disposal of nuclear waste*. University of California Press.

WEHRENS, R., PUTTER, H. AND BUYDENS, L.M.C. 2000. The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems,* Vol. **54**(1) 35-52.

YOUNG, G.A. 1994. Bootstrap: More than a stab in the dark? *Statistical Science,* Vol. **9**, 382-415.