UNIVERSITY OF STRATHCLYDE

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

Uncertainty in geoscience interpretation: Statistical quantification of the factors that affect interpretational ability and their application to the oil and gas industry

by

Euan John Macrae

A thesis presented
in fulfilment of the requirements
for the degree of Doctor of Philosophy

February 2013

Declaration of Originality

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: *Euan Macrae*

Date: 11/02/2013

Abstract

Understanding the subsurface through geological modelling is extremely important to modern civilisation, e.g. the extraction of resources and the geological storage of wastes. Geological data are commonly sparse, with the result that geological models are under-constrained and multiple structural interpretations are often valid. Geoscientists are also affected by cognitive biases, so individual interpretations may not be equally likely. A better understanding of how geoscientists should be trained, and what interpretational approaches are most effective, is therefore required.

An international sample of more than 700 geoscientists, with varying technical backgrounds, and experience levels, was collected. Six reference experts were then recruited to interpret the same seismic image, allowing a scoring system to be formed to evaluate respondents' interpretations. Statistical analysis of the sample showed that respondents' experience was more influential than their education and work environment in terms of producing a valid interpretation. However, interpretational techniques relating to 'thinking about geological time' were even more influential than respondents' experience. The fact that certain techniques were statistically significant in addition to respondents' experience shows that training is important regardless of experience level.

In addition to the large sample, a separate workshop experiment, utilising a control group, was conducted with 49 industry geoscientists. Analysis of the data from the workshop identified a causal link between 'considering the geological evolution' and 'producing a valid interpretation'.

Finally, based on the results, and the analysis of relevant literature, an interpretation workflow was derived for the oil and gas industry. The workflow mitigates cognitive biases, improves team work, validates multiple interpretations and captures interpreters' evolving assumptions. Thus, this research advances the understanding of how risk arising from uncertainty in geoscience interpretation can be mitigated, and how geoscience teaching and practice can be improved.

Dedication

To Lesley-Anne

Acknowledgements

My first degree was in mathematics and statistics, so it was quite daunting to undertake a Ph.D. in a different subject area, although I did not realise this at first… The research looked straightforward, but I soon realised that I had so much more to learn (e.g. *"the more you learn, the more you realise you don't know"*). Combining geoscience with psychology, statistics and a touch of philosophy was very challenging but also rewarding. At the end of this thesis, looking back, I'm very pleased that I chose to do a Ph.D., but I could not have completed it without the support from everyone below, and possibly from others too.

I would like to thank my supervisors Zoe Shipton, Clare Bond, Rebecca Lunn and Marian Scott. Each of you has contributed massively, in different ways. Thank you for being so generous with your time, our discussions motivated me to work harder and explore the ideas further.

Thank you to NERC and Midland Valley Exploration Ltd (the industry sponsor) for funding this research. Thanks especially to Alan Gibbs and Roddy Muir (Midland Valley) for your time and the valuable conversations we had. The continued contact with Midland Valley gave me very useful insights into the oil and gas industry.

I would like to thank the Faults and Fluid Flow (FAFF) research group for the meetings and fieldtrips we had, giving me a multi-disciplinary perspective on my research. Particularly, I would like to thank Yannick Kremer for the (culinary) adventures over the years!

As this research was based heavily on human participation, I would like to acknowledge the hundreds of unnamed geoscience professionals for their participation and interest. It was great to discuss the research with you at the many conferences, fieldtrips, courses, seminars and workshops I somehow managed to attend.

I also was able to do lots of travelling, including: attending the EGU conference in Vienna, attending the AAPG ACE conference in New Orleans (courtesy of Midland Valley), field-assisting in Utah (courtesy of Yannick Kremer's Ph.D. project), and participating in a student exchange to the China University of Geosciences (Wuhan), courtesy of the Scotland-China Higher Education Research Partnership (British Council). Thanks for these brilliant, eye-opening trips into the unknown.

Thanks to Mike Shand (GIS Cartographer, University of Glasgow) for making the world map that was used in Question 21 of the questionnaire.

Most of all, I would like to thank my wife, Lesley-Anne, and my family, for being so patient with me over the last four years. Yes, it's finally finished.

Table of Contents

Chapter 1 – Introduction

Understanding the subsurface through geological modelling is extremely important to modern civilisation, e.g. the extraction of resources and the geological storage of wastes. However, geological data are inherently under-constrained because they are often sparsely distributed and because the subsurface is heterogeneous. Thus, multiple interpretations of geological datasets are possible. Geoscientists are also affected by cognitive biases when making an interpretation so individual interpretations may not be equally likely. A better understanding of how geoscientists should be trained, and what interpretational approaches are most effective, is therefore required.

In this thesis, I establish how geoscientists can become more effective by identifying what 'types' of geoscientists, and what interpretational techniques, are best in an interpretation exercise. Although the conclusions affect multiple industries, and are applicable to other types of under-constrained data, this thesis mainly concerns the oil and gas industry.

An international sample of more than 700 geoscientists was achieved during 2009 to 2011 (Chapter 3). Responses were collected at conferences, in university departments and in geoscience companies, with substantial industry participation. As a quality control measure, the collected sample was filtered to ensure that inexperienced respondents were removed, leaving 444 geoscientists, with a median experience of 10 years, for analysis. The demographics (age and gender) of the sample were also confirmed to represent the underlying population of geoscientists.

The survey employed a 'background' questionnaire that collected detailed information on respondents' education, work environment and professional experience. The questionnaire also included an interpretation exercise, which was the same for all respondents, allowing comparisons to be made between respondents' interpretation, their backgrounds and the techniques used.

To form a scoring system for respondents' interpretations, six reference experts with different technical backgrounds and a median of 24.5 years of experience were recruited (Chapter 4). The reference experts were asked to complete the interpretation exercise and explain their interpretation in detail. All six interpretations honoured the data and were geologically valid, i.e. possible under the laws of geology. Although all six reference experts' interpretations fitted one of two tectonic concepts, there were differences in the fault geometries, the timing of faults and even the terminology used.

The scoring system evaluated respondents' interpretations by quantifying 'how similar' they were to at least one of the reference experts' interpretations. A multivariate statistical analysis of respondents' backgrounds, the interpretational techniques used and respondents' similarity score was then conducted (Chapter 6). The analysis quantified the effect of individual factors and identified what aspects of respondents' backgrounds, and the interpretational techniques used, were most associated to producing a similar interpretation to the reference experts.

Although the survey produced statistically significant results, I wanted to test for evidence of causality and determine just how beneficial one particular technique was. Hence, a separate workshop experiment was conducted with 49 oil and gas industry geoscientists, with a median of 17 years of experience (Chapter 7). The participants were split into two groups and, unknown to them, were given different instructions, allowing a comparison of the effect of the different instructions on their interpretations to be made. Qualitative data was also collected during the workshop experiment via group discussions and written questionnaires. Participants were asked to describe how they approached the interpretation exercise and what they thought 'common practices in industry' were.

Finally, by integrating the results of this research with relevant literature from geoscience and psychology, an interpretation workflow was derived (Chapter 9). The workflow accounts for, and mitigates, the risk arising from the uncertainty in geoscience interpretation, and is particularly suited to the oil and gas exploration industry. Hence, this thesis documents how geoscientists can become more effective, and informs geoscience training, to ensure that the next generation of geoscientists are able to meet the challenges of the twenty-first century.

Chapter 2 – Literature Review

In this chapter, I review literature that is relevant to uncertainty in geoscience interpretation. I explain why geological data are uncertain, how geological models are made and then tested. I note the importance and impact of individual and group biases. Parametric and conceptual uncertainties are compared. I explain why conceptual uncertainty is often more important to assess in geoscience interpretation than parametric uncertainty, and explain why considering multiple interpretations of a dataset can be beneficial.

2.1. Geological modelling of the subsurface

The subsurface has many important uses in modern civilisation (e.g. extraction of hydrocarbons, minerals and rare earth metals; geological storage of CO_2 and of radioactive wastes). Better understanding the subsurface is important to the oil and gas industry because the success rates of exploratory drilling and the production performance of hydrocarbon reservoirs can be potentially be improved. Greater efficiency is also environmentally beneficial. In the following sub-sections, I summarise why geological data are uncertain, what geological models are, how seismic data are interpreted, how 2D interpretations of seismic data can be validated and the impact of producing an ‘invalid’ geological model.

2.1.1. Geological data are uncertain

Geological data (i.e. our measurements) are inherently under-constrained, i.e. uncertain, because they are sparsely distributed and because the subsurface is heterogeneous. Therefore, remote sensing techniques are often used for data collection. Examples of geological data include: seismic data, petrophysical data, drilling data, geochemical data and field outcrop data. Geological data are often sparsely distributed because of the high cost and scale of data collection and technological constraints. For example, a 3D seismic survey can cover hundreds of square kilometres and cost millions of pounds to complete (Davies *et al.*, 2004)[1]. An exploratory well (e.g. to gather petrophysical data using logging tools) can also cost

[1] The acquisition cost of 3D seismic data in the UK North Sea has also fallen from $70,000–100,000 per square kilometre in 1982 to $10,000–20,000 in 2002 (Davies *et al.*, 2004).

millions of pounds to drill, but effectively produces a 1D dataset of the subsurface since the volume of interest is much larger than the borehole volume, e.g. a ~20 centimetre diameter well is used to make inferences on the geology kilometres away. Thus, seismic surveys cover a much larger area than borehole data, but borehole data give a much higher sampling frequency, e.g. borehole data are collected every ~5 centimetres of depth; while for seismic data, the maximum vertical resolution is ~10 metres vertically and ~100 metres horizontally. The resolutions are limited by the available technology and the acquisition cost. Hence, the high cost and large geographical areas explain why geological data have a limited resolution.

Our measurements of the subsurface are also under-constrained because of its natural heterogeneity, e.g. rock properties can be highly variable over short distances. Eaton (2006) indicated that heterogeneity is a controlling influence on reservoir production and can include:

> *"Variations in grain-size, porosity, mineralogy, lithologic texture, rock mechanical properties, structure and diagenetic processes".*

Due to subsurface heterogeneity, a high-sampling density is desired, but can be costly. Heterogeneity is not always represented in geological models due to technological constraints and the difficulty of accurately modelling it.

Frodeman (1995) described geology as being *"interpretative"* and noted that traditional lab-based sciences (e.g. physics) use controlled experiments to derive new knowledge under known conditions. However, the interpretation of geological data is different[2] and requires human observation, which may be erroneous or biased (Chadwick, 1975, 1976). The challenges of the 'sparsity of data' and 'heterogeneity' are likely to have influenced Davis (2002) who wrote that geoscience is based on *"observation"* and the *"interpretation of sparse datasets"*.

2.1.2. A description of geological models

Geoscientists interpret multiple data sources to build models of the subsurface. The act of analysing and explaining geological data is termed an 'interpretation', whereas

[2] E.g. *"Geology is an experiment that was only run once and lasted 4.6 billion years"* Zoe K. Shipton.

a ‘geological model’ is the end result of an interpretation. The final geological model represents a series of geological processes that acted through time. Geological interpretation and modelling both require the use of geological knowledge to make sense of the under-constrained data. Hence, geological models aim to represent the subsurface but are dependent on the knowledge used to create them.

The uncertainty in interpretation comes from geoscientists’ limited understanding of the subsurface. For example, a fault either exists or does not exist. It is the knowledge of the fault (or of the fault properties) that is uncertain, not the fault itself. Statisticians Box and Draper (1987) stated that *“essentially, all models are wrong, but some are useful”.* Just like statistical models, geological models are never truly accurate, only (hopefully) sufficiently accurate for their purpose. Hence, there is no ‘correct’ interpretation of a geological dataset. Geoscientists should therefore be aware of the uncertainties in their interpretations and try to mitigate them, e.g. to represent the subsurface as realistically as possible.

2.1.3. The interpretation of seismic data

Seismic reflection data are one of the most common types of geological data interpreted to create geological models (Bacon *et al.*, 2003). Seismic data are derived from acoustic signals fired from a source on the surface and then reflected back by underlying strata. Comparison of the signals’ travel times and amplitudes allow the subsurface to be imaged. However, seismic data first need to be computer processed to migrate the data to their correct spatial location and to remove any seismic artefacts (e.g. multiples). The final seismic dataset represents the subsurface, but needs to be interpreted by geoscientists due to its limited resolution.

Seismic images are often displayed in a vertically exaggerated state to account for the fact that the vertical resolution is greater than the horizontal resolution. However, vertical exaggeration distorts the geometrical structures in the seismic data, adding a layer of complexity to the interpretation. Stewart (2011) found that seismic images are usually vertically exaggerated 2x to 6x, through a review of recently published research.

Due to their improved resolution and decreased acquisition costs, 3D seismic data are now more commonly collected than 2D seismic data. 3D seismic data have been used to identify new geological features that had not been previously identified due to their three dimensional shapes (Cartwright and Huuse, 2005). The role of the seismic interpreter has also changed over time. Davies *et al.* (2004) stated that:

> *"The modern interpreter must truly be a multidisciplinarian, well versed in subjects as diverse as petrophysics and sequence stratigraphy. Continued professional training is thus a priority in such a demanding environment".*

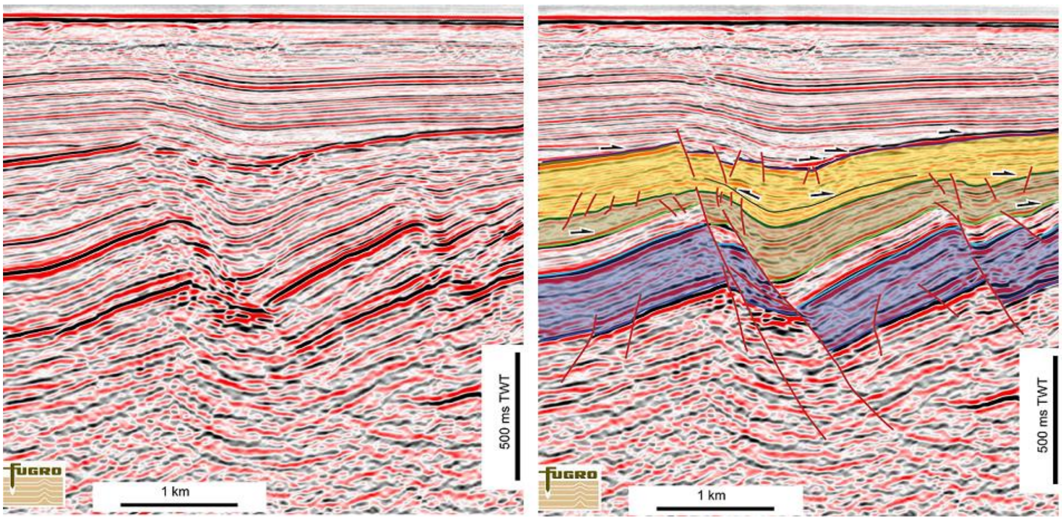
In addition to petrophysics (e.g. Archie, 1950) and sequence stratigraphy (e.g. Payton, 1977; Williams, 1993; Bertram *et al.*, 1996) geoscientists should have an understanding of sedimentology (e.g. Wadell, 1932), structural geology (e.g. Marshak and Mitra, 1988) and geophysics (e.g. Jakosky, 1950; Gao, 2009). Structural geology and geophysics can be used to determine the structural framework of a formation of interest. The rock properties (e.g. of a reservoir) can then be estimated through analysis of borehole data. 'Seismic interpreters' are hence, geoscientists with differing specialities that include petrophysics, sequence stratigraphy, structural geology and geophysics.

In the oil and gas industry, based on informal discussions with geoscientists, it is typical for seismic interpreters to analyse geological data over weeks or months, depending on the time constraints and size of a project. The results gained can then be used for other analyses. Alternatively, the same geological feature can be investigated simultaneously using different approaches. The aim of geological analysis is often to identify hydrocarbon prospects or to improve production rates, but many challenges exist, such as the integration of results between disciplines. Hence, a key challenge is to ensure effective knowledge transfer and teamwork.

This research will focus on the interpretation of seismic images (Chapter 3). In an interpretation, the tectonic concept(s) are likely to have a larger impact on the resulting geological model than, say, changing the geometry of faults. However, the tectonic setting might be well-constrained in some cases, e.g. in a mature production area.

Figure 2.1 presents an example of a structural interpretation of a seismic image. The left seismic image is uninterpreted while the right seismic image has been interpreted. The interpreter has correlated reflector horizons (black lines) across the seismic image, coloured in stratigraphic units (yellow and blue packages), identified faults (red lines) and identified onlapping relationships (black arrows). The interpreted seismic image thus conveys a large amount of information, describing possible relationships within the seismic image. However, other geoscientists may not agree with the interpreter's (subjective) decisions and may interpret the seismic image differently.

Figure 2.1– Illustration of the structural interpretation of a seismic image. The seismic image on the left is uninterpreted, while the seismic image on the right has been interpreted. Image owned by Fugro, and courtesy of the Virtual Seismic Atlas: http://www.seismicatlas.org. Interpreter was Prof. Robert Butler.



The structural interpretation thus forms the framework for the geological model and can then be populated with rock properties. The parameters may be static (e.g. averages) or have pre-defined distributions.

2.1.4. How are 2D interpretations validated?

Although the 'correct' interpretation of a geological dataset is unknown, it is possible to identify 'invalid' interpretations. I define 'invalid' interpretations to be those that are not possible under the principles of geology, i.e. cannot occur naturally. In addition to being 'valid', interpretations also need to 'honour the dataset'; meaning that interpretations cannot contradict the observed data, e.g. an extensional fault cannot be added to an interpretation if there are no discontinuities in the reflector horizons.

Due to the uses of geological models, it is important to confirm that geological models are valid and honour the dataset.

The main method to validate interpretations of 2D seismic images is to check whether the interpretation (i.e. a cross-section) 'balances'. Cross-section balancing is a formal process to validate an interpretation in terms of its bed lengths or areas. Balanced cross-sections were first introduced by Chamberlin (1910) and then formalised by Bally *et al.* (1966), Dahlstrom (1969) and Elliott (1983). There is a large literature on the topic of balanced cross-sections, which is reviewed by Groshong *et al.* (2012). Recent additions include Judge and Allmendinger (2011), who investigated the error propagation of input parameters on horizontal shortening estimates. Also, Woodward (2012) advocated using additional data sources (e.g. stratigraphic and rock fabric information) to improve balanced cross-sections, to assist in the understanding of *"real physical problems"*.

Based on the work of Dahlstrom (1969) and Elliott (1983), Groshong *et al.* (2012) stated that a 'valid' balanced cross-section should be:

> *"(1) accurate, i.e., it must fit the available data constraints; (2) admissible, i.e., it must conform to structural geometries recognized in local or analogous areas (usually natural, sometimes experimental or theoretical); (3) restorable, i.e., it can be returned to a pre-deformational geometry (single-step or sequential); and finally, (4) the restoration must display "balance" of some definable property, e.g., bed lengths or areas".*

Hence, cross-section 'restoration' is one of the requirements of a 'valid balanced' cross-section. My definition of 'valid' (i.e. 'an interpretation that is possible under the principles of geology') is less restrictive than the definition by Groshong *et al.* (2012). From here on, 'valid' will refer to my definition rather than their formal definition. I used a different definition of 'valid' to better suit the interpretation exercise I set survey respondents (Chapter 3) since they did not have sufficient time to use cross-section balancing techniques.

Restoration is the process of sequentially undeforming a cross-section, returning the beds to their pre-deformational geometries. Bond *et al.* (2012) determined that geoscientists who used the technique of 'geological evolution' (i.e. restoring the seismic image via sketches or writing) were almost 90 times more likely to produce a 'correct' interpretation of a synthetic seismic image (or to identify inconsistencies in their interpretation) than geoscientists who used none of the defined techniques.

Although cross-section balancing and restoration do not analyse the geology in 3D, they are still used frequently in oil and gas industry. Groshong *et al.* (2012) stated that 2D balancing is used to validate concepts and to assist with 3D volume balancing:

> *"In industry, 2D balance is used extensively as described in these papers to validate concepts as well as individual interpretations, and to build workflow templates that are then used in 3D interpretation or to guide 3D balancing".*

Validating an interpretation (via cross-section balancing or restoration) increases the likelihood that invalid interpretations will be identified and rectified or even discarded.

2.1.5. The impact of producing an invalid interpretation

As geological models play a role understanding the sub-surface, the impact of producing an invalid interpretation can be large, e.g. serious financial and environmental penalties. When the geology is unexpected, the impact might be failed exploration wells or unsuitable waste storage sites. In these cases, another well might need to be drilled or another storage site might need to be identified. Opportunities could also be missed, e.g. not discovering a hydrocarbon accumulation. However, using an invalid geological model might not always affect the outcome of decisions. For example, an exploration well could still find economic volumes of hydrocarbons even if the geology is unexpected.

The following example quantifies how often exploration wells do not discover hydrocarbons:

> The 'success' rate[3] of exploration wells quantifies, to some extent, how accurately geological models describe subsurface petroleum systems. Loizou (2003) stated that the UK Atlantic Margin area had a success rate of

[3] The success rate for oil exploration wells is the percentage of wells drilled over a certain time period that can produce hydrocarbons at some pre-determined flow rate (e.g. 1,000 barrels of oil per day).

1 in 7 exploration wells (14.3%) prior to 1995, and that the success rate increased to 1 in 5 wells (20%) since 1995. On the other hand, in the USA, the success rate of exploration wells from 1950 to 1970 was about 20%, but this has increased to about 60% during the period 2004 to 2010 (U.S. Energy Information Administration, 2011, Table 4.6). There are vast numbers of exploration wells drilled that are unsuccessful, each costing approximately $1–20 million[4], but can cost as much as $100 million for deepwater operations. The costs are highly variable and depend on various factors, including whether the well is drilled onshore or offshore, the water depth at the drilling location, the reservoir depth and the subsurface complexity. Production wells usually have higher success rates since the basin is better known due to prior exploration knowledge.

If the uncertainty in geological models is systematically accounted for, and mitigated, it is likely that better decisions could be made.

2.2. Cognition and teamwork in geoscience

The interpretation of geological data is affected by heuristics and cognitive biases because geological data are under-constrained and its interpretation requires expert input, which is subjective. Biases in geoscience interpretation are often a negative phenomenon. Interacting with others (e.g. teamwork) can also influence individuals' judgements. Over the last decade biases have become better acknowledged in geoscience, e.g. Baddeley *et al.* (2004), Bond *et al.* (2008), Hall (2010) and Rowbotham *et al.* (2010). In this section, heuristics are defined and cognitive biases in geoscience are described.

2.2.1. Heuristics in psychology

Heuristics are the 'rules of thumb' used by individuals to simplify judgemental operations (e.g. the estimation of probabilities). Heuristics are applied subconsciously when decisions are needed quickly or when there is incomplete information (e.g. in situations of uncertainty). Different types of heuristics were identified by Tversky and Kahneman (1974):

[4] Numbers are based on informal discussions with geoscientists during 2009 to 2012.

2.2.1.1. Anchoring and adjustment

The ‘anchoring and adjustment’ heuristic is the tendency to make an estimate by starting with an initial value and then adjusting the value up or down to give a final estimate. The starting value effectively anchors the final estimate. Tversky and Kahneman (1974) give an example of this heuristic where two separate student groups were given five seconds to estimate an arithmetic product. One group of students was asked to estimate $1\times 2\times 3\times 4\times 5\times 6\times 7\times 8$; while the other group was asked to estimate $8\times 7\times 6\times 5\times 4\times 3\times 2\times 1$. Interestingly, the first group’s estimates were much lower than the second group’s estimates (a median of 512 compared to a median of 2,250), showing that both groups of students anchored on the first few numbers in their product and then adjusted it to get the final estimate. The correct answer is 40,320.

2.2.1.2. Availability

The availability heuristic is the inclination to judge the frequency of an event by how easily individual examples of it can be recalled (Tversky and Kahneman, 1974). For example, individuals are more likely to overestimate the number of plane crashes over a given period of time if they have recently read (or heard about) plane crashes. The more vivid an event, the easier it is recalled. Slovic *et al.* (1982) also noted that the memorability of nuclear disasters enhances the public’s fear of nuclear power stations, a prime example of this heuristic.

2.2.1.3. Representativeness

The representativeness heuristic is the inclination to assess the probability that ‘A’ belongs to ‘B’ by considering how similar A is to B. For example, Kahneman and Tversky (1973) gave an uninformative description of a professional person to a group of assessors and asked them to estimate the probability that the person belonged to one of two groups. The base rates for the two professions were also given (70% lawyers and 30% engineers). The experiment showed that the majority of assessors failed to take the base rate into account and gave probabilities around 0.5. However, there is a 70% chance that the professional person is a lawyer because 70% of the population were said to be lawyers. Hence, the assessors used the representative heuristic in reaching their probability estimate and ignored the base rate of each profession.

2.2.2. Cognitive biases in geoscience

Cognitive biases arise from the incorrect processing of information and are the subconscious result of using heuristics. Some biases (e.g. 'anchoring') are named after the heuristic. The following biases that affect geoscientists are described: anchoring, confirmation and overconfidence. Hall (2010) described other problem areas that can affect geoscientists in non-interpretive situations, such as big numbers, permutations, probability and randomness. However, these are not considered in this thesis.

2.2.2.1. Anchoring bias

As with the heuristic, described above, anchoring bias is the tendency to make an estimate by starting with an initial value and then adjusting the value up or down to give a final estimate (Tversky and Kahneman, 1974). The 'anchor' can also be beliefs or dominant approaches. In geoscience, anchoring bias is highly relevant since geoscientists may anchor on particular interpretation, analogue or concept, limiting their creativity:

- Bentley and Smith (2008) promoted the use of multiple deterministic models in reservoir engineering instead of anchoring on a 'base case' model. They noted that the spread of results was greater between different conceptual models than it was when adjusting the parameter values only.
- Delfiner (2008) stated that the choice of analogue can act as an anchor when evaluating prospects, and that the suitability of chosen analogues and their domain of validity (i.e. the circumstances where the analogues are valid) should be explicitly stated.

Anchoring bias should therefore be mitigated where possible to increase the likelihood that other concepts are considered.

2.2.2.2. Confirmation bias

Confirmation bias is the inclination to emphasise data that support prior opinions while dismissing contradictory data. One of the earliest descriptions of confirmation bias was given by Bacon (1620):

> *"The human understanding when it has once adopted an opinion (either as being the received opinion or as being agreeable to itself) draws all things else to support and agree with it. And though there be a greater number and*

weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects; in order that by this great and pernicious predetermination the authority of its former conclusions may remain inviolate…"

Similarly, Chamberlin (1890) noted that:

"The mind lingers with pleasure upon the facts that fall happily into the embrace of the theory, and feels a natural coldness toward those that seem refractory".

These quotes describe confirmation bias generally and highlight the fact that it has been acknowledged for centuries. Chamberlin (1890) suggested that considering multiple hypotheses might decrease the effect of confirmation bias since different hypotheses are considered. Confirmation bias was also reviewed by Nickerson (1998) who concluded that individuals more naturally seek data in support of a hypothesis than data that refute it. Bond *et al.* (2008) stated that the interpretation of geological data may be influenced by confirmation bias.

2.2.2.3. Overconfidence bias

Overconfidence is the tendency to be more confident in a judgement than objective testing shows is warranted (e.g. Oskamp, 1965). A typical example is that individuals might state their confidence in knowing the answer to a general knowledge question to be higher than their past success rate, e.g. an individual might be '90% sure' they know the answer even though they are only correct 50% of the time. Harvey (1997) reviewed the literature on overconfidence and concluded that it is likely to stem from how individuals weigh information for and against a given proposition.

Geoscientists have also been found to be overconfident in their interpretations. Rankey and Mitchell (2003) investigated the uncertainty in the predictions of reservoir properties by six geoscientists from seismic attribute analysis. In their experiment, Rankey and Mitchell (2003) stated that:

"Although interpreters commented that the interpretation was relatively straightforward, significant differences in horizon interpretation were present in some areas".

The interpreters were confident in their own judgement even though their interpretations were different. Similarly, Rowbotham *et al.* (2010) noted, in the context of reservoir modelling, that:

> *"We commonly observe biases in the quantification of uncertainty due to cultural influence and overconfidence in our ability to estimate uncertainty, leading to disappointment when reality lies outside predictions".*

Thus, overconfidence bias is not easily identified by the individual during the interpretation, but can be detected at a later stage. Overconfidence can also be 'built into' interpretations via software and attractive imagery, while simpler solutions are sometimes ignored (Chellingsworth *et al.*, 2011).

2.2.3. Group interaction in geoscience

Group interaction (e.g. teamwork) is often a key part of geoscience interpretation, whether in developing the interpretation or in its communication to decision-makers. Herding bias affects group interactions because individuals' judgements can be influenced by others. One structured approach to reducing herding bias is to use an 'expert elicitation' process. Herding bias and expert elicitation are now described.

2.2.3.1. Herding bias

'Herding bias' is a group bias whereby individuals imitate other group members' judgements or behaviours without centralised coordination. One of the first descriptions of herding bias was by Smith (1759) in economics. Herding bias has also been observed in animal behaviour (e.g. Hamilton, 1971) and in some financial institutions (e.g. Jain and Gupta, 1987). Recently, the literature on herding bias was reviewed by Raafat *et al.* (2009), who stated that it would be beneficial to analyse herding bias in many contexts, which are often detached.

Herding bias can reduce the effectiveness of group decision-making when a non-expert uses their personality to act authoritatively. For example, O'Hagan *et al.* (2006) stated that:

> *"One risk is that any individual may have undue influence because of the force of personality rather than actual expertise".*

Polson and Curtis (2010) conducted an expert elicitation process with four geoscientists to investigate how they arrived at a consensus view on decisions within their interpretations. Participants were given multiple 2D seismic reflection profiles from the Firth of Forth (UK) and asked to estimate the likelihood of existence of three subsurface features within the seismic dataset. Using the expert elicitation process, Polson and Curtis (2010) identified herding bias in action, where geoscientists agreed on a consensus that did not truly represent their individual opinions:

> *"Individual judgements can be contradictory, and that group interaction radically alters individual perceptions"*.

Thus, the herding bias might have not been identified if expert elicitation had not been used. In relation to Polson and Curtis (2010), Curtis (2012) stated that:

> *"Consensus positions clearly only represent the group opinion at one instant in time, and may not represent the true range of uncertainty about the issue".*

These quotes illustrate that personality plays a role in group interactions and that consensus views may not represent all individuals' true opinions. Ideally, interpretation teams should contain geoscientists with a range of prior knowledge (Bond *et al.*, 2008). However, the benefit of having a greater range of experience in a team would be lost if individuals do not make judgements independently.

2.2.3.2. Expert elicitation

To minimise the effect of herding bias, an expert elicitation process can be used (Baddeley *et al.*, 2004). Expert elicitation is a structured and iterative process, led by a facilitator, to capture and amalgamate knowledge from a group, usually experts. Ye *et al.* (2008) suggested that expert elicitation should be carefully designed to fulfil its specific objectives, and different types of experts (*"generalists, specialists and normative"*) should be present in a team to complement one another. Generalists are defined to understand the wider research aims and how they relate to each other; specialists understand the details of an area of interest, while normative experts assist generalists and specialists in expressing and combining their knowledge in a useful way (Ye *et al.*, 2008).

The application of expert elicitation in geoscience includes:

- Rankey and Mitchell (2003), who quantified the impact that seismic interpretation can have on the prediction of reservoir properties from seismic data.
- Arnell *et al.* (2005), who assessed the likelihood of ‘rapid’ climate change.
- Ye *et al.* (2008), who evaluated five recharge models for a regional groundwater flow system.
- Aspinall (2010), who assessed the time-to-failure of dams.
- Polson and Curtis (2010), who assessed the likelihood of existence of a reservoir, seal and fault.

As an example, details of Aspinall (2010) are now given. Aspinall (2010) used the ‘Cooke method’ of expert elicitation (Cooke, 1991) to amalgamate the time-to-failure for dams from eleven experts. Experts’ judgements were ‘weighed’ by how accurately a set of ‘seed’ questions with known answers were answered, ensuring that ‘more knowledgeable’ experts are given more authority. Aspinall (2010) found that the Cooke method produced time-to-failure values that were twice as long as the ‘equal-weighted’ solution with greater uncertainty ranges, perhaps reducing overconfidence. However, it was noted that the choice of experts and the choice of seed questions could affect the outcome of the elicitation process.

Curtis and Wood (2004) presented an elicitation process that optimised the elicitation of probabilistic information from experts in real time. In their process, relative likelihoods of different probabilities of interest are captured (e.g. ‘model A is more likely than model B’), instead of asking experts to come up with an absolute value (e.g. an X% likelihood that this model is realistic), which can be problematic (e.g. Hall, 2010). Curtis and Wood (2004) stated that their process maximises ‘expected information’ using experimental design theory and mitigates overconfidence.

This section outlined some of the most common biases (anchoring, confirmation, overconfidence and herding) that can affect geoscientists and geoscience teams when making an interpretation. Even if an elicitation process is not used to mediate group interactions, Baddeley *et al.* (2004) advised that geoscientists should still be aware of common heuristics and biases when modelling the subsurface.

2.3. Uncertainty in geoscience: an overview

Uncertainty is a concept that describes 'how much is unknown' about an object of interest, system or nature itself. Relating to the oil and gas industry, Capen (1976) noted that many people find uncertainty difficult to assess, possibly due to limited exposure. However, in the last twenty years, the quantitative analysis of uncertainty has had greater coverage. In this section, I describe the common sources of uncertainty in geoscience and then compare two types of uncertainty analysis.

2.3.1. Sources of uncertainty in geoscience

Mann (1993) and Bárdossy and Fodor (2001) provided descriptions of the sources of uncertainty in geoscience:

1. The inherent variability in nature. For example, a measurable property of a physical object (e.g. the porosity of sandstone) will be different at different spatial locations.

2. Sampling limitations. Gathering a truly representative sample of a geological phenomenon is unachievable as the majority of the rock volume is inaccessible, e.g. underground. Geoscientists' direct observations are mostly on rocks near the surface.

3. Observation error. Since environmental conditions vary at outcrops, e.g. the coverage of vegetation and differences in the weather affect what can be seen. Some localities will thus be more challenging to sample than others. Also, perhaps due to the confirmation bias, Chadwick (1975, 1976) observed that geologists incorrectly recalled details of geological features that they had observed in the field. Antiforms were remembered more easily than synforms and cleavage fans were remembered as presented in text books rather than as they are in reality.

4. Measurement error, which is caused by human error when taking measurements. Measurement error also includes the incorrect calibration of equipment and incorrect rock sample preparation.

5. Errors from the incorrect mathematical evaluation of data, which includes collecting too few measurements (e.g. rock samples).

6. Propagation of errors. Individual measurement errors in multivariate models will propagate through the model based on the relationships between the variables.

7. Conceptual and model uncertainty. Pre-existing geological concepts are necessarily applied to enable the interpretation of geological data. However, since geological data are under-constrained, the choice of concepts to apply in an interpretation is a source of uncertainty. If inappropriate concepts are used, all further analyses could become erroneous.

Uncertainty therefore originates in the data, the method of analysis and the cognitive biases that affect the interpretational process. One method to reduce uncertainty in geoscience interpretation is through further data collection and/or better technology, but this can be costly.

2.3.2. Parametric uncertainty

One of the most commonly analysed types of uncertainty in the oil and gas industry is ‘parametric uncertainty’:

> Parametric uncertainty is the uncertainty in the ‘true’ value of parameters from an equation of interest, e.g. the average porosity, water saturation or permeability of a reservoir. These values exist in reality but are unknown. Therefore, each parameter needs to be estimated to allow the equation to be computed. Parameter estimates can be in the form of statistical distributions to represent the range of possible values and their respective likelihoods of occurrence. Common distributions used in geoscience are: normal, lognormal, triangular and uniform. For example, a ‘normal’ distribution implies that parameter values close to the middle of the distribution are much more likely to occur than parameter values at the tails of the distribution; whereas, a ‘uniform’ distribution implies that all possible parameter values are equally likely to occur. The choice of distribution used for a particular parameter is based on geological knowledge.

Parametric uncertainty can be probabilistically quantified via computer simulations (e.g. Monte Carlo simulations) where the equation of interest is computed many thousands of times using different values for the parameters. Three studies from reservoir modelling are referenced as examples of the analysis of parametric uncertainty:

- Egermann and Lenormand (2004) quantified the uncertainty in the relative permeability and capillary pressures.
- La Pointe and Fox (2011) presented different approaches for characterising uncertainty in fractured reservoir models, including an analysis of parametric uncertainty and an analysis of alternative conceptual models.
- Liu *et al.* (2011) studied the parametric uncertainty in petrophysical parameters of a Gulf of Mexico reservoir as part of a larger analysis, which included multiple conceptual models and an assessment of economic viability.

One criticism of the analysis of parametric uncertainty is that it only explores part of the uncertainty space (Bond *et al.*, 2008; Rowbotham *et al.*, 2010), since only one conceptual model is usually considered. Any geological model developed using an inappropriate concept may not only render further analysis wasted effort, but could have negative financial implications (Bond *et al.*, 2008). It is therefore beneficial to also consider multiple conceptual models of a dataset (discussed below). For example, in groundwater modelling, Refsgaard *et al.* (2006) explained that 'traditional' uncertainty analyses focus on parametric uncertainty and the input data, but that the principal source of uncertainty was often the conceptual model. Similarly, Ye *et al.* (2008) stated that:

> *"Conceptual model uncertainty can be significant, ignoring it (focusing only on parametric uncertainty) may result in biased predictions and underestimation of uncertainty".*

Conceptual uncertainty can often be more influential than parametric uncertainty. However, if a valid concept is used, parametric uncertainty can still obscure the data (Woodward, 2012). Hence, the analysis of both types of uncertainty is useful.

In this section, as well as sources of uncertainty in geoscience, parametric and conceptual uncertainties were discussed. Parametric uncertainty is commonly analysed in subsurface modelling, but conceptual uncertainty may play a larger role than many geoscientists realise. The next section contains a review of conceptual uncertainty.

2.4. Conceptual uncertainty in geoscience

Geological concepts, combined with observations, can be used to identify structural features (e.g. faults, horizons and anticlines) and stratigraphic features (e.g. reflector terminations, carbonate reefs and river deltas) in an interpretation. A geological 'concept' is a theory that explains one or more relationships within the data. However, since the data are under-constrained, more than one concept may be suitable. For example, it is possible to confuse carbonate reefs with anticlines based on seismic data alone. The uncertainty in the geological concept applied in an interpretation (which is an underlying assumption) is termed *"conceptual uncertainty"* (Bond *et al.*, 2007). A reflector discontinuity in a seismic dataset could be explained by the concept of 'faulting' or by the concept of 'superposition', e.g. an unconformity. Both concepts may 'honour the data' (i.e. do not contradict it) but imply different geological processes. Recently, conceptual uncertainty was illustrated by Resor and Pollard (2012) who considered different explanations (i.e. concepts) for rollover faulting in the hanging-wall of a normal fault.

Geoscientists, even with many years of experience, have been shown to interpret the same data differently, which is a form of conceptual uncertainty since the process of interpretation is reliant on geological concepts. This section presents examples of uncertainty in geoscience interpretation and hence, shows that conceptual uncertainty is fundamental to investigate within geoscience interpretation. For example, if invalid concepts are applied during an interpretation all further analyses could become erroneous.

In the following sub-sections, I review the literature on conceptual uncertainty in two tranches, defined by the scale of the study. I define a 'large-scale' study to be one with 100 geoscientists or more. Large sample sizes allow factors to be quantified statistically, whereas small-scale studies grant the investigators more time to analyse participants' behaviours. However, it can always be argued that the findings of a small-scale could be the result of chance via the selection of participants. Would the results still exist if different participants had been selected? On the other hand, if the aim is not to derive results that are statistically significant with regard to the underlying population, but to confirm that some phenomenon is present, then a small dataset may be sufficient. A practical advantage of small-scale studies is that they can be conducted in time-pressured situations, while studies with large sample sizes take longer.

2.4.1. Small-sample studies

Rankey and Mitchell (2003) found that six interpreters' predictions of seismic attributes varied considerably in 'areas of uncertainty' and advocated evaluating the impact of end-member models. Likewise, Rowbotham *et al.* (2010) noted that understanding the effect of uncertainty on commercial decision-making is a key importance.

Uncertainty was present in Polson and Curtis (2010)'s expert elicitation process where four participants did not agree on the probability of existence of defined subsurface features. Even though a consensus view was reached, not all participants agreed to it when asked individually. There was also considerable uncertainty in the experts' initial likelihood estimates, and one expert investigated a different geological feature than what was asked. Hence, the authors advised that *"misunderstandings and incorrect assumptions"* may go undetected if geological interpretations are not quality controlled.

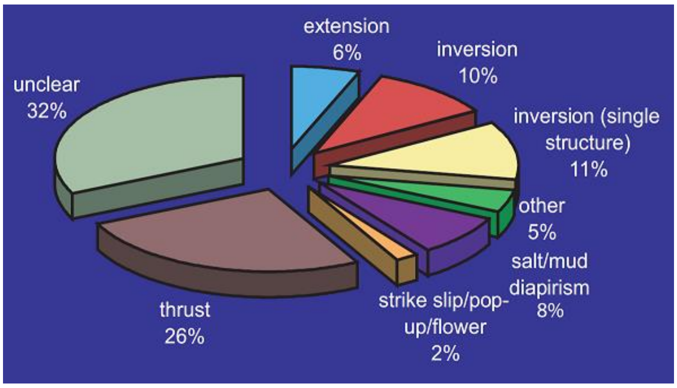
Torvela and Bond (2011) compared 24 participants' seismic interpretations to existing theoretical models for deepwater fold-thrust belts and found that the majority of interpretations were consistent with current theoretical models. However, the interpretations that were less consistent were a better representation of reality. Therefore, use of the current theoretical models over-simplified the interpretation. The authors also found substantial positional uncertainty in the interpreted faults and horizons.

2.4.2. Large-sample studies

Bond *et al.* (2007) surveyed 412 geoscientists from academia and industry in the first large-scale study to directly assess uncertainty in geoscience interpretation. Participants were asked to complete a questionnaire about their education and professional experience, and were then required to interpret a synthetic seismic image. The aim of the study was to capture the range of interpretations of a single dataset to illustrate conceptual uncertainty.

The synthetic seismic image was based upon a known geological model (tectonic inversion of normal faults by later thrusting) and forward-modelled. Hence, a ‘correct’ answer existed in this case, unlike the interpretation of real seismic data. Bond *et al.* (2007) observed that only 21% of participants interpreted the ‘correct’ tectonic setting (Figure 2.2), while the most common interpretation was thrusting (26%). Due to the strict classification scheme, 32% of interpretations had to be classified as ‘unclear’.

Figure 2.2 – From Bond *et al.* (2007). The pie chart shows the range of tectonic styles in 412 interpretations of a synthetic seismic image. Only 21% of the respondents (11% + 10% = 21%) identified the original tectonic setting (inversion).



In Bond *et al.* (2012), the authors considered only the ‘expert’ participants of the original dataset, which totalled 184 geoscientists. Experts were defined to be those participants who ranked themselves as being ‘proficient in structural geology’ or ‘a structural geologist’ within the questionnaire. Using a multivariate statistical analysis, Bond *et al.* (2012) showed that ‘having a Masters or Ph.D. degree’ significantly improved the chances of identifying the correct tectonic setting, regardless of the interpretational techniques used.

The most effective technique was found to be ‘evolutionary thought’, which only 18 of the 184 experts applied. Evolutionary thought was defined to be:

> *“Sketches or writing that show the evolution of the geological architecture (structure and/or sedimentation through time)”*.

The other (less influential) interpretational techniques were: ‘feature identification’, ‘horizon interpretation’, ‘sticks’ and ‘annotation’. When comparing the geoscientists who applied ‘evolutionary thought’ to the geoscientists who used no interpretational techniques, Bond *et al.* (2012) found that using the technique of evolutionary thought made geoscientists almost 90 times more likely[5] to attain the correct answer. There was also evidence that two geoscientists who applied evolutionary thought and did not identify the correct answer knew that they were wrong.

Hence, the technique of ‘evolutionary thought’ helped geoscientists attain the correct answer and allowed two other geoscientists to identify their mistakes during the interpretation. In concluding, Bond *et al.* (2012) advised geoscientists to consider the geological evolution of seismic images as a matter of routine, and noted that it should also be the foundation of a fuller model validation such as cross-section balancing.

2.4.3. What further research is needed?

Rankey and Mitchell (2003) and Polson and Curtis (2010) described experiments where small groups of interpreters interpreted the same geological dataset differently. Although more information was known about the participants, there were too few to achieve statistical significance.

Bond *et al.* (2007) demonstrated that the range of interpretations of a single dataset can be large; while, Bond *et al.* (2012) identified which geoscientists were most effective in terms of their education, professional experience and the interpretational techniques that were used. However, the conclusions that can be drawn from these studies are limited because the questionnaire captured limited amounts of detail about the respondents (nine questions about education and professional experience). For example, the questionnaire did not differentiate between Master’s and Ph.D. degrees.

Further research should thus, employ a large-sample survey to quantify which aspects of geoscientists’ backgrounds, and what the techniques used, are

[5] 89.8 was the odds ratio from the statistical analysis, comparing geoscientists who had used evolutionary thought against the geoscientists who had used none of the techniques.

associated to producing a valid interpretation. Using this knowledge, geoscience teaching could be improved and mitigation workflows can be developed.

2.5. Multiple interpretations of data

The risk of applying a wrong concept in a geological interpretation can be mitigated by considering multiple hypotheses (multiple interpretations of data) since alternatives are considered. Chamberlin (1890) advocated considering ‘multiple hypotheses’ during scientific investigations to avoid confirmation bias, e.g. to avoid having *“parental affection for a favourite theory”*. Multiple interpretations can also reduce anchoring bias (Bentley and Smith, 2008). As above, geological data are under-constrained and uncertainty is therefore ubiquitous.

Bond *et al.* (2008) contended that choosing one interpretation early in the workflow should be avoided as it increases the associated risk since other alternative interpretations are ignored. By considering multiple models through the workflow, Bond *et al.* (2008) explained that information is maximised and uncertainty is minimised. The main benefit of ‘considering multiple interpretations’ through the workflow is that the alternatives can be re-assessed without having to restart the interpretation process. Depending on the context, restarting the interpretation process could take weeks or months, e.g. if a full reservoir simulation is required. Being able to re-assess different interpretations quickly is important in time-pressured situations.

Bond *et al.* (2008) suggested that an interpretation team with a broad range of prior knowledge will increase the likelihood that alternative interpretations are considered. Also, since the ‘best’ interpretation is often unclear, attempting multiple interpretations might give a better understanding of the data.

In addition, the following research promotes the use of multiple models in geoscience interpretation:

- Bentley and Smith (2008) promoted the use of multiple deterministic models in reservoir engineering instead of anchoring on a best-guess model to reduce anchoring bias.

- Rowbotham *et al.* (2010) encouraged the identification, communication and use of multiple deterministic models as part of the standard seismic interpretation workflow.
- Chellingsworth *et al.* (2011) advocated the use of alternative models to avoid anchoring on an initial model and adjusting model parameters.

Hence, considering multiple conceptual models may yield benefits in interpretational efficiency and mitigate the anchoring and confirmation biases.

2.6. Conclusions

Effective subsurface modelling is essential to the oil and gas industry, but there are substantial barriers to overcome. The barriers include that geological data are under-constrained and that interpreters are influenced naturally by cognitive biases when dealing with these data. Data quality is driven by technology and acquisition costs, while cognitive biases affect geoscientists who use their experience and prior knowledge during an interpretation. Individual and group biases are both commonplace in geoscience interpretation even though they are not often acknowledged while making an interpretation.

One of the most common types of uncertainty analysis is the assessment of parametric uncertainty. I contend, however, that conceptual uncertainty can be more important since it is an underlying assumption that constrains all further analyses. Conceptual uncertainty has been acknowledged by recent literature as a fundamental uncertainty that should be considered. One current method to mitigate the risk arising from conceptual uncertainty is to assess multiple interpretations of data, which also mitigates the confirmation and anchoring biases.

This thesis documents how geoscience interpreters can become more effective at a time when hydrocarbon exploration is becoming more complicated. For example, fourway structural closure traps are being analysed less often by explorationists, and more complex reservoirs are being targeted (e.g. high pressure, high temperature structures). Due to technological advancements, geological datasets are more detailed than ever due to their increasing size and the number of parameters available. Therefore, training geoscientists to be more effective, and managing interpretations effectively, is of the utmost importance to meet the challenges of the twenty-first century.

Chapter 3 – Methodology

In this chapter, I explain why a survey was the best option to address the research questions of interest. I discuss relevant sampling theory and questionnaire design issues. The questions are listed and the reasoning for them explained. I then discuss the interpretation exercise that was chosen to be in the questionnaire, which was a key component of the survey, along with the responses to the questionnaire questions. Finally, I summarise the data collection and quality control measures that were implemented to ensure a high-quality sample.

3.1. Survey introduction

To investigate the factors that affect the interpretation of geological data, I created a questionnaire and collected responses from a large sample of geoscientists during 2009 to 2011. The questionnaire captured information on respondents' backgrounds and contained a seismic interpretation exercise. I named the questionnaire the *"Freyja questionnaire"* to give it a brand, which I hoped geoscientists would remember and even place confidence in. The name 'Freyja' comes originally from Norse mythology. Freyja was chosen to imitate the informal names given to previous research projects by Dr Bond and Professor Shipton: 'Odin' (Bond *et al.* 2007; 2012), 'Thor' (Bond *et al.*, 2011) and 'Vör' (unpublished). In this thesis, I will refer to the questionnaire as the 'background questionnaire'.

3.2. Questionnaire sampling strategy

This section presents relevant survey theory and discusses the types of questionnaire that could be implemented.

3.2.1. Survey theory

Surveys are undertaken to collect opinion, appraise ideas or to gain knowledge as they allow investigators to gather data on the population of interest in an efficient manner (Levy and Lemeshow, 1999). Hence, the resulting samples are, by definition, a subset of the underlying population, which cannot usually be directly measured due to its size and scattered distribution. I took the following principles into account when planning the survey:

- The size of the sample needed to be as large as possible to ensure accurate estimates of the population parameters (the object/questions being investigated).
- Geoscientists needed to complete the interpretation exercise independently from one another.
- The sample of geoscientists needed to represent the underlying geoscientist population.
- The occurrence of missing data needed to be kept as low as possible.

3.2.1.1. Sample size

While small-scale studies are detailed, the numbers of participants traditionally used are not large enough to be able to capture the impact of individual variables. The main research aim was to statistically quantify the effect of individual factors in geoscience interpretation. Thus, a large sample was needed (hundreds of geoscientists).

If the sample is too small then any results could have been due to chance. For instance, in a small sample, individuals with a ‘stronger than normal’ opinion could have been sampled, distorting the view of the whole population. A sample therefore has to be large enough for the results to be representative of the population, whilst being affordable in terms of resources.

Statistical theory dictates how large a sample needs to be to guarantee a desired ‘precision’ of the population attributes of interest (Cochran, 1977). A parameter is defined to be the population attribute that is being investigated, e.g. the sample is collected to gain information about aspects of the underlying population. The ‘precision’ measures how well an estimate of the parameter $\hat{\theta}$ matches the population parameter θ, which is unknown. Hence, if $|\theta - \hat{\theta}| \approx 0$ then the estimator $\hat{\theta}$ approximates θ well and is said to be a ‘precise’ estimate. $\hat{\theta}$ is a random variable which varies from sample to sample. Hence, if multiple samples were collected, different values for $\hat{\theta}$ would be found; which is an example of sampling error.

Prior to data collection, an 'error bound' B can be specified so that $\hat{\theta}$ lies within $\theta \pm B$ with high probability, e.g. $P(\theta - B < \hat{\theta} < \theta + B) = 0.95$. The sample size is therefore chosen large enough so that, on average, the probability of $\hat{\theta}$ being within the error bound is 95%. Cochran (1977, p.73) presents the formula that connects the error bound with the sample size (Equation 1):

Equation 1 – From Cochran (1977, p.73). The equation connecting the 95% confidence error bound (B), sample size (n) and proportion of interest (p), for a large unknown population.

$$n = \frac{4p(1-p)}{B^2}.$$

To guarantee an error bound of 5%, using Equation 1, I show that the desired sample size for this research was 400 respondents:

$$n = \frac{4p(1-p)}{B^2},$$

$$n = \frac{4 \times 0.5 \times (1 - 0.5)}{0.05^2} = \frac{1}{0.0025} = 400.$$

The proportion of 0.5 was used since it maximises the numerator. Hence, if the proportion of interest (p) is any other value, the error bound is guaranteed to be smaller than 5% for a sample of 400 respondents. Equation 1 can also be rearranged to determine an error bound, given the sample size.

3.2.1.2. Independence

The statistical analysis assumes that all respondents completed the interpretation exercise independently. With this in mind, I wrote on the questionnaire that *"this exercise must be completed independently"*. In addition, all respondents were given this instruction verbally before starting the exercise. Where the questionnaire was distributed in seminars, I asked geoscientists to space themselves out. However, due to space constrains in some locations, respondents were working in close proximity to their colleagues and it may have been possible for them to see each other's interpretations. Even if some geoscientists did copy, or were influenced by, their neighbours' interpretations, the risk of this negatively affecting the results is lowered since 40% of the respondents (283 of 703) were filtered out due to having inappropriate experiential backgrounds or being too inexperienced (see section 3.7.1).

3.2.1.3. Missing data

Respondents with missing data were excluded from the analysis if the question which they missed out was being analysed. To emphasise that respondents should complete all questions, I added *"Please complete all questions"* to the front of the questionnaire, and respondents were also verbally told to 'complete all questions' before starting the questionnaire. The amount of missing data is noted where appropriate. I also considered using the statistical technique of missing data imputation (e.g. Little and Rubin, 2002) to fill-in values for missing data, based on respondents' answers to other questions. For example, imputation may have been useful for Q21 where 75 respondents failed to record the locations where they had investigated the geology. It may have been possible to use responses to other questions to estimate how many geographical locations the respondent had investigated the geology at. However, I decided against using imputation due to time constraints.

3.2.1.4. Self-selection bias

Self-selection bias can occur when individuals who agree to take part in the survey have a systematically different opinion to those individuals who do not participate, meaning that the sample becomes non-random. Self-selection bias is common (Heckman, 1979). In my experience, geoscientists who chose not to participate probably did not see it as a valuable use of their time, e.g. one geoscientist said *"you cannot afford my time"*, implying that they were too busy or did not see the survey as being useful. However, respondents who may not have participated had the questionnaire been handed out at a conference took part in the seminar format. In general, it is easier for potential respondents to decline a questionnaire than to leave a seminar. The number of respondents who received the questionnaire at conferences is compared to the number of respondents who received the questionnaire in seminars is discussed in section 3.6.2. I also targeted different demographics of the population whilst at conferences, e.g. young and old geoscientists. However, since the sample was filtered, the effect of self-selection bias, if there is any, should decrease.

3.2.2. Type of questionnaire

One main survey design decision was whether the questionnaire would be an electronic-based or paper-based exercise. The electronic-based questionnaire would need to be placed on a website, would be interactive and the interpretation exercise would have to be completed digitally. Alternatively, a paper-based questionnaire would require getting the physical questionnaire to the geoscientists. Each format has its own advantages, which I have outlined below.

There are three main advantages to using an electronic questionnaire:

- I would be able to achieve a far greater sample size allowing the statistical analysis to produce more accurate estimates of population parameters of interest.
- I would not need to manually input the questionnaire data, i.e. respondents' background information would be filled-in digitally and input to a spreadsheet.
- There would be no possibility of transcription errors.

However, I decided to use a paper-based questionnaire for three principal reasons:

- A paper-based questionnaire would allow me greater control over its distribution, giving me greater confidence in the quality of the sample. For example, if an electronic questionnaire had been used, it would have been possible for the questionnaire to be sent between colleagues.
- I (or someone acting on my behalf) was there to ensure that the data collection process was completed properly, e.g. that the questionnaires were completed independently.
- Inputting the questionnaires manually would give me a far better understanding of the geology in the data set than if the interpretation was completed digitally.

I decided that having a sample of the highest quality was the most important factor and thus wanted complete control over the questionnaire distribution. Although I would have gained a larger sample using an electronic questionnaire, the sample I collected was still more than large enough to attain statistical significance. In addition, although there was the chance of transcription errors, I wanted to gain as strong an understanding as possible of the geology in respondents' interpretations. Hence, the task of inputting the questionnaire and interpretation data, which took over three months, was time well spent. For information on how I mitigated the chance of transcription errors during data input, see section 6.2.

3.3. Design of the questionnaire

In this section, I discuss the factors I took into account when designing the background questionnaire, how I tested the questions before the data collection, and how I gained ethics approval for the survey. I primarily used Willis (2005) and Bradburn *et al.* (2004) as a guide to design the questions and response categories in the questionnaire.

3.3.1. Length vs. detail

A balance needed to be struck between how much detail about respondents was captured and the length of the questionnaire. Longer questionnaires can be more detailed. However, to maximise the sample size, I wanted to keep the questionnaire short and still collect a reasonable amount of detail about respondents' education, work environment and experience. I decided to limit the background questionnaire to 21 questions in addition to the interpretation exercise. The entire exercise was expected to take between 15 and 25 minutes.

3.3.2. Clarity of questions

The questions had to be as clear as possible to enable respondents to complete the questionnaire quickly and give meaningful answers. It is important for questions to be understood easily (Jenkins and Slack, 1985). For example, the wording of a question can affect the response given, and even small changes in the wording or grammar can impact answers to a question (Bradburn *et al.*, 2004). The questions should not 'lead' respondents by using words with connotations which could affect how respondents might answer. The text size and type font should be large and clear enough for respondents to read quickly without having to strain their eyes (Bradburn *et al.*, 2004). Some respondents may be put off finishing the questionnaire if the questions within it are long and complex. In addition to having effective questions, the response categories also needed to be appropriate for the question so that all respondents fit into at least one answer category.

3.3.3. Testing the questions

Oppenheim (1992) stated that all questionnaires should be tested to determine whether the questions will be understood as intended. Hence, there were eight distinct versions of the background questionnaire over a six-month development period, with the questions being tested on at least four separate occasions by Ph.D. students, academics and industry professionals[6]. Testers were given the questionnaire with no further information and asked what they thought the questions were meant to capture and whether they could identify any confusing questions. The testers gave invaluable feedback regarding the style and phrasing of questions, as well as the aesthetic look of the questionnaire. Since not all respondents' first languages would be English, the questions were checked by non-native English speaking geoscientists to ensure clarity.

One example of ambiguity in the questions is as follows. In the answer categories for Question Qiii, it was found that the meaning of the phrase 'quite confident' with respect to 'confident' was different in the USA and the UK. In the UK, 'confident' would be rated as being more confident than 'quite confident', but in the USA 'quite' has the opposite meaning. In the USA, 'quite confident' is taken be more confident than 'confident'. The phrase 'quite confident' was therefore replaced by 'satisfied'.

Early in the data collection phase (after the first 3 months), I changed the look of the three questions underneath the seismic image to emphasise that the questions were to be completed after the interpretation exercise and were not part of the background questionnaire. I did this by changing the outline of the box from a thin black line to a thicker red outline. I also made the following instruction larger and in bold:

> *"Complete the three questions below after completing the seismic interpretation exercise"*.

[6] The students and academics mainly came from the Universities of Strathclyde and Glasgow. Apart from a few hand-picked individuals, the professionals came from Midland Valley Exploration Ltd.

It is therefore likely that these changes made the questions clearer for respondents, but this could not be tested.

3.3.4. Ethics approval

Guidelines for best practice in research mandate that all research involving human subjects must be approved by an independent ethics committee. The importance of ethics is also noted by Bradburn *et al.* (2004), who note the following ethical principles for surveys: respondents should have the right to privacy, respondents should give their informed consent to take part, and respondents' questionnaire responses should be kept confidential. Accordingly, I submitted an ethics form (Appendix 2) and the questionnaire to the University of Glasgow ethics committee in July 2009. It was approved, and also by the University of Strathclyde ethics committee in August 2011[7].

The departmental ethics committee at the University of Glasgow noted the following issues:

- Student respondents should not get penalised in their degree work (consciously or sub-consciously) by a lecturer seeing their interpreted seismic image.
- Professional respondents should not get penalised by their employer, i.e. by their boss seeing their interpreted seismic image.
- All participants should be told why I am asking for their help, who is being invited to take part, and the survey's background information.
- Participation should be voluntary and the geoscientist's own choice.

I mitigated the first two bullets by ensuring that respondents' anonymity was protected. Anonymity ensured that even if a lecturer or employer saw a completed questionnaire, they would not know who the individual was. I stressed to all individuals that participation in the survey was anonymous and voluntary. No compensation was given to respondents. Respondents also did not need to sign a consent form as this would have prevented anonymity; their permission was assumed since they had chosen to take part.

[7] I started my Ph.D. at the University of Glasgow but moved to the University of Strathclyde in June 2010 due to supervisor relocation.

To provide the necessary information, I produced an information sheet for respondents. I also added an outline of the research and my contact details to the front of the questionnaire. For student groups, the lecturer was not allowed to observe students while they completed the interpretation exercise. I wanted to avoid the situation of a lecturer seeing certain students as ‘being more keen’ than others, based on how they filled-in the questionnaire. In the few cases where a lecturer was present, they took part in the survey as well.

3.4. List of questions

The background questionnaire contained 21 questions about respondents’ education and experience, and three questions about the interpretation exercise itself. This section lists the questions that were included in the background questionnaire. I also explained why each question was chosen. Please refer to the background questionnaire in the Appendix to view the answer categories and to see how the questions were displayed.

Questions 1 and 2

1) Gender?

2) Age?

Questions 1 and 2 were an effective opening to the questionnaire because respondents can answer them almost without thought. Respondents’ genders and ages were directly available so they did not have to calculate anything, the information was ready. Respondents might be less likely to complete the questionnaire had a more complex question been asked first (Bradburn *et al.*, 2004).

Kali and Orion (1996) documented that teenage males outperformed teenage females in spatial thinking and visualisation, which are needed in geological reasoning. However, it was also shown that these differences decreased with experience and training. Since the target population for the survey was professional geoscientists, gender is likely to have little influence on respondents’ interpretational ability, the measure of interest for this research. After analysing the main dataset, I observed that ‘gender’ was non-significant in the statistical analysis (Appendix 5), which implied that the gender of geoscientists did not affect their interpretational ability.

Since experience increases with age, respondents' ages may be positively associated to their interpretational abilities. Respondents were asked to select their age from bins of ten years to ensure that there are reasonable numbers in each category. However, I determined that 'age' was not significantly associated to interpretational ability (Appendix 5).

Questions 3 and 4

3) What degrees have you completed?

4) What subject / topic area(s) were they in?

Question 3 captured the University degrees that respondents had attained, and Question 4 captured their subject area(s). Knowing what degrees were completed was useful because education is likely to affect interpretational ability. Some respondents gave their full course title, while others gave the topic area. After the data collection, the subject / topic area(s) from Q4 were manually grouped into similar subject / topic areas. However, the categorised versions were not used since the variables were non-significant in the statistical analysis (Appendix 5).

Question 5

5) Number of years of relevant experience (those relating to geoscience) since attaining your highest degree?

Question 5 captured the 'years of experience' since the completion of respondents' highest University degree. Relevant experience was defined to be 'those years relating to geoscience' to avoid the inclusion of non-relevant experience. The number of years of experience was important to capture as it was a direct measure of respondents' experiences.

Question 6 and 7

6) Which of the below describes where you have worked in the past 24 months?

7) Which of the below describes your background most accurately?

Question 6 captured the work environment from the last 2 years, while Question 7 captured their previous work environment (i.e. their entire geoscience career, excluding the two most recent years). Willis (2005) noted that the phrase ‘in the past 2 years’ can mean either 2 calendar years, where respondents’ employment areas would be counted from the 1st of January, or 2 years to that point in time’. Therefore, if I had used the phrase ‘in the past 2 years’, some respondents may have only counted events since the 1st of January and ignored relevant events before the 1st of January. I used ‘in the past 24 months’ to make the question clearer.

Questions 8 and 9

8) What best describes your experience in structural geology?

9) What best describes your experience in seismic interpretation?

Questions 8 and 9 captured respondents’ experiences in structural geology and seismic interpretation. Respondents’ experience levels had to be self-assessed to preserve anonymity. The response categories for Question 8 and 9 were originally: ‘expert’, ‘intermediate’ and ‘beginner’. However, the wording of these could have encouraged respondents to overestimate their abilities, e.g. some respondents may have been drawn to select the ‘expert’ category without sufficient expertise. I therefore changed the response categories to ‘specialist’, ‘good working knowledge’ and ‘basic knowledge’, which are less judgemental. I also added the category of ‘no knowledge’ to be able to filter out the respondents who had inadequate experience in structural geology or in seismic interpretation.

Mid-data collection, I learned that rating scales should ideally have an odd number of categories in them, to allow respondents to ‘move away’ from the middle option (Bradburn *et al.*, 2004). Since there were only four options, respondents who rated themselves in the ‘middle’ either had to place themselves in the ‘good working knowledge’ or the ‘basic knowledge’ category, which may have been difficult. However, as the data collection was well underway, I decided not add another response category. If I had originally included a fifth response category, I would have yielded a better distinction between respondents who had ‘middle’ levels of experience. Nevertheless, I still found that respondents were divided effectively when comparing the categories ‘specialist’ and ‘basic knowledge’.

Question 10

10) How often do you interpret or use seismic images?

Question 10 asked how often respondents interpreted or used seismic images. The phrase ‘interpret or use seismic images’ was chosen to make the question applicable to other geoscientists who work with seismic images often but do not actively interpret them. An unnoticed flaw in this question was that it would not distinguish between an accomplished seismic interpreter who now spends their time in management and a geoscientist who just does not work with seismic images. In this example, each individual might select one of the ‘least frequent’ options even though their seismic interpretation experience levels would be very different. I expected increased frequency of seismic image interpretation to be a positive influence on respondents’ interpretational ability.

Questions 11, 12 and 13

11) Rank the following areas of geoscience to show which areas you have been most active in over the last 24 months.

12) Rank the following geological settings by duration to show where you have worked in the past 24 months.

13) Rank the following geological settings by duration to show where you have worked in your entire geoscience career.

Questions 11, 12 and 13 captured the geoscience areas and tectonic regimes that respondents had been most active in. Questions 11 focussed on the areas of geoscience, while Questions 12 and 13 focussed on the tectonic regimes respondents had worked at ‘in the last 24 months’, and in respondents’ entire geoscience careers. Respondents were asked to rank multiple response categories, in the order of where respondents had been ‘most active’:

“In the following questions (Q11–13) please use rankings to indicate your answers.

Please note; you do not need to rank areas/geological settings in which you have never worked – only rank options in which you actually have some experience.

Equal ranks are allowed.

In all questions, 1 = most active / worked there most, and lower rankings (2, 3, 4, etc…) = less active / worked there less".

While ranking questions take slightly longer to complete, they often yield information that is more detailed. Respondents were likely to have been active in multiple geoscience areas and tectonic regimes, so rankings were an effective question type to have used. Questions 11, 12 and 13 did yield useful information but three main problems were found, e.g. some respondents:

- Ticked the geoscience areas (or tectonic regimes) where they had been active without actually ranking them.
- Used rankings which did not start at 1. For example, certain respondents started with much lower rankings (e.g. 4 or 5), perhaps to indicate that they worked in that area infrequently and in no other areas. Although, I provided space in which respondents could provide alternative geoscience areas or tectonic regimes.
- Assigned percentages to each box, which gave a sense of the magnitude between their selected areas or tectonic regimes.

In the cases where these questions were misunderstood, I used the available information to rank the categories myself. For instance, in Question 11, if a respondent had only given rankings of, say, 4 and 5, to 'basin modelling' and 'geochemistry' respectively, I would have changed those rankings to 1 and 2, respectively. In most cases, it was clear which categories the respondent had wanted to emphasise. In retrospect it now seems that I should have stressed that respondents should read the instructions before starting Questions 11, 12 and 13. Question 12 and 13 were answered in the same way for many of the respondents. Hence, I used Q13 in the analysis.

The information from this question was used to determine whether respondents were affected by anchoring bias. However, no evidence that respondents anchored on their 'most-worked' tectonic setting was found (Appendix 2).

Questions 14, 15 and 16

14) Who was your first geoscience related employer after finishing your highest degree?

15) Have you completed an industry graduate training course/programme?

16) If so, what was the duration of the course/programme, and what format did it take?

Question 14 asked for respondents' first geoscience-related employer, Question 15 asked if they had completed a graduate training course and Question 16 asked for the duration and format of the graduate training course (if undertaken). As geoscience companies train their graduates differently I decided to capture respondents' first geoscience-related employer. However, Question 14 could not be used in the analysis since the responses were very diverse (more than 200 separate employers, including universities and companies).

Many geoscience graduates participate in industry-based training courses that are organised by their employers to develop key geoscience skills. Graduate training courses are likely to be the first instances where geological data is interpreted at a professional level. The courses are usually years long and intensive. Hence, I decided to test whether there was evidence to suggest that graduate training courses improved respondents' interpretational abilities. Respondents gave a wide range of durations for their graduate course. Some respondents noted durations of less than 6 months and others quoted weeks. The shortest course duration was reported as being 2 days. Hence, not all respondents considered the same 'types' of graduate courses as I had anticipated. I did not use Q16 in the analysis because Q15 was non-significant (Appendix 5).

Questions 17 to 20

17) Have you been on a seismic interpretation course (not including university training)?

18) If so, what was the duration of the course?

19) Have you been on a structural geology course (not including university training)?

20) If so, what was the duration of the course?

Questions 17 and 19 asked whether respondents had attended a seismic interpretation course or a structural geology course, independent of their University training. Questions 18 and 20 asked for the durations of any courses attended to be able to validate responses. I decided to exclude University training courses as it can be hard to distinguish between normal degree experience, which is already captured in previous questions, and other focussed training courses which respondents might have attended.

Question 21

> *21) Please mark geographical locations where you have investigated the geology for more than 2 weeks in your entire geoscience career. (This should include everything; Ph.D. thesis, projects, scientific studies, fieldwork, etc.)*

Question 21 asked respondents to mark geographical locations onto a map where they had investigated the geology. Question 21 captured respondents' ranges of experience. The question stated that 'everything' should be included, e.g. thesis work, projects, scientific studies, fieldwork, etc... The map allowed more detail to be captured in a short timeframe and was expected to have triggered memories of past locations that had been worked at, e.g. many locations might have been forgotten or not included if respondents had been asked to write a list. I arbitrarily chose the cut-off duration to be '2 weeks' as this was not too short and not too long, e.g. it would include a reasonable number of locations and not miss out any important ones. The fact that the threshold was consistent for all respondents was more important than the choice of threshold.

Post-data collection, I found out that 75 respondents had missed out Question 21, which was more than anticipated. One respondent told me that they had thought it was 'just a picture' when I pointed out that they had not completed it. If I had removed the black line from around the edge of the question and increased the text size, the completion rate might have been higher.

The question is also slightly ambiguous as some respondents may have found the phrase 'geographical locations where you have investigated the geology' to be confusing. I had intended respondents to note the geographical locations where the geology physically was (e.g. locations of basins, oil fields and outcrops), rather than the location in which they studied the geology from, e.g. remotely. Therefore, a geologist who was investigating the geology of the South China Sea from their office in the UK would be expected to mark the 'South China Sea' on their map, rather than their office's location. Given the context of the questionnaire, it is thought that the respondents understood that I was interested in the locations of the geology.

The questions underneath the seismic image

Qi) How long did you spend interpreting the seismic image?

Qii) Would you have liked more time?

Qiii) What is your confidence; in your interpretation? in the linkage of faults?

I used Questions Qi and Qii to detect that respondents had enough time to complete the interpretation exercise. The analysis found that Qi and Qii were non-significant (Appendix 5), meaning that the time spent on the interpretation exercise, or whether participants' wanted 'more time' was not associated to respondents' interpretational ability.

Question Qiii asked for respondents' confidences in their interpretation; firstly in their interpretation overall, and then secondly, in the linkage of their interpreted faults. Respondents generally marked their confidence 'in their interpretation' to be very similar to their 'confidence in the linkage of their faults'. Question Qiii was also non-significant (Appendix 5), implying that confidence was not associated to interpretational ability.

3.5. Choice of interpretation exercise

The background questions captured information on respondents' backgrounds, while an interpretation exercise was needed to measure respondents' interpretational ability. Measuring interpretational ability allowed me to determine what backgrounds and interpretational approaches were associated to better interpretations. Due to time and logistical restraints, no attempt was made to have respondents repeat the interpretation exercise again (e.g. 12 months later) to test whether they produced a similar interpretation. In this section, the different types of interpretation exercise I considered are discussed.

3.5.1. Types of interpretation exercise

I chose to use seismic data as the experimental medium for this research to represent geological data in general. From here on, 'interpretational ability' will refer 'producing a valid interpretation of the seismic data', where *"valid"* is defined to be 'possible under the laws of geology'. The scoring system is explained in Chapter 4. Interpreting a seismic image is analogous to the interpretation of any type of geological data, e.g. geological maps or borehole logs. Talking about seismic interpretation in an interview (see Chapter 4), reference expert Prof. Stewart said:

> *"The fact that it's seismic to me is an irrelevance. I see this as being absolutely analogous to any other piece of geoscience interpretation" (lines 452 to 453).*

As in Chapter 2, all geological data are 'under-constrained', meaning that multiple interpretations are possible. Therefore, the results of this research are expected to be transferable to the interpretation of other types of geological data.

A 2D seismic image was chosen for this research instead of a 3D seismic dataset because 3D data would take respondents longer to interpret, add more complexity to the analysis and potentially decrease the sample size. A 2D seismic image could also be easily presented in questionnaire format. There are three main types of 2D seismic interpretation exercise that I considered using in the background questionnaire:

- One single seismic image for all questionnaires. I would give all respondents the same seismic image to interpret.
- Multiple seismic images. I would select two or three different seismic images and each respondent would be given one, at random, to interpret. The look of the questionnaire would be identical for all respondents (apart from the seismic image across pages 2 and 3).
- A seismic image with a choice of 2D interpretations. Instead of interpreting the seismic image, I would give respondents a seismic image and a choice of possible interpretations.

However, I decided to use one seismic image for the following four reasons:

- The interpretation of a single seismic image was the simplest of the types of interpretation exercise and would allow me to make direct comparisons between all respondents.
- The sample size would be maximised. If multiple seismic images had been used then the sample size would have effectively been divided by the

number of seismic images used and a much larger a sample would be needed to uphold the 5% error bound on estimates.

- Using one seismic image made the analysis quicker as considering the multiple interpretations of multiple seismic images requires more work.
- The uncertainty in the scoring system for respondents' interpretations would be minimised. The scoring system was consistent for all respondents as there was only one seismic image. If I had used multiple seismic images, the uncertainty in the scoring system would have increased.

Although using multiple seismic images would have increased the geological scope of the survey, gaining a large sample size was more important. I also considered using one 2D seismic image with a choice of 2D interpretations to choose from, but interpreting a seismic image from scratch addresses the research questions more effectively. Using one 2D seismic image was therefore the best option for this research.

The instruction for the interpretation exercise was kept simple because I did not want to instruct respondents on how they should approach the exercise, which allowed them to interpret the geological features that were important to them. The instruction was chosen to be: *"please interpret the whole seismic image"*, allowing respondents to use the approach that seemed most suitable. No regional context and no vertical scale were provided to ensure that respondents had no expectations about where the image was from and what geology to expect, i.e. to mitigate anchoring bias. For example, if I had said that the seismic image was from the Gulf of Mexico then some respondents would have assumed that there must be salt in the section (since the geology in the Gulf of Mexico is known to be influenced by salt tectonics).

3.5.2. What type of seismic data: real or synthetic?

After choosing 2D seismic data to represent geological data in general, I then had to decide whether I was going to use real or synthetic data. A synthetic seismic image is a forward-modelled seismic image based on a known structural model. In this scenario, the original geological model is known and can be considered the 'correct' interpretation to which respondents could be compared, e.g. Bond *et al.* (2007).

However, a ‘real’ seismic image is often richer in stratigraphic character than synthetic data. In their review of Bond *et al.* (2012), one of the reviewers noted that:

> *“One of the weaknesses of the study [Bond et al. 2012] is the reliance on one synthetic seismic profile. Even though I have over 30 years’ experience interpreting seismic data and visiting structural analogues in the field, I am not ashamed to note that I would no doubt have fallen into the majority of “experts” who “misinterpreted” the data in this study. The synthetic profile in this study has no similarity to any real seismic line I have ever worked on. Synthetics are good at showing the overall form and general seismic response, but they cannot duplicate the richness of the couplets as they respond to nuisances in rock rheology. The reflectors lack any real character and therefore the interpreter cannot make reliable correlation of reflectors across faults”.*

As I wanted to make the interpretation exercise similar to interpretation in industry, I decided to use real seismic data instead of synthetic data. If the chosen seismic image had been from a particular geological setting, I could have compared respondents’ interpretations to theoretical models, e.g. Torvela and Bond (2011).

One downside to using real data was that the ‘correct’ interpretation is always unknown. However, this fact is also true in industry interpretation as well. In reality, the ‘correctness’ of an interpretation depends on its usage. It sometimes does not matter whether an interpretation is ‘correct’ if sufficient volumes of hydrocarbons can still be produced. The varying definition of ‘success’, within industry, means that interpretations only need to be sufficiently correct for the intended purpose.

3.5.3. Choice of seismic image

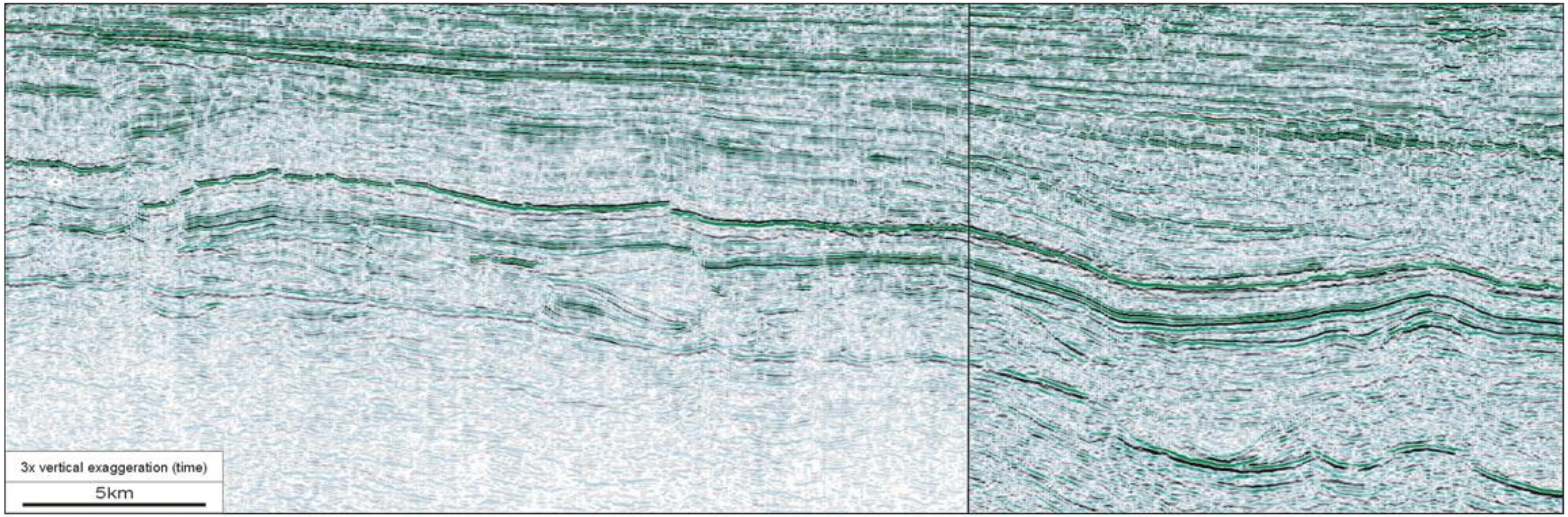
The choice of the specific 2D seismic image for the interpretation exercise was also an important decision. The chosen seismic image should represent a ‘typical’ unknown seismic image which professional interpreters are routinely faced with, to ensure that results gained from the analysis can be applied back into industry. For example, the seismic image could not have been widely known to respondents, nor could it be so obscure that the extrapolation of the results would be irrelevant to geoscience interpretation in industry.

Professor Shipton and Dr Bond considered different seismic images on the Virtual Seismic Atlas[8] and selected a 2D reflection seismic image (c. 43km in length) from the UK North Sea (peripheral graben system, blocks 20/20 and 21/16 area). The chosen seismic image was published and interpreted by Stewart (2007). The published version of the seismic image was time-migrated and presented at 6x vertical exaggeration. In the background questionnaire, the seismic image was presented at 3x vertical exaggeration. Due to the differing vertical exaggerations and the fact that the seismic image was one of 25 figures in Stewart (2007), respondents would be unlikely to remember the interpreted seismic image even if they had read the publication. I have used Stewart's interpretation as a reference interpretation to which I could compare respondents' interpretations (Chapter 4).

Figure 3.1 presents the chosen seismic image, which was displayed on pages 2 and 3 in the background questionnaire. I also added a box in the bottom left hand side of the image which said *"3x vertical exaggeration (time)"* to ensure that respondents knew the seismic image was vertically exaggerated and in time, i.e. not a depth-migrated seismic image.

[8] The Virtual Seismic Atlas: http://www.seismicatlas.org (last accessed: 16/1/2013).

Figure 3.1 – The seismic image that was presented in the background questionnaire and interpreted by six reference experts and over 700 geoscientists. The instruction along the top of the seismic image was: *"please interpret the whole seismic image"*.



3.5.4. How similar is the exercise to seismic interpretation in industry?

One aim of this research was to derive an interpretation workflow for the oil and gas industry that mitigates the risk arising from the uncertainty in geoscience interpretation. Hence, it is important to consider how the interpretation exercise relates to interpretation in the oil and gas industry.

Gao (2009) noted that 3D seismic data is interpreted in 2D sections, but also advised that a 2D interpretation does not make use of the full value of 3D data:

> *"Although the 3D seismic data offer a unique opportunity to make seismic observations and geologic interpretations in 3D space, most 3D seismic data are displayed and interpreted in a 2D manner, leaving the critical advantage and potential value of 3D seismic data underused".*

In addition, the geoscientists who participated in the industry workshops (Chapter 7) indicated that it is normal to interpret 2D seismic images to assist in the interpretation of 3D datasets. 32 out of 49 participants stated that they build a template model (e.g. a cartoon) to aid their 3D interpretation either 'always' (4 participants), 'often' (11) or 'sometimes' (17). One workshop participant, a senior manager in geophysics, also commented that it was common to interpret a single 2D seismic image to aid the 3D interpretation.

3.5.5. What 'type' of interpretation was expected?

An interesting concern that was raised by a respondent was that the interpretation exercise presented in the questionnaire seemed to be of a 'structural' nature and did not allow for other types of interpretations (G. Bertram, 2009; pers. comm., Appendix 2):

> *"It is often said that the geologist who has seen the most rocks is the best geologist and I think this applies to seismic interpreters as well. We are all products of our own experiences and our views are shaped by our backgrounds. The same could be said for your exercises. The concept of seismic interpretation implied in the questions (and the space provided for answers) is very much influenced by simple seismic surface correlation and structural interpretation. (Is this the background Midland Valley influence?) Basic correlation is only part of what is done (or ought to be done) in a seismic interpretation. As you build your analysis of a data set your structural*

model must go hand in hand with your understanding of the stratigraphic and lithological relationships. To attempt an interpretation concentrating on only one aspect of the geology is to restrict the data available to you".

Similarly, Gao (2009) noted that 'seismic interpretation' also includes:

"Data selection and conditioning, structure and facies characterization, prospect evaluation and generation and well-bore planning", and "seismic structures, facies and hydrocarbon systems".

The uncertainty in knowing what interpretational approach was expected to be used in the interpretation exercise was noted by respondent 303:

"I do not understand what you mean by 'interpret'. I have drawn in faults".

Respondent 303 was not clear on how to approach the seismic interpretation exercise, but still chose to interpret faults to identify the discontinuities in reflector horizons. During the filtering process, respondent 303 was removed from the sample due to being inexperienced.

Given the concerns noted above, the following three decisions were made to ensure that the interpretation exercise was applicable to different 'types' of seismic interpretation, rather than just the correlation of reflectors:

- Respondents were not told how to interpret the seismic image and could use whatever approach they deemed applicable.
- Multiple reference experts with different technical backgrounds were recruited.
- Experts could choose any type of feature as I did not specify what 'key features were'. As in Chapter 4, the instructions that were given to the reference experts were:

    "The term 'features' is meant to be vague, but can include different types of faults, horizons, sedimentary packages, folds, etc. Feel free to add other types of features. (These are the features that help define the tectonic setting and/or stratigraphic setting of your interpretation)".

Thus, I remained objective by allowing the respondents to interpret the seismic image using the approach that seemed most appropriate to them, by recruiting multiple reference experts, and giving the reference experts the freedom to define their own key features.

In any case, a ‘structural’ interpretation was completed by all respondents, to some extent, since they interpreted at least one horizon or fault. Far fewer respondents included, say, sedimentological features in their interpretation. Hence, if sedimentological features were used as the basis for my scoring system, I would have had to exclude a large proportion of the dataset since so few respondents used that approach. I did not count up how many respondents used different approaches due to difficulties in defining the approaches.

3.6. Data collection

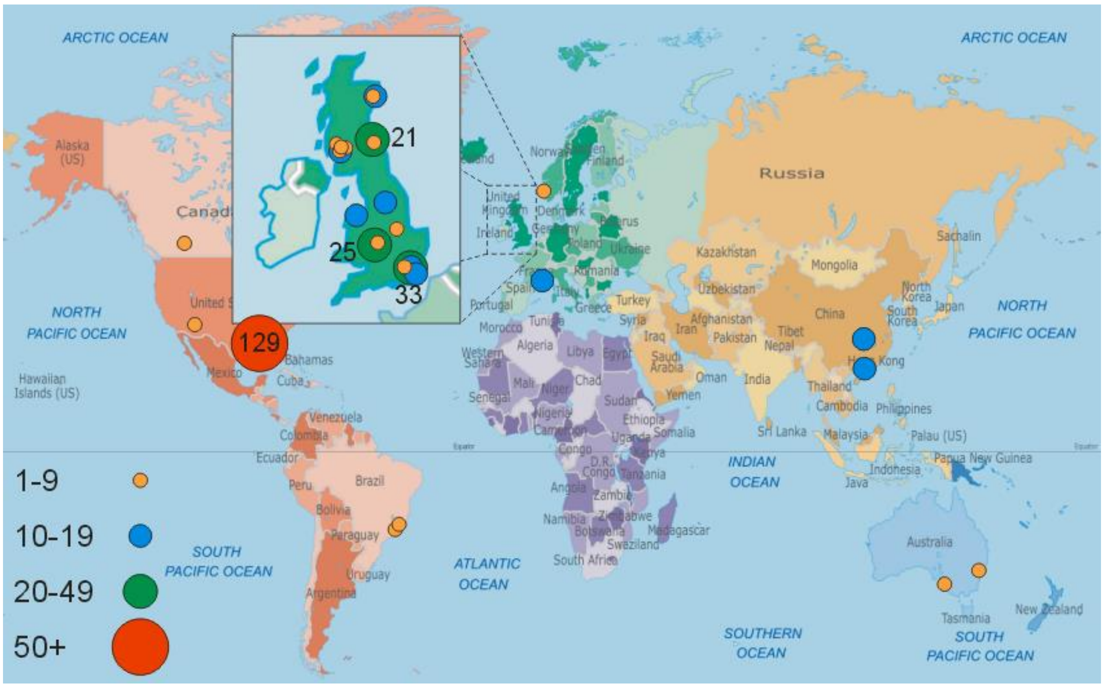
The survey was planned and executed with the quality of the dataset being the first priority. I thus, made the questionnaire as clear as possible. I collected the main dataset internationally and in a range of working environments to ensure a diverse cross-section of respondents. I collected a large sample to ensure that the statistical error bound on estimates of the population was <5%. Pens and/or coloured pencils were provided for respondents to answer the questions and interpret the seismic image.

3.6.1. Sampling locations

I distributed the questionnaire at conferences (including workshops), energy companies and in University departments. The respondents thus fell into natural groups of where they had received the questionnaire. I referred to each of these groups as a ‘batch’. Each batch was given a unique batch number, written at the top right of the questionnaire before it is distributed. The batch number enabled me to identify where the questionnaire was originally distributed, when it was distributed and by whom, should it be returned to me at a later date with no other information. In some cases, selected individuals distributed the background questionnaire on my behalf. For example, geologists employed by Midland Valley Exploration visited two oil companies in Australia, where they distributed questionnaires to industry personnel on my behalf. See Appendix 2 for the full list of sampling locations for the survey (e.g. the conferences and seminars where the questionnaire was distributed).

Figure 3.2 presents the geographical locations where the background questionnaires were distributed over the period 2009 to 2011 for the filtered sample of 420 respondents. As explained in section 3.7.1, 283 respondents were filtered out of the sample due to being 'inexperienced' or having 'inappropriate' backgrounds. Hence, 703 interpretations of the seismic image were collected in total. I distributed around 1400 questionnaires and hence, achieved a response rate of about 50%. I did not record precisely how many questionnaires were distributed and focussed on maximising returns. The 24 participants from the workshop 'control' group were added to the sample at a later date (see Chapter 7), giving a total of 444 respondents for the statistical analysis in Chapter 6 (e.g. 703 – 283 + 24 = 444). The 'filtered' sample of 444 respondents will hereby be known as the 'main dataset'.

Figure 3.2 – World map showing approximate locations where the background questionnaire was distributed. The size and colour of the dot, as defined in the legend (bottom left), reflects the number of respondents recruited at each location. For green and red dots, the numbers of respondents are noted for each location.



Twenty-two of the 29 sampling locations for the survey sample were located within Europe and America. As such, it was likely that the sample consisted of mostly European and American geoscientists. However, it should be emphasised that conferences, seminars and workshops attracted international delegates as well. I did not collect nationality or ethnicity data from respondents and hence, it was not possible to say what proportion of respondents were nationals from specific countries or continents. Table 3.1 shows that 390 respondents received the questionnaire in Europe or the USA, and only 54 respondents received the questionnaire elsewhere. Thus, for this research, the underlying population was those geoscientists, from academia and industry, who regularly worked in Europe or the USA.

Table 3.1 – The number of respondents recruited at each sampling location. 390 respondents (87.8%) were sampled in Europe and the USA; while, only 54 respondents (12.2%) were sampled elsewhere (grey boxes).

Sampling location	**No. of respondents sampled (%)**
Europe	253 (57)
USA	137 (30.9)
China	27 (6.1)
Brazil	14 (3.2)
Australia	9 (2.0)
Canada	4 (0.9)
Total	*444*

3.6.2. Sampling environment

The sampling environment batches fitted into two categories that described the ‘average conditions’ for respondents. It was therefore important to compare these two environments in case they systematically affected respondents’ interpretational abilities. Table 3.2 shows the sampling environment and the number of respondents in each category.

Table 3.2 – The numbers of respondents that were sampled at conferences, in energy companies and in university departments. The 'sampling environment' is noted for each sampling location.

Sampling location	Number of respondents (%)	Sampling environment
Conferences	240 (54.1)	'conference'
Energy companies	93 (20.9)	'seminar'
University departments	111 (25.0)	'seminar'
Total:	444	

The two sampling environments are defined in Table 3.3. In the 'seminar' environment, respondents generally filled in the questionnaire away from the pressures of their desk (e.g. in a seminar, workshop or 'lunch and learn' type event), making it easier for respondents to focus on the exercise. For example, batches of questionnaires were sent to managers in companies who then organised an hour's seminar where geoscientists interpreted the seismic image and then discussed their interpretations. I was not present in these instances but prepared a detailed instruction sheet.

Respondents in the 'conference' environment received the questionnaire at a conference but were not required to complete it right away, e.g. some respondents filled it in during coffee breaks, while others took it home and returned it the next day. Hence, there is uncertainty in the actual sampling environment for 'conference' respondents. I assumed that 'seminar' respondents completed the questionnaire in a quieter and more focussed environment than 'conference' respondents. Pre-data collection, due to the additional time and quieter environment, I expected the 'seminar' respondents to produce better interpretations than the 'conference' respondents.

Table 3.3 – A description of the two sampling environments that I characterised respondents into. The 'seminar' respondents spent 5 minutes, on average, longer on the interpretation exercise than the 'conference' respondents.

Sampling environment	Description of sampling environments	Median time spent on the interpretation
'Conference'	Background questionnaires were distributed at conferences. The exercise was generally completed in a noisier atmosphere, e.g. during coffee breaks. Some respondents completed the questionnaire out with the conference.	10
'Seminar'	Background questionnaires were distributed at a seminars or workshop events. The exercise was completed in a quiet atmosphere.	15

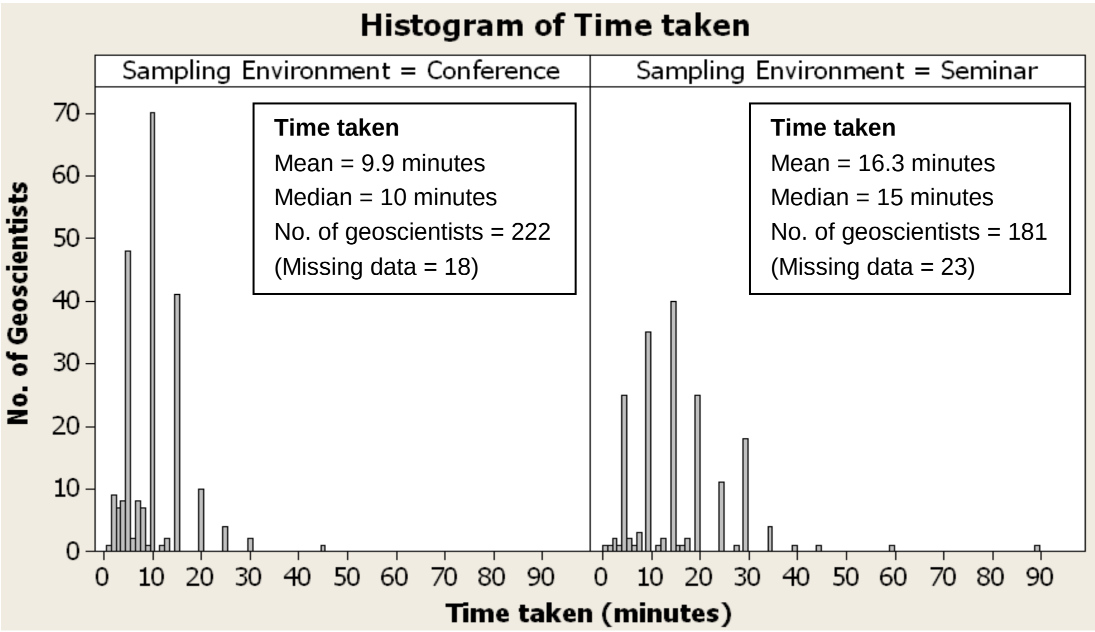
Thus, I decided to test whether there was any evidence that 'conference' respondents had been disadvantaged. The analysis showed that 'seminar' respondents did not produce better interpretations than 'conference' respondents (p=0.989, Appendix 7). (The scoring system that evaluated respondents' interpretations is explained in Chapter 4). The analysis actually showed that, on average, the 'conference' respondents produced better interpretations than the seminar group. However, this anomaly was explained after I took into account the differing levels of experience within the two sampling environment groups. 'Conference' respondents typically had more experience than 'seminar' respondents, possibly because experienced geoscientists were more willing to participate in the survey whilst at a conference, rather than attend a seminar during office hours. Although, it is also possible that more experienced respondents were sampled at conferences because they chose to take part and inexperienced respondents systematically declined (e.g. self-selection bias), but this could not be tested. Hence, the sampling environment was not an important factor, and the 'conference' respondents were not disadvantaged.

3.6.3. Time spent on the exercise

The time taken to complete the seismic interpretation exercise was collected on the background questionnaire (this did not include the time spent filling in the questionnaire). The 'time taken' was collected to test whether it was associated to producing a valid interpretation, e.g. perhaps respondents who produced inferior interpretations did not have enough time. However, the analysis showed that the ‘time taken’ was non-significant (Appendix 5). Hence, the time spent on the interpretation exercise was not linked to whether respondents produced a valid interpretation. I also observed that the ‘wanted more time’ variable (Qii) was non-significant (Appendix 5).

The distributions of the time taken are illustrated in Figure 3.3, split by sampling environment. It is interesting that many respondents filled-in ‘rounded’ times, e.g. ‘15 minutes’ instead of ‘14 minutes’ or ‘13 minutes’, which explains the multi-peaked distributions. These data are highly skewed, with most respondents taking between 5 and 20 minutes, and a few respondents taking more than 40 minutes. The maximum time taken was 90 minutes, and the median time taken for the sample as a whole (i.e. both sampling environments) was 10 minutes.

Figure 3.3 – The time taken to interpret the seismic image for each ‘sampling environment’ group. ‘Seminar’ respondents generally spent longer on their interpretation than conference respondents. The bin size is 1 minute.



3.7. Data filtering and representativeness

In this section, I explain how the sample was filtered to remove respondents who were inexperienced or who had inappropriate backgrounds, and show that the sample represented the underlying population of geoscientists.

3.7.1. Filtering out inexperienced respondents

I distributed the questionnaire to as many geoscientists as possible to gain a large sample. Hence, some of the individuals who completed the questionnaire might have been inappropriate for the research. For example, some geoscientists might have had no experience in seismic interpretation or in structural geology but still took part in the survey because I handed them a questionnaire without realising that their experiential background was inappropriate. Ultimately, I wanted the results of the research to be applicable to professional geoscientists. Therefore, it was important to filter out any geoscientists who had inappropriate backgrounds for this exercise, and also those geoscientists who were too inexperienced. I decided that the minimum experience, for inclusion in the analysis, would be 2 years (after the completion of their highest degree).

Thus, the filtering criteria used were: ‘<21 years old’, ‘No university degree’, ‘<2 years of experience after the completion of their highest degree’, ‘no experience in seismic interpretation’, or ‘no experience in structural geology’. In total, 283 respondents fell into at least one of these categories and were excluded from the analysis, leaving 420 respondents. The 24 control workshop participants (Chapter 7) were then added to the dataset, giving a grand total of 444 respondents for the analysis. The workshop control participants completed the seismic interpretation exercise with the same instructions, and under the same conditions, as the ‘seminar’ respondents in the main dataset. All statistics and data hereon in refer to the filtered sample of 444 respondents unless otherwise stated.

3.7.2. Underlying geoscience population

To ensure that the survey results were representative of the underlying geoscientist population, I wanted the sample to be similar in terms of age and gender. However, since it is impossible to know the age and gender statistics for the population precisely, I approximated them with age and gender statistics from four large geoscience organisations. These figures were used to validate the sample. The four geoscience organisations that were contacted provided anonymous age and gender data from their membership lists (Table 3.4). The data was given freely but under the assurance that the data would be used only for checking that the main dataset was truly representative of the range of ages and genders in the population. It is likely that some geoscientists were members of multiple organisations.

The membership data from the four organisations were likely to be reasonably representative of the underlying population since the organisations were large and based in Europe or the USA. The data were collected in 2009, which was the same year I started the survey. Since the geoscience organisations cover slightly different subject specialities, no particular organisation was more reliable than the others with regards to the true but unknown age distribution and gender proportions in the wider population. Hence, it was assumed that the data provided by the four organisations were of equal quality, and that the membership data from the four separate organisations is more reliable when taken together than any one organisation on its own. In some of the graphs below, the totals are different due to missing data in the organisations' membership lists and non-response to the background questionnaire.

Table 3.4 – The four geoscience organisations that provided anonymous age and gender data for their membership list. These data are used to validate the demographics of the main dataset.

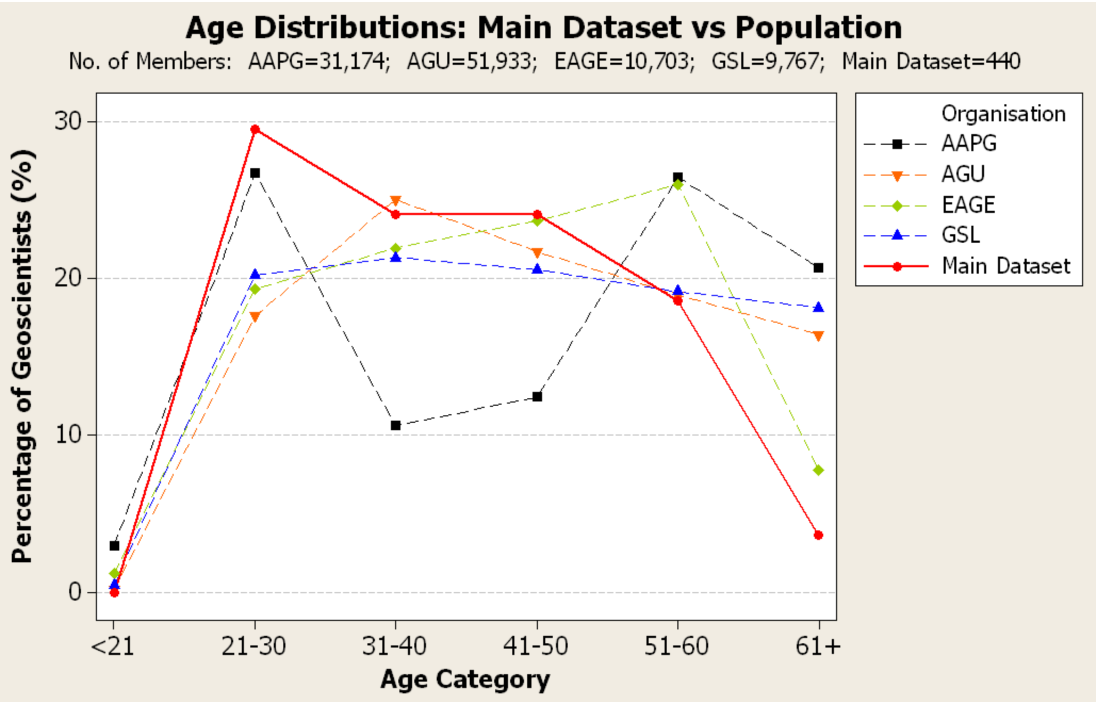
Organisation	**Acronym**	**Subject areas of interest (as taken from website)**	**Date of data collection**	**No. of members**
American Association of Petroleum Geologists http://www.aapg.org/	AAPG	Geoscience	31st December 2009	35,627
American Geophysical Union http://www.agu.org/	AGU	Earth and space science	30th September 2009	57,185
European Association of Geoscientists and Engineers http://www.eage.org/	EAGE	Geophysics, petroleum exploration, geology, reservoir engineering, mining and civil engineering	2009	13,703
Geological Society of London	GSL	Geoscience	2009	9,930

http://www.geolsoc.org.uk/				

3.7.3. Demographics (age and gender) of the sample

The percentages of geoscientists in each age category for the four geoscience organisations, and for the main dataset, are shown in Figure 3.4. It can be seen that the main dataset slightly over-sampled the 21-30 year olds and slightly under-sampled the 61+ year olds. The variation between the geoscience organisations themselves is also noteworthy. The percentage of 31-50 year old members in the AAPG organisation was much lower than that of the other geoscience organisations. The EAGE had a much smaller percentage of 61+ year old members than the three other geoscience organisations. Considering the fact that there is clear variation in the age distributions for the geoscience organisations, it can be said that the main dataset adequately represented the underlying population of geoscientists in terms of respondents' ages.

Figure 3.4 – Age distributions for the four geoscience organisations and the main dataset. The main dataset was a good match to the geoscience organisations, but slightly over-sampled the '21-30' year olds and under-sampled the '61+' year olds.



The impact of filtering out the least-experienced respondents improved the match between the age distributions of the main dataset and the underlying geoscientist population. By removing the 283 least-experienced respondents via the filtering criteria, the percentage of respondents who were aged 21-30 was reduced from 48.9% to 29.6%, which was closer to the population value. The reason for so many of the original sample being aged ‘21-30’ (48.9%) was simply because it was easier to sample younger geoscientists, e.g. they seemed more willing to participate than older geoscientists. Also, the sampling strategy involved distributing the questionnaire in University departments, which might have attracted greater numbers of students to take part.

Table 3.5 shows that the average percentage of female members for the geoscience organisations was 18.9%. The percentage of female respondents in the main dataset was 21.2%, which is reasonably close to the population estimate of 18.9%. It is interesting to note that the EAGE had only 16.1% female members, while the AGU had 23.4% female members.

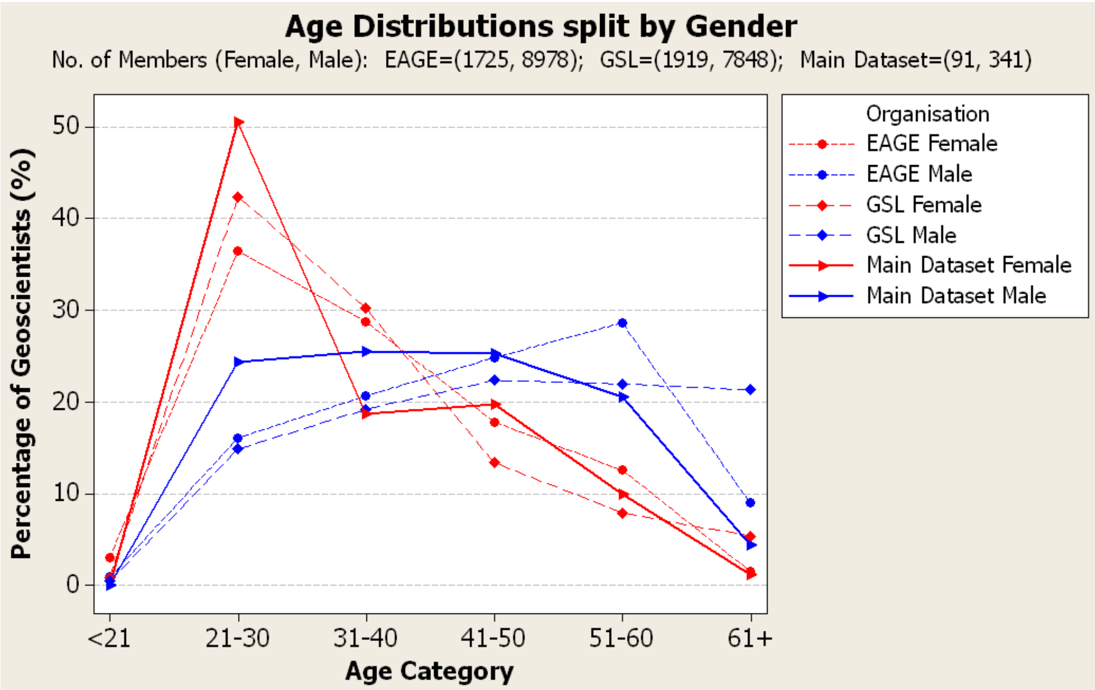
Table 3.5 – Gender proportions for the four geoscience organisations and the main dataset. The main dataset was a good match to the geoscience organisations with approximately the same proportion of female members.

Organisation	Female (%)	Male (%)	No. of members
AAPG	16.5	83.5	35,627
AGU	23.4	76.6	47,875
EAGE	16.1	83.9	10,703
GSL	19.5	80.5	9,924
Mean %	*18.9*	*81.1*	
Main dataset	21.2	78.8	433

Figure 3.5 shows the age distributions of respondents, split by gender, for the EAGE, GSL and main dataset. The AAPG and AGU did not provide cross-referenced data and thus, could not be included here. For the EAGE, GSL and main dataset, the majority of female geoscientists were aged 21-30, with relatively fewer older geoscientists. The male geoscientists were more equally split over the age categories. For the female geoscientists, the main dataset over-sampled the 21-30

year olds and under-sampled the 31-40 year olds. For the male geoscientists, the main dataset slightly over-sampled the 21-30 and 31-40 year old categories and under-sampled the 61+ category. However, these differences are relatively small compared to the uncertainty in the age distributions for the geoscience organisations.

Figure 3.5 – Age distributions for the main dataset and two geoscience organisations, split by gender. In all cases, the majority of females in each group were aged 21-30 (red lines). The spread of male geoscientists between the age categories (blue lines) was more equal.



The above comparisons show that the main dataset adequately represented the age distribution and gender proportions seen in the four geoscience organisations; and therefore, adequately represented the underlying population of those geoscientists who work in Europe or the USA.

3.8. Conclusions

This chapter detailed the methodology used to collect a large sample of geoscientists' interpretations of a single dataset and also their education and professional background. The questionnaire was designed to be clear and was put through numerous testing phases. A real 2D seismic image was chosen as the interpretation exercise; it represented all types of geological data since it was under-constrained and multiple interpretations were possible. Respondents were not instructed on how to interpret the seismic image and could use whatever approaches they deemed applicable.

The questionnaire was distributed to more than one thousand geoscientists, spread over five continents, during 2009 to 2011. A sample of 703 questionnaires was achieved. The active sample size was reduced to 444 geoscientists due to the filtering process used, but ensured a high-quality sample as inexperienced geoscientists were removed. By comparing the sample demographics (age and gender) to the statistics from four large geoscience organisations, I found that the collected sample was a good match to the underlying population.

Apart from Bond *et al.* (2007), large samples demonstrating uncertainty in geoscience interpretation are seldom captured due to time restraints. Therefore, the collected sample is a valuable resource. One of the key strengths of the sample is the high level of detail known about the respondents. Respondents' interpretational abilities, in the form of whether they produced a valid interpretation of the seismic image, will be assessed. The background factors and interpretational techniques that were most associated to producing a valid interpretation of the seismic image will be determined.

Chapter 4 – Reference Experts

This chapter explains how I recruited the six reference experts and formed a scoring system to assess respondents' interpretations. The reference experts were invited to interpret the seismic image, explain their interpretation in detail and then provide key geological features that were integral to their interpretation. The key features were then used to create the response variable for the main analysis of respondents' backgrounds and the techniques used. I also compare and contrast the reference experts' interpretations and derive the geological evolution of their respective interpretations.

4.1. Rationale for using reference experts to score interpretations

To determine what background factors and interpretational approaches had the greatest influence on respondents' interpretations, a scoring system was needed. Since it is not possible to define a unique underlying 'correct' interpretation of the seismic image, I decided to compare respondents to hand-picked reference experts. The chosen experts were highly experienced and respected professionals within their disciplines. The experts interpreted the seismic image in detail and explained their interpretation. The comparison allowed me to validate respondents' interpretations using a 'scale of similarity', allowing me to determine *how similar* respondents' interpretations were. Hence, my approach was different to Bond *et al.* (2007) who used a binary response variable.

In the oil and gas industry, and in other industries, experienced geoscientists are often hired as consultants to assist in the interpretation of geological data. In larger companies, internal highly experienced geoscientists sometimes 'quality control' other geoscientists' interpretations. It was therefore reasonable to compare respondents' interpretations against the reference experts' interpretations as the reference experts were similar, in terms of experience, to typical geoscience experts that are used in academia and industry. I name five of the six reference experts in the next section and discuss their backgrounds. One expert preferred to remain anonymous.

4.1.1. Introducing the reference experts

The reference experts were selected by Prof. Shipton and Dr Bond based on their differing technical experience. The five experts who agreed to be acknowledged are shown in Table 4.1. Although five of the six experts are named, it is not revealed which expert completed which interpretation. I agreed with the reference experts that the interpretations would be anonymised.

Table 4.1 – Details of the reference experts who were recruited to interpreted the seismic image. The experts' respective years of experience (after the completion of their Ph.D.) and their areas of expertise are noted. One expert preferred to remain anonymous.

Reference expert's name (alphabetical by surname)	**Q5: Years of experience, in 2011 (after completion of a Ph.D.)**	**Areas of expertise**
Prof. Joe Cartwright University of Oxford, UK	23	3D seismic data, sequence stratigraphy, structural analysis, basin hydrodynamics and seismic interpretation. http://www.earth.ox.ac.uk/people/profiles/academic/joec
Dr Conrad Childs University College Dublin, Ireland	25	The geometry and evolution of faults and fault systems from outcrop to basin scale. Evolution of the internal structure of faults and fault zones. The hydraulic properties of faults. http://www.ucd.ie/research/people/geologicalsciences/drconradchilds/
Dr Alan Gibbs Midland Valley Exploration, Ltd.,UK	37	Structural geology and the applications of structural geology for the oil and gas industry. http://www.mve.com/about/people/dr-alan-gibbs
Dr Mark Rowan Rowan Consulting, Inc., USA	20	Styles and processes of salt tectonics, salt-sediment interaction, the geometry and kinematics of fold-and-thrust belts, and the applications to petroleum exploration.
Prof. Simon Stewart Saudi Aramco, Saudi Arabia	18	Geophysics, reflection seismic interpretation, structural geology, well planning and operations geology. Structural Interpretation of reflection seismic data. Fault patterns. Salt

		tectonics. Structural styles. Impact craters. Mud volcanoes.
Anonymous	24	Sedimentology

Table 4.1 shows that the reference experts' areas of expertise were focussed around seismic interpretation, structural geology and tectonics. The anonymous expert specialised in sedimentology, although their list of expertise has been shortened to preserve anonymity. Cartwright, Childs and Stewart were academics; while, Gibbs and Rowan worked in industry. Both Cartwright and Stewart had previous experience working for super-major oil companies and Childs had experience consulting for oil companies. Gibbs had previous experience in academia before moving into industry. Thus, the reference experts had a wide range of experience that better represents the underlying population than if I had only recruited only one expert.

I initially compared respondents' interpretations to Prof. Stewart's published interpretation of the seismic image (Stewart, 2007), but decided to contact other reference experts to capture a range of interpretations of the dataset. Prof. Stewart was interviewed twice in March 2010 to establish contextual information, e.g. where the seismic data was collected, and to gain a detailed understanding of his published interpretation. The interview also covered work practices in the oil and gas industry. The interviews (151 minutes in total) were audio recorded, professionally transcribed, and then checked by me. Quotes from the interview transcription are referenced by line number. The full interview transcript is included in Appendix 3.

4.1.2. Additional information on the seismic image

Prof. Stewart had additional information as he had encountered the seismic image whilst working for a super-major oil company. Prof. Stewart confirmed that over a period of two years, he worked on the area and had access to more data (and time) than the respondents in the main dataset, such as multiple 3D surveys:

> *"For the purposes of putting together a published interpretation that you've got on the table today, almost certainly I would have just done that [the published seismic interpretation] in PowerPoint as I'm preparing the paper. But that was on the back of me having worked in this area on and off for a couple of years on various 3D surveys"* (lines 1160 to 1163).

Prof. Stewart said that he thought that his interpretation could be derived without the additional information he had access to, as there was enough evidence in the seismic image. Therefore, any respondent should potentially be able to produce a comparable interpretation. Dr Gibbs also had additional knowledge about the seismic image. Dr Gibbs stated (pers. comm.):

> *"The seismic image is from a data set that we had access to and worked on for a client project several years ago",* and *"what I did was obviously influenced by my knowledge of that earlier work and the structural and regional setting".*

However, it should be stressed that Dr Gibbs did not take an active role in this research. Dr Gibbs was given a verbal 'progress update' every 6 to 12 months and also asked to help with specific tasks, such as giving feedback on presentations and being a reference expert. The other four reference experts did not have additional information of the seismic image.

4.1.3. How I elicited the experts' interpretations

The reference experts' interpretations were elicited independently and I thus, did not conduct an expert elicitation process since no group interaction was possible and herding bias could not occur. The reference experts were each sent copies of the background questionnaire and seismic image by post. The experts were asked to interpret the seismic image and identify key geological features in their interpretation, which would form the basis of the scoring system for respondents' interpretations.

The given instructions were as follows:

> *"i. Please complete the background questionnaire (the A4 sheets).*
>
> *ii. Please spend at least 30 minutes interpreting the seismic image (the A3 sheet).*
>
> *iii. Provide a short summary of your interpretation underneath the seismic image. Your interpretation will be one of six reference interpretations and it is hence important that you give as thorough an interpretation as possible.*
>
> *iv. Identify and list between 3 and 8 key geological features that are essential to your interpretation, on the back of the A3 sheet. The term 'features' is*

meant to be vague, but can include different types of faults, horizons, sedimentary packages, folds, etc. Feel free to add other types of features. (These are the features that help define the tectonic setting and/or stratigraphic setting of your interpretation).

v. Complete the questions in the red box underneath the seismic image.

vi. Return all worksheets to me in the addressed envelope, which is enclosed for your convenience. Please email me to confirm when you have posted your interpretation back".

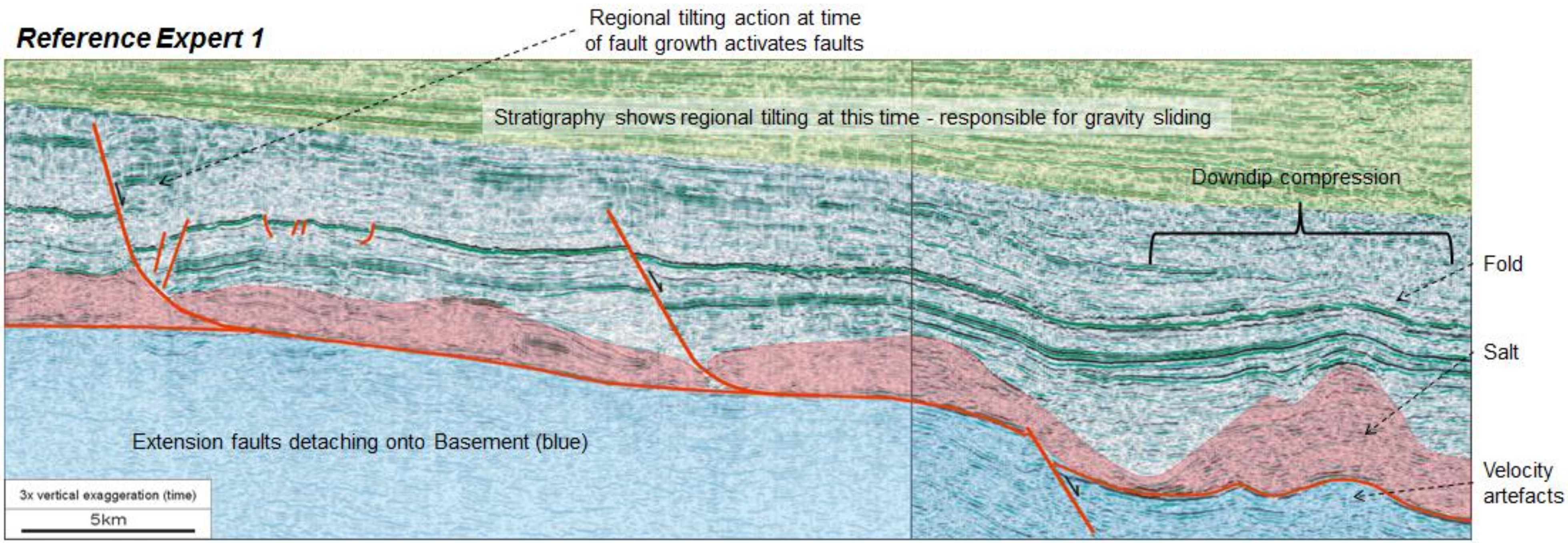
The reference experts were asked to spend at least 30 minutes completing the interpretation exercise in addition to explaining their interpretations. The experts' interpretations are presented and discussed in the next section.

4.2. Introducing the reference experts' interpretations

Digitised versions of the six reference experts' interpretations are shown on the following pages with the respective expert's summary of their interpretation underneath. I reproduced the experts' interpretations accurately, although I did not include all of the detail in Expert 4 and 6's interpretations for purposes of clarity. Un-annotated photographs of the original interpretations can be found in Appendix 3. I also added annotations to the experts' interpretations to highlight the key aspects from their explanation notes. The explanations were copied verbatim and then any grammatical corrections were made, e.g. clearer punctuation. Expert 1's explanation notes were included in a personal communication and are presented verbatim. I added comments in square brackets to Expert 1's notes to better link the notes to the interpretation.

The interpretations below have been anonymised and listed in an approximate progression, in terms of structural style, from a detached listric normal fault interpretation to a strike-slip interpretation. Because the reference experts used different labelling systems, a legend was included (Figure 4.7). The names given to horizons and packages in this legend will be used from here on. The legend was based on the interpretation by Expert 4 as they had interpreted the greatest number of horizons.

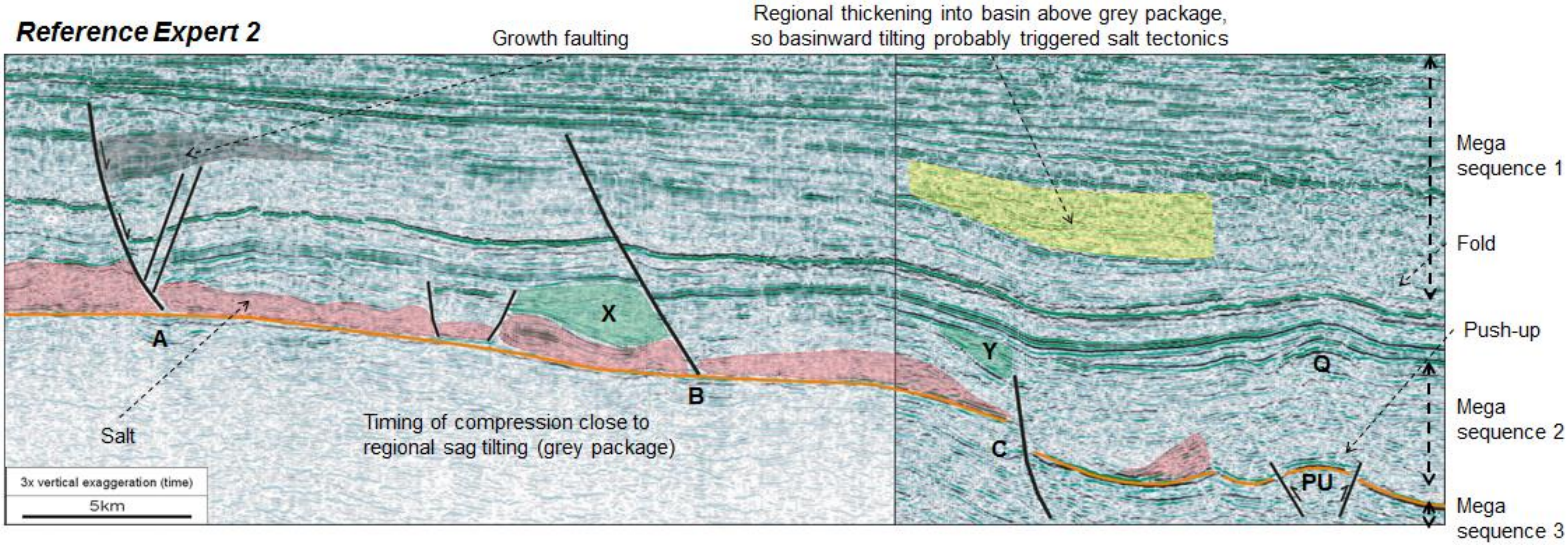
Figure 4.1 – Digitised interpretation for reference expert 1 with their explanation notes. Expert 1 interpreted listric normal faults detaching on the salt basement, with downdip compression to the right. *“Velocity artefacts”* were interpreted on the right hand side. The expert’s interpretation focussed on the main structural features without much interpretation of smaller features. To better link the notes to the interpretation, I added package labels in square brackets.



Expert’s interpretation notes

“The base detachment surface is largely unfaulted and has got a regional dip, which given the vertical exaggeration, is a couple of degrees. There looks to be a significant salt layer [pink] across the seismic image. There is also a fold structure on the right-hand side. This is the classic recipe for detachment tectonics: updip extensional domains and downdip compressional domains, which we have here. The ‘bumps’ on the right-hand side of the Basement horizon are probably velocity artefacts. The large faults are listric and sole out onto the Basement [blue]. A late regional tilting action probably activated the listric faults and caused the gravity sliding – timing shown by stratigraphy [yellow]. Tilting occurs at time of growth on listric faults”.

Figure 4.2 – Digitised interpretation for reference expert 2 with their explanation notes. Expert 2 interpreted three *"mega sequences"* showing the timing of packages. The left, middle and right faults are extensional, detaching on horizon 1. A fold on the right was identified. They identify packages X and Y as showing earlier movement.

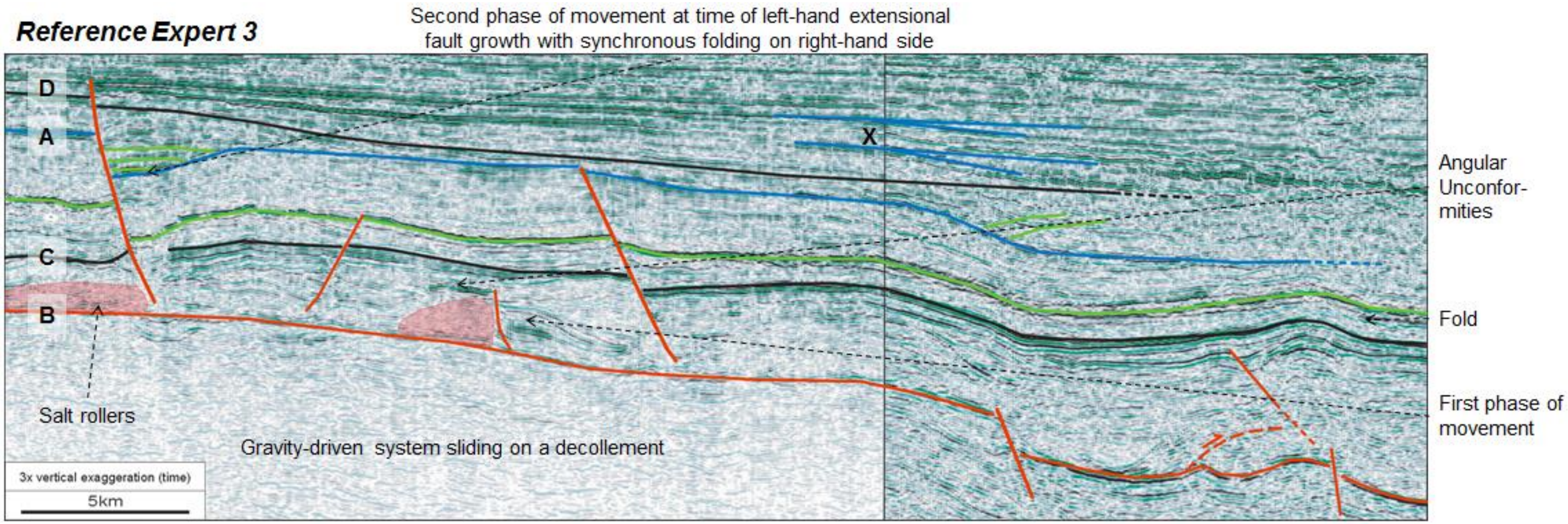


Expert's interpretation notes

"This section is of a portion of a salt-floored basin. Top salt is very hard to pick, and I suspect that some of the upper evaporites may be reflective. I based my Top Salt pick on displacements on the major salt withdrawal faults A, B. It's also possible that the salt has largely welded. Basement fault C, has a modest offset, so this is a 'platform' area, rather than a major extensional (graben) feature. Growth faulting is evident on A, B in the upper megasequence 1 with a clastic divergent wedge (grey). Regional thickening into the basin (sag-style) evident above this grey package – so basinward tilting probably triggered salt sections (latest phase, updip extension). Anticline Q may involve some basement compression. PU – push up structure, possibly inverted pre-salt faults. Timing of compression is close to regional sag tilting (grey package). I have largely ignored the pre-salt data as the data quality is poor. Packages 'X' and 'Y' – maybe an earlier phase of movement".

"Detachment surface (base salt)" is also noted on the back on the interpretation.

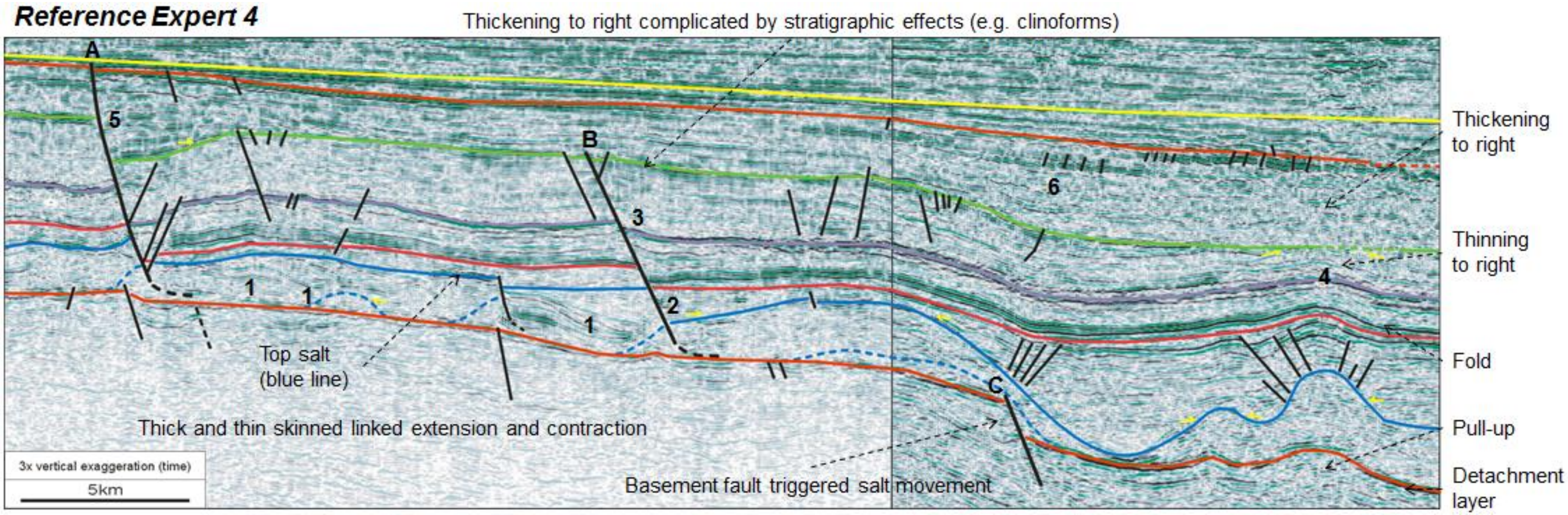
Figure 4.3 – Digitised interpretation for reference expert 3 with their explanation notes. Expert 3 interpreted two phases of movement. The left, middle and right faults were interpreted as being extensional, sliding on a décollement. A fold on the right is identified. Horizon 2 was interpreted as a *"locally angular unconformity"*.



Expert's interpretation notes

"According to my interpretation this is a mainly gravity driven system with sliding on a décollement (B). I have interpreted two phases of movement. The first is indicated by normal faults and steep bed dips between horizons B and C. Lower amplitude zones at this level are interpreted to be salt or some other relatively mobile material which may form 'rollers', and forms the décollement for the second phase of extension. This deformed B-C interval is overlain by A-C deposited in a tectonically quiescent period with C forming a locally angular unconformity, truncating underlying dipping reflectors. A second phase of deformation initiated at A with extensional faulting on the left of the section and synchronous folding on the right indicating gravitational sliding towards the right on the décollement B. This deformation occurred mainly in the interval between the deposition of A and D. Growth strata associated with both normal faulting and folding occur in this interval. It is possible there is some thrusting associated with the fold above the level of the décollement. This second phase of deformation may have been triggered by movement on a normal fault at depth which offsets B, and thickness variations in A-D directly above this fault suggest that it was active at this time. Local truncations (labelled X) above horizon D indicate that this fault may have had some activity later than D, as has the normal fault on the left of the section".

Figure 4.4 – Digitised interpretation for reference expert 4 with their explanation notes. Expert 4 interpreted the left, middle and right faults to be extensional. The left and middle faults are detached on horizon 1, offsetting the salt layer. A fold on the right was identified. The right side of horizon 1 is interpreted to be a pull-up.



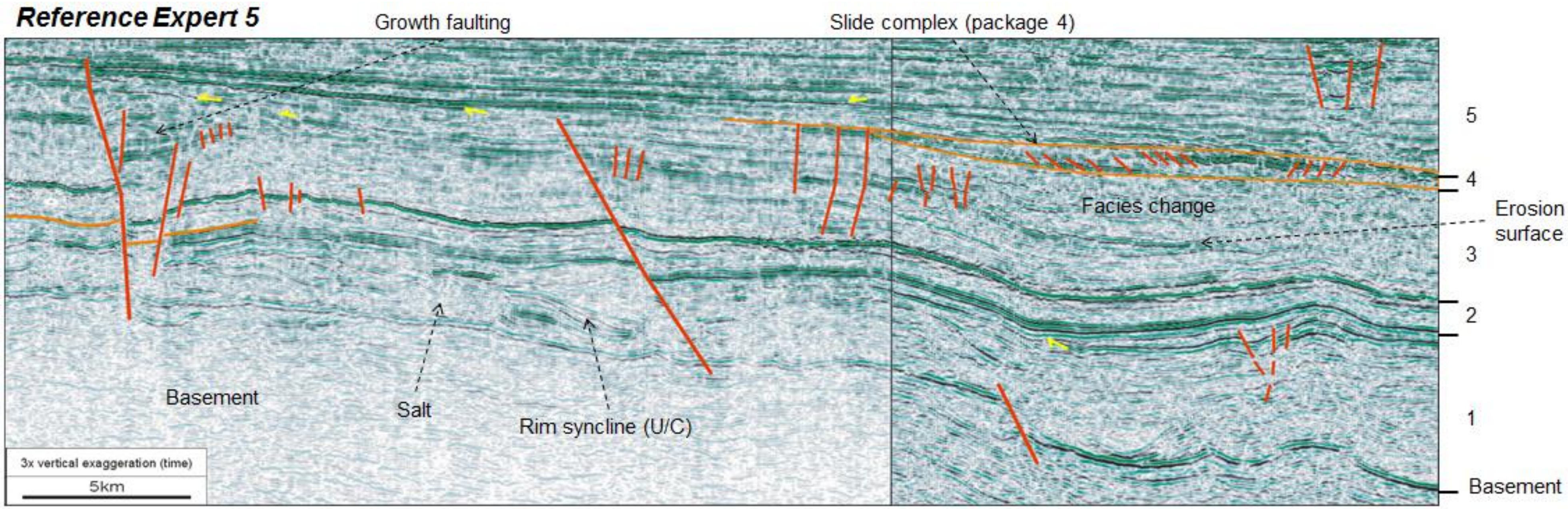
Expert’s interpretation notes

“Interpretation: Area of both thick-skinned and thin-skinned linked extension and contraction. Salt detachment indicated by: 1) progressive, shifting depocentres ‘1’, 2) drape geometry over fault C, and 3) velocity pull-up on time data.

However, top salt very difficult to identify. Most people would pick the dashed teal line so that growth geometries ‘1’ are in suprasalt section. But because I think this is the North Sea, with little acoustic impedance contrast between the Zechstein and Triassic, the solid teal is more likely to be the top evaporites. In this case, salt movement was already active during ongoing evaporite deposition, as known from various salt basins. *Evolution:* 1. Thick-skinned extension (fault C and others) during at least teal to red time, probably also during evaporite time, triggering salt movement. Extension decoupled, with pre-salt extending primarily at fault C and suprasalt extending mostly at fault B ‘2’. 2. Postrift gravity gliding (detached on salt) due to differential thermal and loading subsidence (larger faults to right?) – extension between red and green time on fault B ‘3’, coeval shortening, thinning, and extensional truncation at

‘4’ (fold covered by salt). 3. Ongoing thin-skinned extension between green and yellow, but shifted to fault A ‘3’. Ramping over fault C thickens overlying section ‘6’, coeval shortening probably off-section to right. 4. Minor differential subsidence post-yellow.

Figure 4.5 – Digitised interpretation for reference expert 5 with their explanation notes. The left, middle and right faults are interpreted to be extensional. The left and middle fault cut horizon 1. Expert 5 interpreted a facies change in package 4 (top of their package 3), and a *"slide complex"* on horizon 5.

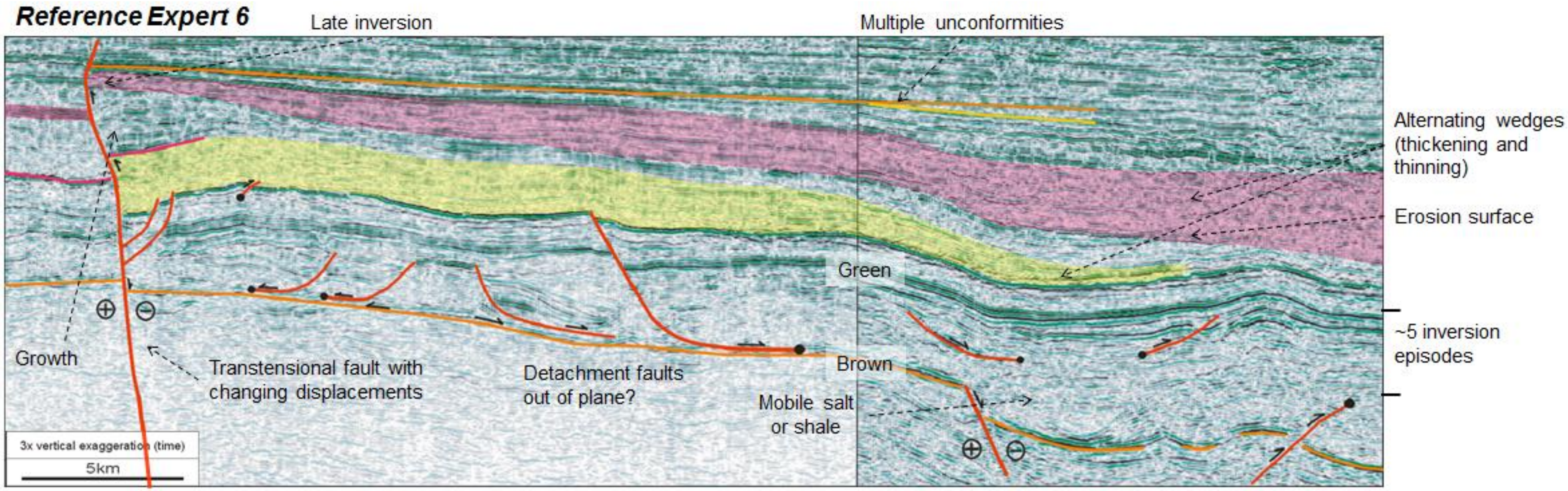


Expert's interpretation notes

"1) Identified key faults/discontinuities first, then packages 1-5. 2) Looked for truncations/onlap of reflectors particularly in upper part. 3) Identified slide complex thrusted toe and possible extension at back.

Packages: 1) Strong basement (top) reflector offset by faults in places – top marked by strong continuous reflector package. Internally lots of unconformities and discontinuities which look salt related (rim synclines) to me. An alternative would be a faulted deepwater channel levee complex but the geometries and contacts do not look right. Overall package thins to left suggesting basin was to right. Possible truncation just below top of 1. Minor faulting (extension) in this unit to right. 2) Twin relatively coherent set of reflectors traceable across section, cut by extensional faults. 3) Package with significant variations in reflector character – suggesting facies changes. Wedge of sediment on left side related to movement of fault, with onlap onto wedge – onlap appears to be conformable in the basin. Numerous extensional faults in this package. Facies change to right could be shale (but very coherent) or a mass transport complex (MTC) – unlikely as does not have well-defined edges. Unconformity truncation at base of this on right. 4) Slide complex which wedges out to left. 5) Packages of overlapping wedges to left, subtle truncation of packages. *Faults:* Major – steep extensional fault on left which cuts through most units – has character of negative flower structure but depends on h vs. v scale. *History of packages:* 1) Salt movement and deposition. 2) Deposition across area. 3) Extension and instability. 4) Slide. 5) Onlap, possible minor fault movement".

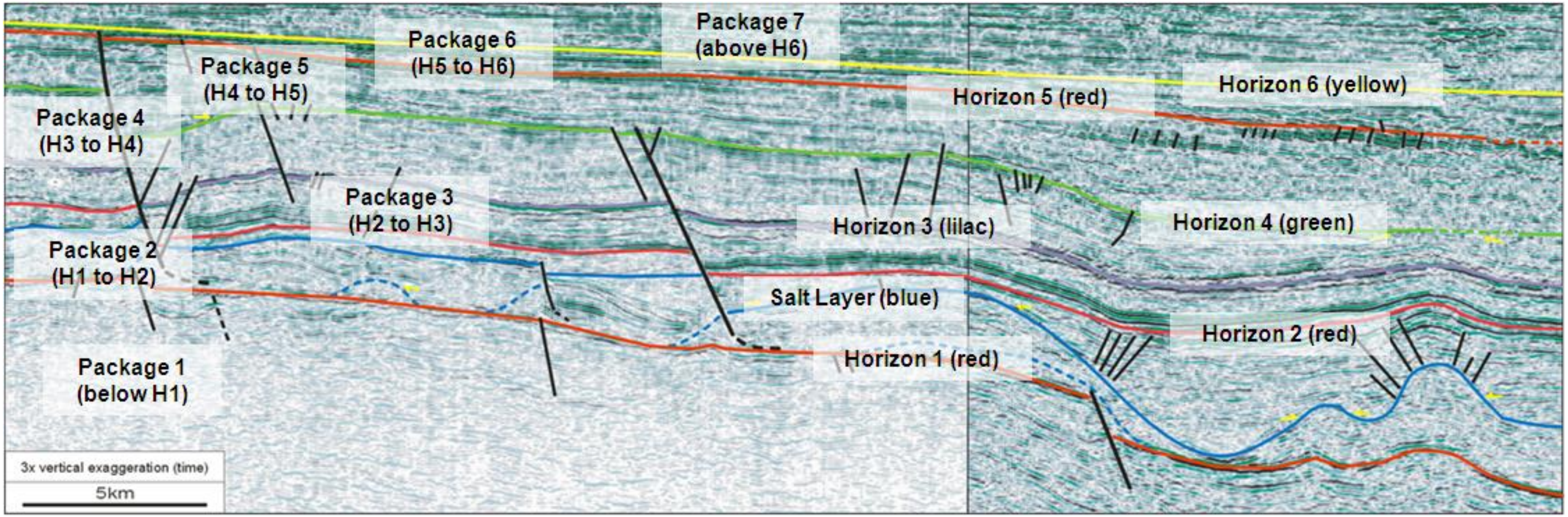
Figure 4.6 – Digitised interpretation for reference expert 6 with their explanation notes. Expert 6 interpreted the left fault to be transtensional. The left fault offsets horizon 1 with the right hand side downthrown, whereas horizon 4 is the left hand side is downthrown. The middle fault is a listric normal fault and detached on horizon 1. Expert 6 identified packages of thickening and thinning, and 5 inversion episodes on the right of package 3. There is a compressional fault offsetting horizon 1.



Expert's interpretation notes

"I did not spend a lot of time on this as developing a valid interpretation requires reference to additional in-line and cross line data. Confidence relates to this rather than to lack of confidence in picking features. The interpretation presented is to develop and note key features and concepts that need to be resolved with reference to additional data. Features that stand out as key to understanding the area: 1) Major long lived fault at left is steep with changing displacements – strongly suggests a major strike-slip component. The thickness alternation across the fault also confirms this. 2) The stratigraphic sequence shows numerous sediment wedges with a mix of erosion and downlap surfaces separating the units. This indicates that the basin is strike-slip influenced. Either we are dealing with a strike-slip basin sensu-stricto or a mix mode basin with a strong trans-tensional component. 3) There is a faulted system that appears to detach on the acoustic basement (brown on the section). This could be a salt or shale related detachment system and this would potentially develop into a valid single line interpretation. Not clear from the seismic if these faults are hard or soft-linked – I'm inclined to think they may be soft-linked. 4) In the lower sedimentary sequence Brown to Green some faults look as if they could be reverse and there are a number of folds which are not obviously normal roll-overs – could be inversion or maybe out of plane rotational components. 5) Reflector quality changes quite a lot across the section and this could be due to a major dip component out of section. Taken together I would look at this being a single line from a transtensional or mixed mode basin system. The section does not include the basin boundary fault and there is no controlling detachment faulting contained within this section. I would also like to see a longer regional line to place this in context of the basin as a whole".

Figure 4.7 – The legend for the seismic image to unify the labelling systems used by the reference experts. The names given in this legend to the horizons, faults and packages will be used from here on.



Legend for the seismic image

This legend shows the main horizons, faults and packages, as picked by the reference experts.

The main horizons are: *"horizons 1 to 6".*
The main packages are: *"packages 1 to 7".*
The main faults are: *"left fault"*, *"middle fault"* and *"right fault"*.
The fold on the right-hand side between horizons 2 and 3 is referred to as the *"fold"*.

'Packages' are defined to be the layers of sediment between reflector horizons.

4.2.1. Similarities and differences between experts' interpretations

It was of great interest to note the similarities and differences between experts' interpretations. For instance, Experts 1 to 5 interpreted the left fault as being extensional; while Expert 6 interpreted it as being transtensional. Experts 1 to 4 interpreted the left fault to be detached on horizon 1, while Expert 5 interpreted it to be deeper, offsetting horizon 1. Having diversity in the reference experts' interpretations was consistent with Shanteau (2000) and O'Hagan *et al.* (2006) who note that it is common for experts to disagree. The reference experts' interpretations were all valid as per my definition ('possible under the laws of geology') based on the given data. The interpretations were also treated equally in the statistical analysis.

Expert 3 was the only expert to interpret horizon 2 as a *"locally angular unconformity"*, while Expert 5 was the only expert not to include a *"detachment"* or *"décollement"* in their interpretation. There was also substantial disagreement in how the right side of horizon 1 should be interpreted. Experts 1 and 4 interpreted the curved right side of horizon 1 to be *"velocity artefacts"* or *"pull-ups"*; while Experts 2, 3 and 6 interpreted the curves to be offsets in the reflector horizon and drew in faults to explain the discontinuity. Expert 5 did not interpret the right side of horizon 1 in their interpretation. The middle fault was interpreted by all reference experts as being extensional. However, Expert 5 interpreted it to be offsetting horizon 1, while the other experts either interpreted an extensional fault that soled-out on the detachment, or stopped above horizon 1. The right fault was interpreted by all experts as being extensional, although Expert 6 interpreted transtensional movement.

The correlation of the right side of horizon 4 (as defined in Figure 4.7) to the left side of the seismic image was an important decision since it determined whether the geometry of the left fault was extensional or compressional. For example, if the right side of horizon 4 was correlated to the left side of horizon 4, the fault would extensional, as in Experts 1 to 5s' interpretations. However, if the right side of horizon 4 was correlated to the left side of horizon 3, the fault would be compressional, as in Expert 6's interpretation.

Each expert had a particular 'style' of interpretation. Expert 1 interpreted the main structural features that they found to be important (e.g. *"updip extensional domains and downdip compressional domains"*) with less emphasis on the smaller features, apart from the *"fold"* on the right side. In contrast, Expert 4 noted many smaller features (e.g. *"shifting depocentres"*), in addition to the larger structural features (left, middle and right faults). Expert 6 also noted that:

> *"The interpretation presented is to develop and note key features and concepts that need to be resolved with reference to additional data".*

Hence, Expert 6 acknowledged the fact that they could only do so much with a single 2D seismic image, and that more data is needed to be able to build a comprehensive interpretation.

4.2.2. Differences in the types of language/terminology used by experts

It was interesting to note the different types of language/terminology used by the experts. Although most of the interpretations were reasonably similar in terms of structural style and kinematics, the language used by the experts to describe their interpretation was markedly different.

- Expert 1 described their interpretation in terms of structural geology, while Expert 5 mainly focussed on sedimentary features and used sedimentological terminology. For example, both experts highlight the importance of the thickening to the right-hand side in package 5 and timing, but use different language. Expert 1 said:

  > *"A late regional tilting action probably activated the listric faults and caused the gravity sliding – timing shown by stratigraphy (yellow). Tilting occurs at time of growth on listric faults".*

  While, Expert 5 said:

  > *"Package with significant variations in reflector character – suggesting facies changes. Wedge of sediment on left side related to movement of fault, with onlap onto wedge – onlap appears to be conformable in the basin".*

- The different descriptions of Horizon 1 were interesting. Noted in no particular order, they include: *"base detachment surface"*, *"detachment* *surface (base salt)"*, *"décollement"*, *"salt detachment"*, *"basement (top) reflector"* and *"acoustic basement"*.

- Expert 1 and Expert 4 used different phrases to describe the same seismic feature, *"velocity artefacts"* and *"pull-up"* (respectively).

It is thought that the type of terminology used by experts was influenced by their area of expertise, but this cannot be further developed without breaching the experts' anonymity.

4.3. Evolutionary diagrams of experts' interpretations

The geological evolutions of reference experts' interpretations were investigated to confirm that their interpretations were valid ('possible under the laws of geology'), and to compare what geological times (packages) were focussed on. All experts described their interpretation's geological evolution to some extent (even though they were not asked explicitly for it). Expert 4 gave three evolutionary steps, while all other experts gave some evolutionary steps mixed in with the summary of their interpretation. It was therefore, likely that 'interpretation' and 'geological evolution' were synonymous to most of the experts as they explained their evolution to some extent as part of their interpretation. I used each expert's interpretation (seismic image and notes) to extract their geological evolution.

I created a sequence of seven diagrams for each expert to show the steps in the geological evolution of the interpretations. Each diagram corresponded in time to the deposition of one of the six packages of sediment or the salt layer (package boundaries on the seismic image are defined in Figure 4.7). Each diagram represents an evolutionary step that occurs at the same geological time for each expert, to allow comparisons to be made. The diagrams are not to scale and the packages are not fully representative in terms of relative depths (thicknesses of packages) and dips (angles of domain boundaries); they only illustrate the main structural activity. For example, the individual packages in the evolutionary diagrams were constructed with horizontal boundaries, which is different from the dipping reflectors in the seismic image. The diagrams allow a much clearer understanding of the major similarities and differences between the experts' evolutions and hence, their interpretations. I did not correspond with the respective experts about the evolutionary diagrams, thus they represent the evolution based on my analysis of their interpretations.

Due to the limited time the experts had to complete the exercise, and the complexity of the exercise itself, each expert naturally emphasised different parts of the geological evolution (the parts that were presumably most important to them). Hence, due to this subjectivity, it was not possible for me to include every evolutionary step for each expert as many of their steps were implicit rather than explicit. For example, Expert 1 gives a broad tectonic summary, only summarising the main tectonic events that occur in package 5, but did not explicitly say that each underlying layer of sediment had been deposited in sequence. However, for the layers of sediment to be displaced by the left fault they had to have been deposited previously. In these instances, I had to infer the evolutionary step and this was noted. I was thus conservative in my assumptions. I have also only used arrows (e.g. to show relative fault movement) when the movement was explicit at the particular time step.

It was not clear when the salt movement occurred for some of the experts. Hence, in these cases, I drew a salt diapir on the right of package 2 to coincide with the main fault activity (mostly package 5) as the extensional faults may have initiated compression, salt activity and doming. It was also not clear when the right fault was active for most experts and, as above, I noted where I inferred a geological action in the evolutionary diagrams.

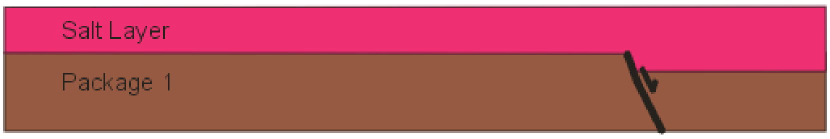
4.3.1. Introducing the evolutionary diagrams

Geological actions were noted in bold, followed by evidence for that action. Evidence comes in the form of direct quotes or in the form of observations of the expert's interpretation. Observations were explicitly noted as observations. For comparative viewing purposes, the evolutionary diagrams have also been reproduced individually at A4 size (see envelope at back of thesis). The six sets of diagrams are presented first and then compared afterwards.

Reference Expert 1

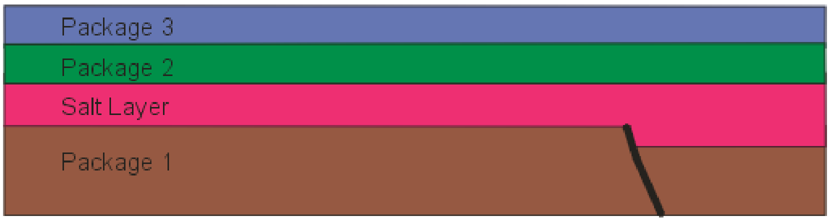
1. Layers of sediment deposited up to top of package 1. (Inferred).

2a. Salt layer deposited on top of package 1. *"There looks to be a significant salt layer (pink) across the seismic image".* **Right fault active.** (Inferred).



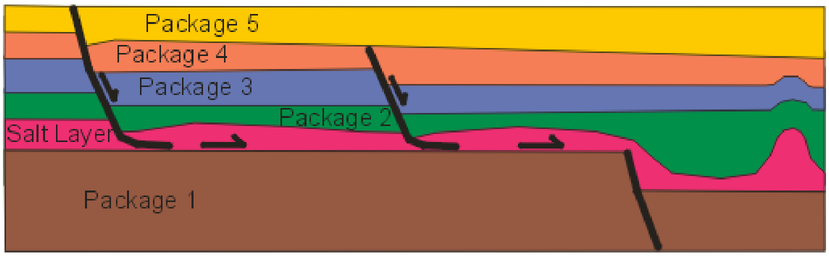
2b. Layers of sediment deposited up to top of package 2. (Inferred).

3. Layers of sediment deposited up to top of package 3. (Inferred).

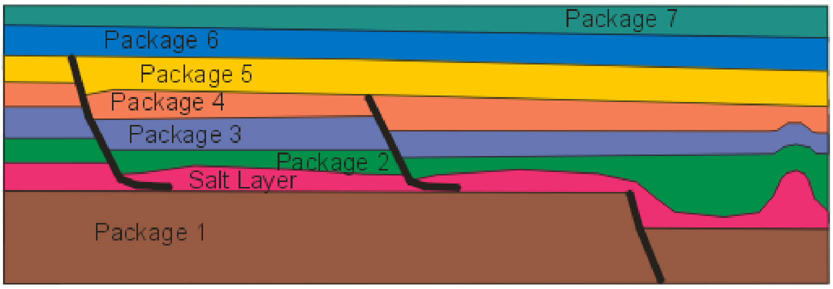


4. Layers of sediment deposited up to top of package 4. (Inferred).

5. Layers of sediment deposited up to top of package 5. (Inferred). **Regional tilting event occurred. Left and middle faults activated and packages 2 to 4 detached on horizon 1 to the right. Growth faulting on the left fault. Fold on right side was generated by downdip compression.** *"There is also a fold structure on the right-hand side. This is the classic recipe for detachment tectonics: updip extensional domains and downdip compressional domains, which we have here. A late regional tilting action probably activated* *the listric faults and caused the gravity sliding – timing shown by stratigraphy (yellow). Tilting occurs at time of growth on listric faults".*



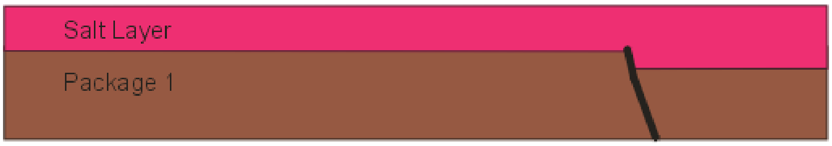
6. Layers of sediment deposited up to top of package 7. Packages 6 and 7 thicken to right due to tilting at previous step. (Inferred).



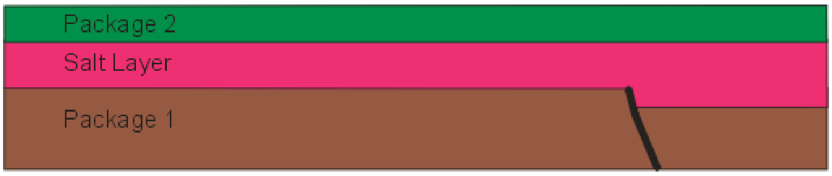
Reference Expert 2

1. Layers of sediment deposited up to top of package 1. (Inferred). **Minor faults active on right of package 1.** Right side of horizon 1: *"Push up structure [is] possibly inverted presalt faults".*

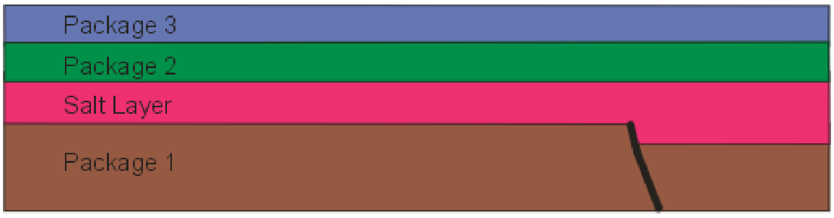
2a. Salt deposited on top of package 1. Horizon 1: *"Base Salt".*



2b. Layers of sediment deposited up to top of package 2. (Inferred).

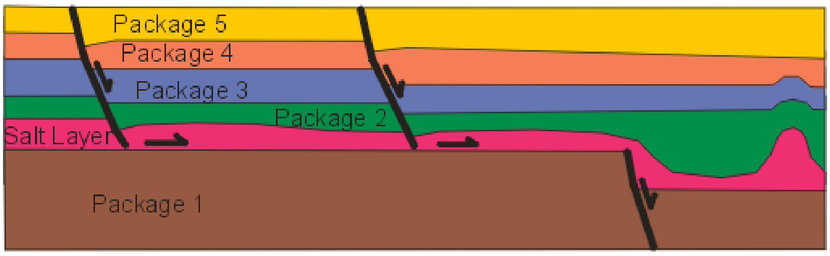


3. Layers of sediment deposited up to top of package 3. (Inferred).

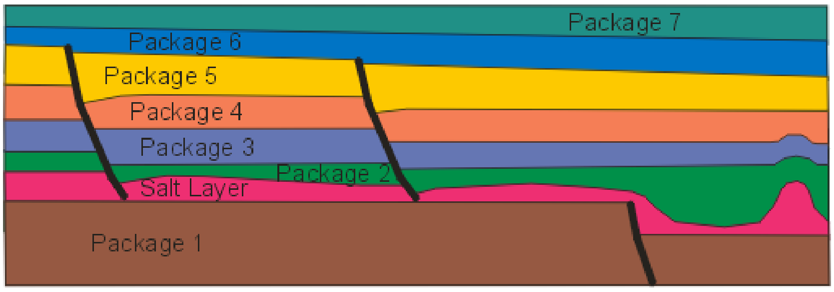


4. Layers of sediment deposited up to top of package 4. (Inferred).

5. Layers of sediment deposited up to top of package 5. (Inferred). **Regional sag tilting event occurred forming a 'steer's head' structure. The sag tilting event activated the left and middle faults, which detached on package 1 to right with folding of packages 3 and 4. Package 5 thickened to right.** *"Regional thickening into the basin (sag-style) evident above this grey package [packages 5 to 7] so basinwards tilting probably triggered salt tectonics (latest phase, updip extension)".* Observation: growth strata have been annotated on horizon 4 of the left fault.



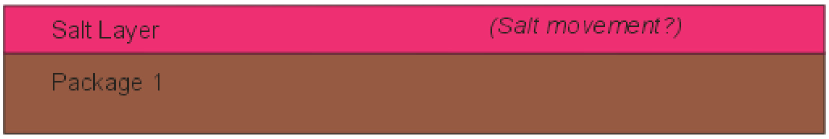
6. Layers of sediment deposited up to top of package 7. Packages 6 and 7 thicken to right due to tilting at previous step. *"Regional thickening into the basin (sag-style) evident above this grey package [packages 5 to 7]".*



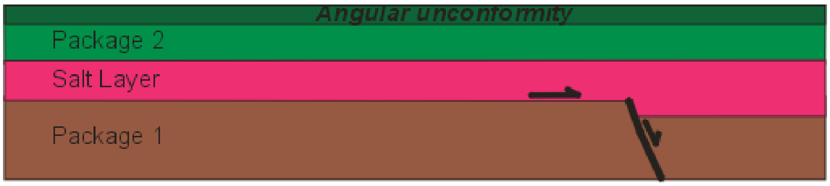
Reference Expert 3

1. Layers of sediment deposited up to top of package 1. (Inferred).

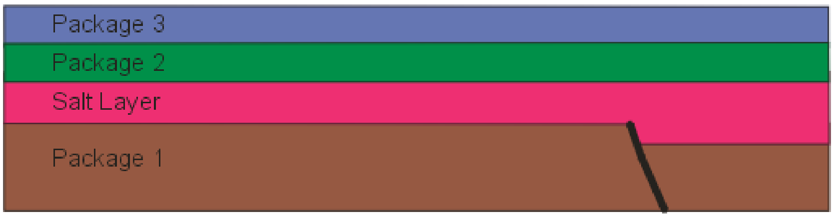
2a. Salt deposited on top of package 1. *"Lower amplitude zones at this level [salt layer] are interpreted to be salt or some other relatively mobile material".*



2b. Layers of sediment deposited up to top of package 2. (Inferred). **Normal faulting; beds in package 2 were deformed (rotated).** *"The first [phase of movement] is indicated by normal faults and steep bed dips between horizons B and C [package 2]".* **Horizon 2 formed an angular unconformity with the beds below.** *"This deformed B-C interval [package 2] is overlain by A-C [packages 3 and 4], deposited in a tectonically quiescent period with C [horizon 2] forming a locally angular unconformity truncating underlying dipping reflectors".*

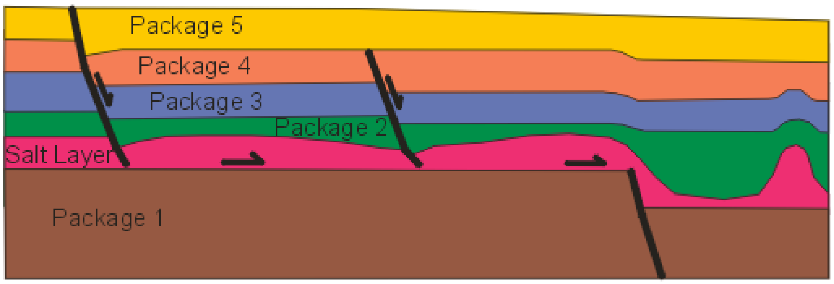


3. Layers of sediment deposited up to top of package 3 in a tectonically quiescent period. *"This deformed B-C interval [package 2] is overlain by A-C [packages 3 and 4], deposited in a tectonically quiescent period with C [horizon 2] forming a locally angular unconformity, truncating underlying dipping reflectors".*

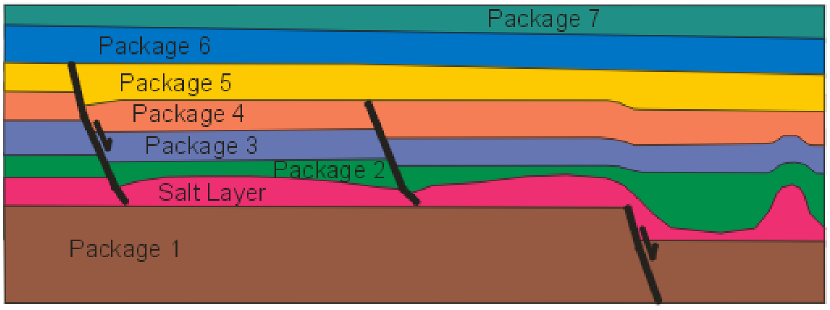


4. Layers of sediment deposited up to top of package 4 in a tectonically quiescent period. (See quote in step 3).

5. Package 5 deposited while left fault was active; growth faulting on left fault and on right monocline. *"Growth strata associated with both normal faulting and folding occur in this interval [package 5]".* Observation: growth strata have been annotated on horizon 4 of the left fault and on the right monocline on horizon 4. **Packages 2 to 4 detached on salt along horizon 1 to right, with coeval folding on the right.** *"A second phase of deformation initiated at A [horizon 4] with extensional faulting on the left of the section and synchronous folding on the right indicating gravitational sliding towards the right on the décollement B [horizon 1]".* **Main deformation occurred in package 5.** *"This deformation occurred mainly in the interval between deposition of A and D [package 5]".* Trigger: *"This second phase of deformation may have been triggered by movement on a normal fault at depth which offsets B [right fault] and thickness variations in A-D [package 5] directly above this fault suggest that it was active at this time".*



6. Layers of sediment deposited up to top of package 7. (Inferred). **Minor movement on left fault and right fault during deposition of package 6.** *"Local truncations (labelled 'X') above horizon D [horizon 5] indicate that this fault [right fault] may have had some activity later than D [horizon 5] as has the normal fault on the left of the section [left fault]".*

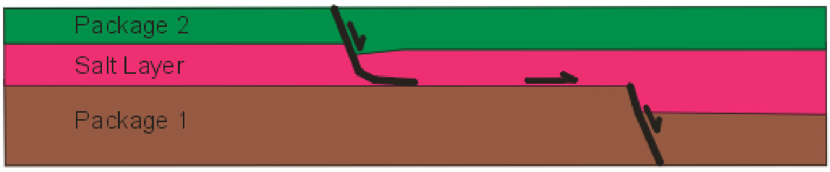


Reference Expert 4

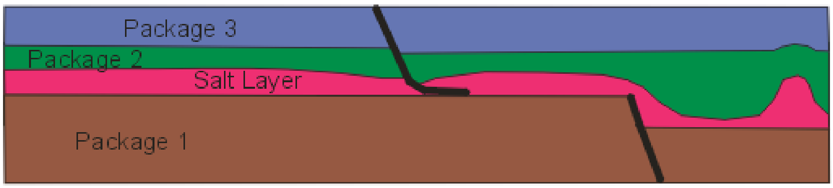
1. Layers of sediment deposited up to top of package 1. (Inferred).

2a. Salt deposited on top of package 1. Package 2: *"Probable top salt"*. **Package 1 extended to right, mainly on right fault.** Observation: salt drape over right fault and depocentres noted. *"Thick skinned extension (fault C and others) [right fault] during, at least, teal-to-red time [top of package 2]. Probably also during evap time, triggering salt movement".*

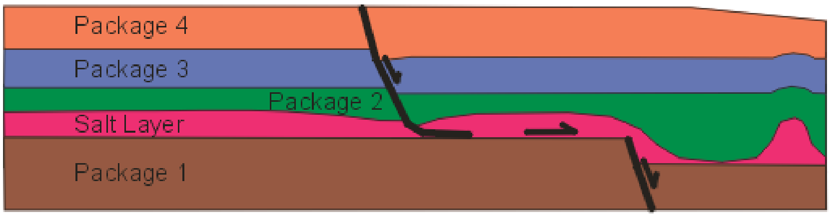
2b. Layers of sediment deposited up to top of package 2. Salt movement as sediment was deposited. *"Salt movement was already active during ongoing evaporite deposition, as known from various salt basins".* **Package 1 extended to right, mainly on the right fault. Package 2 mainly extended to the right on the middle fault, detached on the salt layer on the top of package 1.** *"Thick skinned extension (fault C and others) [right fault] during, at least, teal-to-red time [top of package 2]. Probably also during evap time, triggering salt movement. Extension decoupled with presalt primarily extending at fault C [right fault] and suprasalt extending mostly on fault B [middle fault]".* **Postrift gravity gliding on package 2 due to differential thermal and loading subsidence.** Package 2: *"Postrift gravity gliding (detached on salt) due to differential thermal and loading subsidence (larger faults to right?)"*



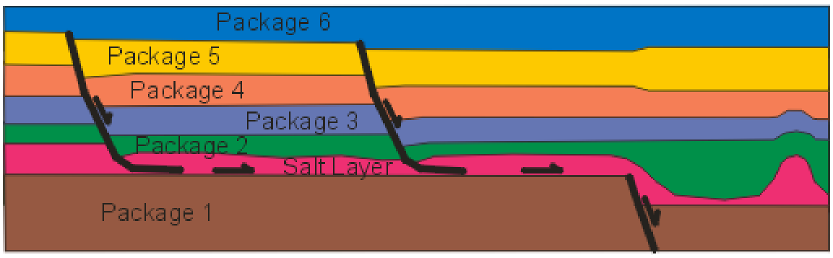
3 + 4. Layers of sediment deposited up to top of package 3. Layers of sediment deposited up to top of package 4. Middle fault active with extension to right on packages 3 and 4, detached along top of package 1. Coeval shortening on right; thinning and erosional truncation over fold. *"Extension between red and green time [packages 3 and 4] on fault B [middle fault], coeval shortening, thinning and erosional truncation".* Observation: toplap arrows drawn underneath right side of horizon 4.



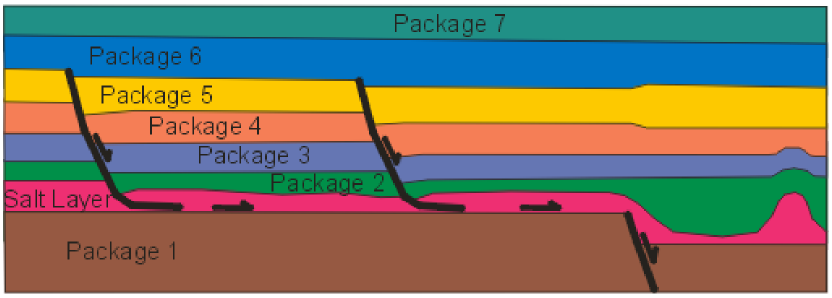
4. See above step.



5. Layers of sediment deposited up to top of package 6. Ongoing extension during deposition of package 5, but main activity shifted from middle fault to left fault. Right of package 5 thickened due to ramping over right fault. Package 6 thickened to right. Further folding on right. *"Ongoing thin-skinned extension between green and yellow [horizons 4 to 6], but shifted to fault A [left fault]. Ramping over fault C [right fault] thickens overlying section, coeval shortening probably off-section to right".* (Trigger for fault activity not mentioned).



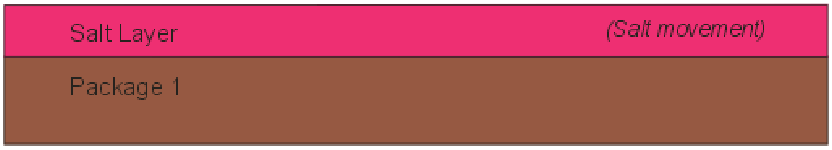
6. Layers of sediment deposited up to top of package 7. Package 7 thickened to right. Minor subsidence with continued infill. *“Minor differential subsidence post yellow [packages 6 and 7]”.*



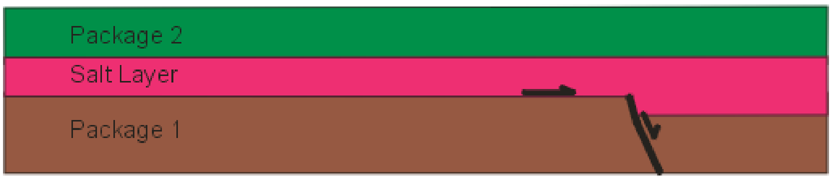
Reference Expert 5

1. Layers of sediment deposited up to top of package 1. Observation: package 1 labelled as *“Basement”.*

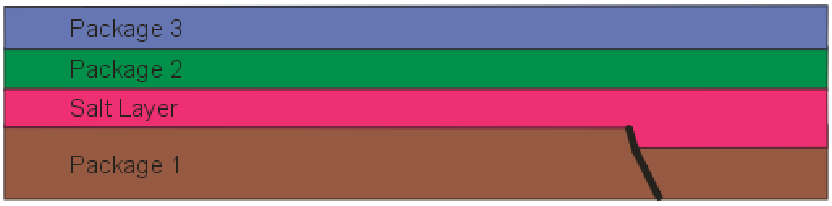
2a + 2b. Salt deposited on top of package 1. Layers of sediment deposited up to top of package 2. Salt movement as sediment was deposited. Salt layer: *“Internally lots of unconformities and discontinuities which look salt-related (rim synclines)”*. Observation: salt layer has *“Salt?”* annotated on it.



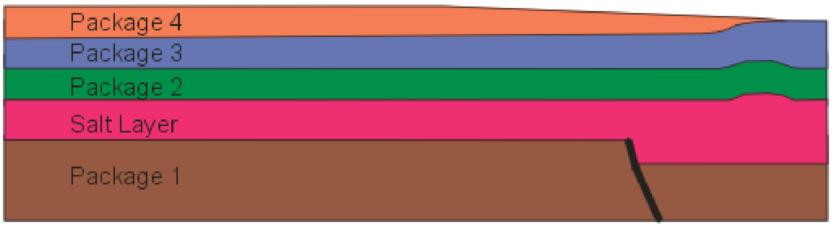
2b. See above step.



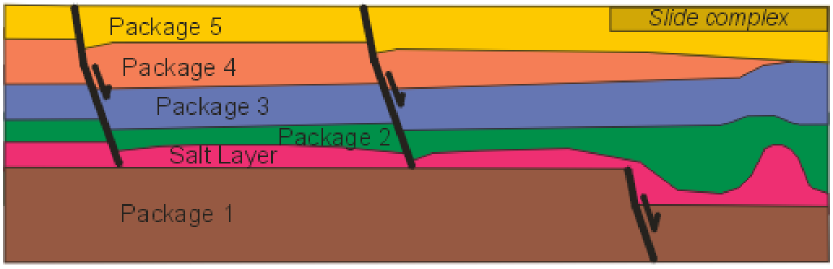
3. Layers of sediment deposited up to top of package 3. Package 3: *"Deposition across area".*



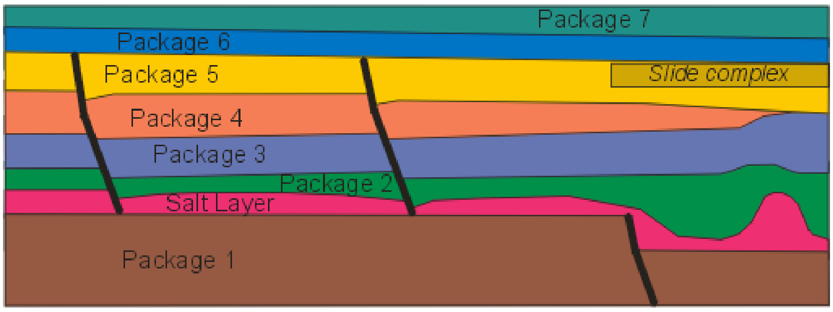
4. Layers of sediment deposited up to top of package 4. (Inferred). **An erosion surface formed at right of horizon 4.** Observation: right side of horizon 4 has *"erosion surface"* annotated.



5. Layers of sediment deposited up to top of package 5. (Inferred). **Left and middle faults were active during deposition of package 5 (growth faulting).** Packages 4 and 5: *"Extension and instability".* Observation: package 5 has *"growth faulting"* annotated on the left fault. **Facies change in package 5 on right.** Package 5: *"Facies change".* **A slide complex at top of package 5 on the right.** Observation: *"slide complex"* was annotated at the top of package 5 on the right. (Trigger for fault activity not mentioned).

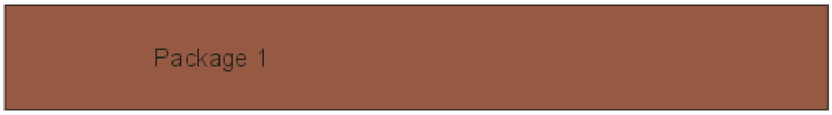


6. Layers of sediment deposited up to top of package 7. (Inferred). **Packages 6 and 7 thicken to right with onlaps.** Packages 6 and 7: *"Package of onlapping wedges to left, subtle truncation of packages".*

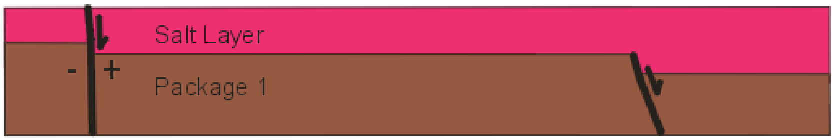


Reference Expert 6

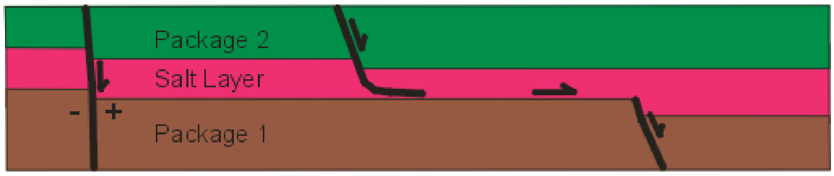
1. Layers of sediment deposited up to top of package 1. (Inferred).



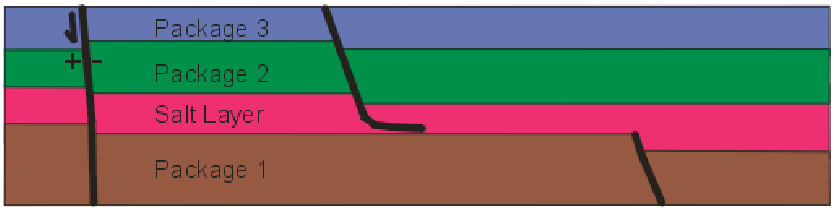
2a. Salt deposited on top of package 1. Package 2: *"Salt or shale in this section".* **Left fault was active (strike-slip) causing offset on horizon 1.** Observation: the left of horizon 1 went up relative to the right.



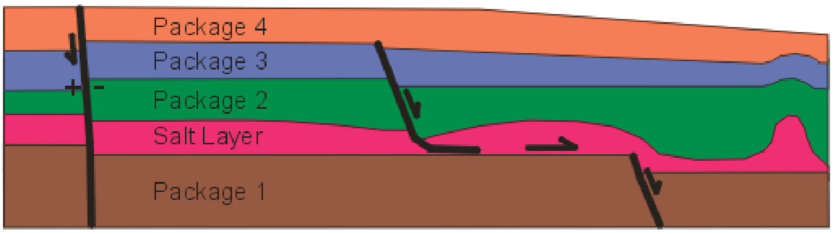
2b. Layers of sediment deposited up to top of package 2. Salt movement as sediment was deposited. Observation: offset in horizon 1 due to strike-slip movement on left fault. **Middle fault was active, detaching to right on top of package 1.** Observation: *"detachment"* annotated on horizon 1; fault geometries drawn in.



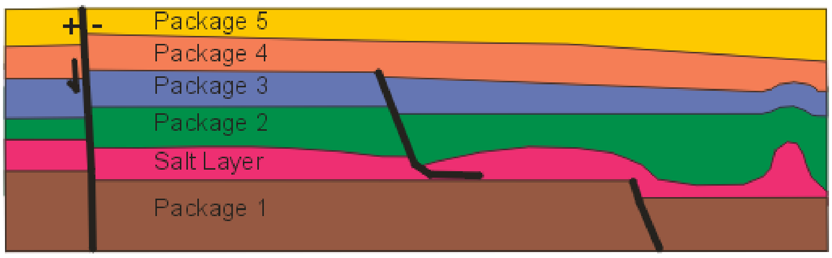
3. Layers of sediment deposited up to top of package 3. (Inferred). **Strike-slip movement on left fault changed direction (polarity change).** Observation: the right side of horizon 3 has gone up relative to the left side.



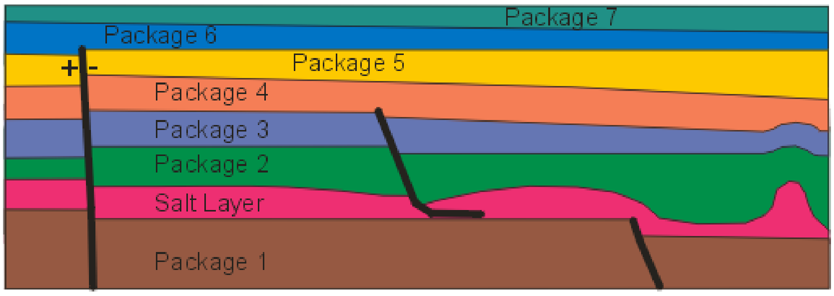
4. Layers of sediment deposited up to top of package 4. (Inferred). **Continued strike-slip movement of the left fault with same polarity.** Observation: horizon 4 has the same relative side offset as horizon 3 (the right side goes up relative to the left side). **Package 4 thinned to the right.** Observation: *"thin"*, with arrow to the right annotated on horizon 4. **Fold on right was generated by compression.** Observation: fault geometries drawn in.



5. Layers of sediment deposited up to top of package 5. (Inferred). **Continued strike-slip movement of the left fault with same polarity.** Observation: horizon 5 has the same relative side offset as horizon 4 (the right goes up relative to the left). **Package 5 was thinning to left on an erosion surface.** Observation: *"thin"* with arrow to the left side annotated in package 5. Observation: *"alternating wedges"* annotated with cartoon showing alternating wedges.



6. Layers of sediment deposited up to top of package 7. (Inferred).



4.3.2. Comparing the evolutionary diagrams

Table 4.2 shows seven geological features in experts' interpretations, or in the evolutionary diagrams, and cross-tabulates which expert(s) interpreted which feature. I chose the seven features to illustrate the variation between experts' interpretations and the evolutionary diagrams. The seven features are labelled 'a' to 'g'. It can be seen from Table 4.2 that the experts did not all interpret the same features in the seismic image. For example, Expert 5 did not interpret the detachment on horizon 1 (feature a), Experts 1 to 3 did not interpret the erosion surface to the right of horizon 4 (feature c), and only Expert 3 interpreted an angular unconformity on horizon 2 (feature g).

Table 4.2 – Cross-tabulation of the seven features (a to g) showing which expert(s) picked each feature. Direct quotes from experts' interpretations, or associated notes, are noted in quotation marks.

	Expert 1	**Expert 2**	**Expert 3**	**Expert 4**	**Expert 5**	**Expert 6**
a. Detachment on horizon 1	Yes	Yes	Yes	Yes	No	Yes
b. Reason for left and middle fault activity in package 5	*"Regional tilting action"*	*"Regional sag tilting"*	*"Movement on a normal fault at depth which offsets B [right fault]"*	No reason given	No reason given	No reason given
c. Erosion surface on right of horizon 4	No	No	No	Toplap arrows drawn underneath right side of horizon 4	*"Erosion surface?"*	*"Erosion surface"*
d. Monocline on horizon 4 above right	No	No	Growth strata have been	*"Ramping over fault C [right fault]*	No	No

fault			annotated on the right of horizon 4	*thickens overlying section"*		
e. Small faults on RHS of horizon 5	No	*"Polygonal faults"*	No	Small faults interpreted but not annotated	*"Slide complex"*	No
f. Interpretation of right of horizon 1	*"Velocity artefacts"*	Faults interpreted	Faults interpreted	*"Pull-up"*	Not interpreted	Faults interpreted
g. Angular unconformity on horizon 2	No	No	Yes	No	No	No

The evolutionary diagrams show that 4 of the 6 experts (Experts 1, 2, 3, and 5) had similar geological evolutions with only subtle differences. Each of these experts interpreted early movement on the right fault (diagram 2b) with continued sedimentation and no major tectonic action until the deposition of package 5, when the left and middle faults became active, acting on packages 2 to 4. The final evolutionary step shows the deposition of packages 6 and 7 with minor thickening to the right side.

Expert 4's geological evolution was similar to the four experts above, but included early movement on the middle fault (at the time of package 2), preceding any activity on the left fault. (The above four experts inferred or said that the left and middle faults became active at the same geological time). Expert 6, however, gave the most different interpretation and evolution compared to the other experts. The evolutionary diagrams for Expert 6 show that the left fault was strike-slip and continuously active throughout packages 1 to 7. The left fault also had a polarity change during the deposition of package 3. The middle fault was listric and detached on salt along the horizon 1 and not active after package 3. Alternating wedges were produced on the right of the section (thickening and thinning) in packages 4 to 7.

Three of the experts (Experts 1, 2 and 4) noted that the 'thickening to the right' in packages 4 to 7 was important, but it is unclear whether they rated the thickening as being as important as Expert 6 did. Expert 6 noted that the thickening indicated that the basin was strike-slip influenced:

"The stratigraphic sequence shows numerous sediment wedges with a mix of erosion and downlap surfaces separating the units. This indicates that the basin is strike-slip influenced".

All experts interpreted salt, but did not agree on the importance of the salt movement during the geological evolution. Expert 1 said that the salt played a major part in the geological evolution, while Expert 6 said that it played a minor part (pers. comm.).

4.4. Using key features to assess respondents' interpretations

Key features in each expert's interpretation were defined in order to form a scoring system by which respondents' interpretations were assessed with a similarity score. As in chapter 2, 'respondents' are defined to be those geoscientists who took part in the Freyja questionnaire survey. Each expert was asked to pick between 3 and 8 key geological features that were integral to their interpretation. To make the comparison between the respondents and experts as objective as possible, I asked each expert to choose their own key features (instead of me choosing the key features based on their interpretations). The instructions the reference experts received were as follows:

"Identify and list between 3 and 8 key geological features that are essential to your interpretation on the back of the A3 sheet. The term 'features' is meant to be vague, but can include different types of faults, horizons, sedimentary packages, folds, etc. Feel free to add other types of features. (These are the features that help define the tectonic setting and/or stratigraphic setting of your interpretation)."

The instructions introduced an element of subjectivity since my definition of 'key features' was purposely vague. However, I considered it important not to influence the reference experts' choice of key features. For example, if I had further defined what 'key features' were meant to be, then I would have biased the experts' choice of their key features. By being vague, I allowed experts to choose the geological features that were most important to them. I also stated in the instructions that other 'types' of features could be added. However, Expert 5 said that 'key features', as I had defined them, could not adequately represent their interpretation of the seismic image:

> *"There is no specific key feature that I used, it is the reflector continuity/discontinuity, seismic facies, reflector termination style, thickness and facies changes between prominent reflectors. My approach is more generic than related to a specific fault or reflector; it's more related to changes in reflection types between prominent reflectors. So I guess my key features would be all the prominent reflectors that can be traced across the line and then changes in thickness and character of the reflectors in between. By tracing the reflectors and seismic packages out it is possible to then define the structural characteristics".*

Expert 5 mostly used sedimentological language in their interpretation and thus, their approach was likely to be the most different of all the experts. However, their resulting interpretation and the evolutionary steps were still similar to what Experts 1 to 4 produced. In contrast, Expert 6 used a similar approach to Experts 1 to 4 and produced a tectonically different interpretation (strike-slip compared to extension). Since Expert 5 could not provide key features for their interpretation, I used key features for five of the six experts (Experts 1 to 4, and Expert 6).

4.4.1. What key features did the reference experts choose?

Table 4.3 to Table 4.7 shows the key features that the five reference experts picked to represent their respective interpretation. The experts each picked between 6 and 9 key features. Logical operators (e.g. OR, AND) were used in three of the experts' sets of key features (these are noted in capitals and bold within the tables).

Logical operators were used to ease the input of the experts' key features. For example, Expert 2 had *"local packages that show unusual thickening patterns (yellow, green and grey)"*, see Figure 4.2, and I broke this down into the following two key features. The first key feature was growth faulting on the left or right of horizon 4 (grey and yellow packages, respectively); while the second key feature was an 'earlier phase of movement' in package 2 (marked as the green packages in Expert 2's interpretation. As I explain below, Expert 4's key features were based on *"common timing"* of two other local features and hence, the 'AND' logical operator is needed.

The ‘acceptance criteria’ for each key feature was a list of possible criteria that I needed to see at least one instance of on respondents’ interpretations to decide that the respondent had interpreted that key feature. The number of respondents from the main dataset who interpreted each feature is noted in the right-hand column.

Table 4.3 – The seven key features selected by Expert 1. The most interpreted key feature was the middle fault (68.9%), and the least interpreted key feature was the regional tilt (0.2%).

Key features picked by Expert 1	Acceptance criteria for feature	No. of respondents
Detached horizon 1	Detachment along horizon 1; explained, annotated or horizontal arrow drawn. E.g. *“detachment”*, *“décollement”*, *“sliding”* and *“decoupling”.*	53 (11.9%)
Salt/shale in package 2	Salt or shale, in package 2 only; explained, annotated or via symbols (^ ^ ^). E.g. *“salt”*, *“shale”*, *“roller”*, *“dome”*, *“pillows”*, *“salt weld”*, *“anhydrite”*, *“ductile unit”*, *“salt withdrawal”*, *“turtle-back”* and *“evaporite”.*	82 (18.5%)
Left fault: normal, downthrown to right (not cutting horizon 1)	Normal to right (not cutting horizon 1); explained or drawn. Fault does not cut horizon 1.	187 (42.1%)
Middle fault: normal, downthrown to right (not cutting horizon 1)	Normal to right (not cutting horizon 1); explained or drawn. Fault does not cut horizon 1.	306 (68.9%)
Compression on right in packages 2 or 3	Compression in package 2 or 3; reverse faults or fold, explained or annotated or via symbols (<|>). E.g. *“anticline”*, *“fold”* and *“four-way dip closure”.*	87 (19.6%)
Regional slope	Indication of there being a regional slope anywhere in the interpretation; explained or drawn (“dip ---->”). E.g. *“basin”*, *“basinwards”*, *“sediment flow”*, *“regional slope”*, *“thickening to centre of basin”*, *“down-slope transport”.*	55 (12.4%)
Regional tilt	Indication of a tilting action at a specific geological time. Distinct from ‘regional slope’ key feature as the tilting action must be clear.	1 (0.2%)

Table 4.4 – The nine key features selected by Expert 2. Four of the key features were the same as Expert 1. The most interpreted key feature was the middle fault (68.9%), and the least interpreted feature was the polygonal faults (2.3%).

Key features picked by Expert 2	Acceptance criteria for feature	No. of respondents
Detached horizon 1	As in Expert 1.	53 (11.9%)
Salt/shale in package 2	As in Expert 1.	82 (18.5%)
Left fault: normal, downthrown to right (not cutting horizon 1)	As in Expert 1.	187 (42.1%)
Middle fault: normal, downthrown to right (not cutting horizon 1)	As in Expert 1.	306 (68.9%)
Compression on right in packages 1, 2 or 3 **OR** pop-up structure	Compression in package 1 or 2 on the right side; reverse faults or fold, explained or annotated or via symbols (<|>) **OR** a 'pop-up' structure on horizon 1. E.g. *"anticline"*, *"fold"* and *"four-way dip closure"*. Includes reverse faults acting on horizon 1.	141 (31.8%)
Growth package (left side **OR** right side)	Explained, annotated or drawn onlap arrows showing (not just highlighted a horizon). Either on left side **OR** right side of horizon 4. E.g. *"syn-rift"* and *"growth faulting".*	76 (17.1%)
Multiple faults on right of horizon 1, downthrown to right	Below horizon 2, explained or annotated; multiple faults in horizon 1, including the right fault.	240 (54.1%)
Earlier phase of movement ('local packages' in package 2)	Explained or annotated in package 2 E.g. *"first phase of movement"* and *"earlier movement".*	23 (5.2%)
Polygonal faults on horizon 5	Explained or annotated on right of horizon 5, e.g. *"polygonal faults"*. Not just small faults interpreted.	10 (2.3%)

Table 4.5 – The eight key features selected by Expert 3. Five key features were the same as Expert 1 and one feature was the same as for Expert 2. The most interpreted key feature was the right fault (74.3%). The least interpreted feature was the unconformity on horizon 2 (2.3%).

Key features picked by Expert 3	Acceptance criteria for feature	No. of respondents
Detached horizon 1	As in Expert 1.	53 (11.9%)
Salt/shale in package 2	As in Expert 1.	82 (18.5%)
Left fault: normal, downthrown to right (not cutting horizon 1)	As in Expert 1.	187 (42.1%)
Middle fault: normal, downthrown to right (not cutting horizon 1)	As in Expert 1.	306 (68.9%)
Compression on right in packages 2 or 3	As in Expert 1.	87 (19.6%)
Growth package (left side **OR** right side)	As in Expert 2.	76 (17.1%)
Right fault: normal, downthrown to right	Drawn, written or annotated.	330 (74.3%)
Unconformity on horizon 2	Noted unconformity or squiggly line drawn (~~~~~~) on horizon 2 only. E.g. *"unconformity"* and *"U/C"*.	10 (2.3%)

Table 4.6 – The six key features selected by Expert 4. Only one key feature was the same as the other experts' key features. Three of the key features were based on the 'common timing' of features and therefore use the AND logical operator. The most interpreted feature was salt/shale (18.5%) and the least interpreted key feature was 'salt layer to horizon 2 extension at middle fault AND thickening over drape fold at right fault' (0.0%).

Key features picked by Expert 4	Acceptance criteria for feature	No. of respondents
Salt/shale in package 2	As in Expert 1.	82 (18.5%)
Growth package (left side **AND** right side)	Explained, annotated or drawn onlap arrows showing (not just highlighted a horizon). On left side **AND** right side of horizon 4.	21 (4.7%)

	E.g. *"syn-rift"* and *"growth faulting".*	
Drape of salt layer over right fault	Continuous horizon annotated not being cut by right fault; drawn, explained or annotated. E.g. *"drape over Basement fault".*	17 (3.8%)
Salt layer to horizon 2 extension at middle fault **AND** thickening over drape fold at right fault	'Extension' drawn, explained or noted on the hanging-wall of the middle fault, between salt layer and horizon 2 (specifically) **AND** thickening of drape fold over right fault; drawn, explained or noted (e.g. *"thickening"*).	0 (0.0%)
Horizon 2 to horizon 4 extension at middle fault **AND** erosion surface above right fold	'Extension' drawn, explained or noted on the hanging-wall of the middle fault, between horizon 2 and horizon 4 (specifically) **AND** 'erosion surface' above fold drawn, explained or noted (e.g. onlap/toplap arrows on horizon 4 indicating an erosional truncation).	14 (3.2%)
Velocity pull-up	Right of horizon 1; annotated or explained. E.g. *"pull-up"*, *"multiple"* and *"seismic artefact".*	11 (2.5%)

Table 4.7 – The seven key features selected by Expert 6. Two features were the same for Expert 1 and one key feature was the same as for Expert 2. The most interpreted feature was the middle fault (68.9%) and the least interpreted key feature was 2+ unconformities (2.0%).

Key features picked by Expert 6	Acceptance criteria for feature	No. of respondents
Detached horizon 1	As in Expert 1.	53 (11.9%)
Middle fault: normal, downthrown to right (not cutting horizon 1)	As in Expert 1.	306 (68.9%)
Compression on right in packages 1, 2 or 3 **OR** pop-up structure	As in Expert 2.	137 (30.9%)
Left fault: strike-slip on (cutting horizon 1)	Drawn, explained, noted or symbols (+, -). Drawn a 'strike-slip' looking fault: flower structure: base is steep with curved faults joining near the top.	73 (16.4%)

	E.g. *"strike-slip"* and *"flower structure".*	
Sediment wedge (thickening and thinning)	Drawn or noted thickening and thinning wedges.	109 (24.5%)
2+ unconformities	2+ unconformities not including and above horizon 1; explained, annotated or drawn squiggly lines. E.g. *"unconformity"* and *"U/C".*	9 (2.0%)
Inversion on right of package 2	Explained, annotated or inversion arrow drawn.	10 (2.3%)

4.4.2. Comparison of key features from reference experts

Experts 1 to 3 and Expert 6 largely focussed on the large-scale structure, while Expert 4 focussed on the geological timing of their interpretation. Twenty-three unique key features were chosen by the experts; 7 key features were picked by multiple experts (Table 4.8), while 16 were chosen by a single expert.

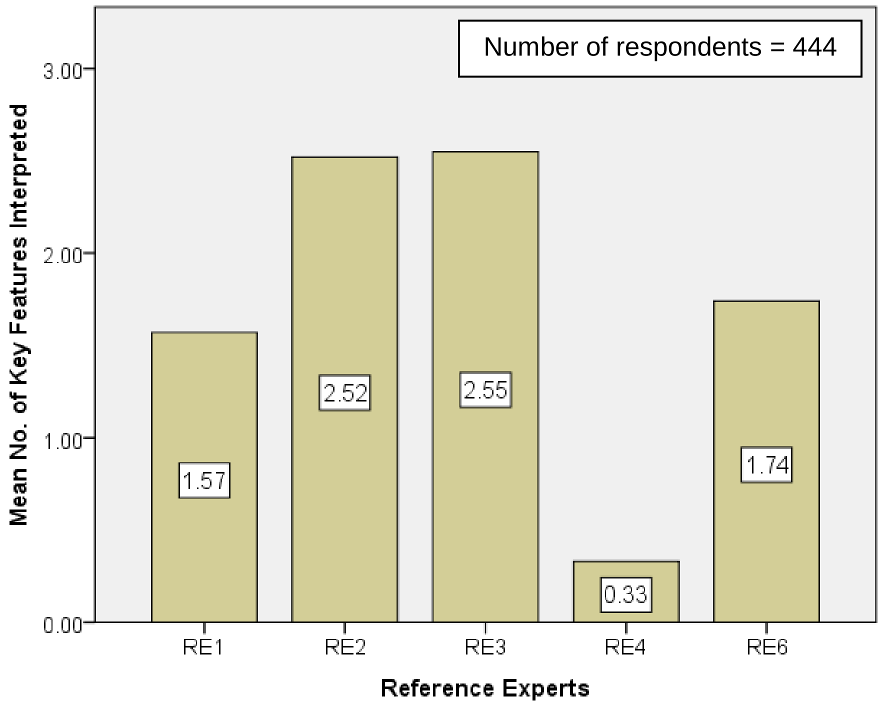
Table 4.8 – Key features that were chosen by six reference experts. Three key features were chosen by 4 of the 5 experts, and Experts 1 to 3 chose three of the same key features.

Name of Feature	Which experts?	No. of experts that chose feature
Detached horizon 1	1, 2, 3 and 6	4
Middle fault: normal, downthrown to right (not cutting horizon 1)	1, 2, 3 and 6	4
Salt/shale in package 2	1, 2, 3 and 4	4
Left fault: normal, downthrown to right (not cutting horizon 1)	1, 2 and 3	3
Growth package (left side **OR** right side)	2 and 3	2
Compression on right in packages 2 or 3	1 and 3	2
Compression on right in packages 1, 2 or 3 **OR** pop-up structure	2 and 6	2

Figure 4.8 shows the mean numbers of key features interpreted by the 444 respondents for the five reference experts. The means for Experts 2 and 3 were 2.52 and 2.55 features, respectively. The means for Experts 1 and 6 were 1.57 and 1.74, respectively; while the mean number of key features for Expert 4 was 0.33

features. Respondents were more likely to interpret the key features that Experts 2 and 3 had chosen, than the key features that Expert 4 had chosen. The maximum possible score was 9 key features, although no respondents achieved this.

Figure 4.8 – Mean numbers of key features interpreted by the 444 respondents for each reference expert. On average, the respondents only interpreted 0.33 of Expert 4's key features. The mean number of key features interpreted for the other experts ranged from 1.57 to 2.55.



The most likely explanation for the lower mean value for Expert 4 was that their key features were stricter than the other experts' features were, since they focussed on the timing of the geological actions (e.g. *"common timing of"*):

> *"1. Stratal geometries between dark blue [horizon 1] and teal [salt layer]*
>
> *2. Drape of teal horizon [salt layer] over fault C [right fault]*
>
> *3. Common timing of teal-red [salt layer to horizon 2] extension at fault B [middle fault] and thickening over drape fold at fault C [right fault] = decoupling*
>
> *4. Common timing of red-green [horizon 2 to horizon 4] extension on fault B [middle fault] and uplift/truncation over most distal structure [fold]*

5. Common timing of green-orange [horizon 4 to horizon 5] extension on fault A [left fault] and thickening over fault C [right fault] = translation over ramp
6. Velocity pull-up".

To simplify the data input, I used the logical operator 'AND' to break down three of Expert 4's key features into their two component parts (*"common timing of"*). Expert 4's feature 3 was not the same as the 'detached basement' key feature as it referred to the right-hand side of the seismic image only. I used the logical operator 'OR' for Experts 2, 3 and 6.

The key features chosen by the reference experts represent their interpretations of the seismic image. It is important to emphasise that the scoring metric was the same for all respondents (see next section), and hence it did not matter that some of the key features were more difficult to identify than others. Hence, no respondents were disadvantaged by the key features that the experts chose.

4.5. Scoring respondents' interpretations via the key features

The sets of key features, selected by the reference experts, represented their interpretations. I used the key features to define a similarity score. Higher similarity scores would imply that respondents were 'more similar' to the reference experts' interpretations, which was assumed to be a desirable quality. For each respondent, I had to decide whether or not they had interpreted each key feature. I thus had a similarity score from each expert, for each respondent. Each similarity score was simply the number of key features that had been interpreted from each expert's interpretation. The central three options to combine the multiple similarity scores were:

1. *Use the most common key features to build a composite similarity score.*
   Under this scenario I would only use the key features that at least two reference experts had identified (Table 4.8) and ignore all other key features. One advantage of this option was that the uncertainty in choosing only one interpretation of a dataset would be mitigated. However, there are issues with combing different reference experts' interpretations (e.g. extension with strike-slip). For example, if respondent X was confident that the structural style of the seismic image was strike-slip, they would be likely to score highly for Expert 6, since Expert 6 interpreted the left-hand side fault and right fault

as being strike-slip. However, Respondent X would also score low for the other four experts, reducing their overall similarity score. Hence, this option would systematically disadvantage the respondents who interpreted the seismic image as being strike-slip because most of the reference experts interpreted it as being extension. This option would have been more suitable if the reference experts' interpretations had been tectonically similar.

2. *Use the mean number of key features interpreted as the similarity score.*
   Under this option, I would take the mean number of key features interpreted to be the similarity score. However, there are again issues with combining different reference experts' interpretations, as above.

3. *Use the maximum number of key features interpreted as the similarity score.*
   Under this option, I would take the maximum number of key features interpreted to be the similarity score. This option would not combine the experts' interpretations, which was an advantage since their interpretations were tectonically different.

I therefore chose option 3 (the maximum number of key features interpreted) since options 1 and 2 would disadvantage respondents who interpreted the seismic image as being strike-slip. Hence, the maximum similarity score achieved by each respondent was therefore taken to be their score in the statistical analysis. The maximum similarity score is referred to as the 'Max RE Score'. As a check, after completing the analysis, I repeated the analysis using scoring option 2, the mean number of key features interpreted. I observed that the final results were almost the same, which implied that the chosen scoring option did not considerably affect the results.

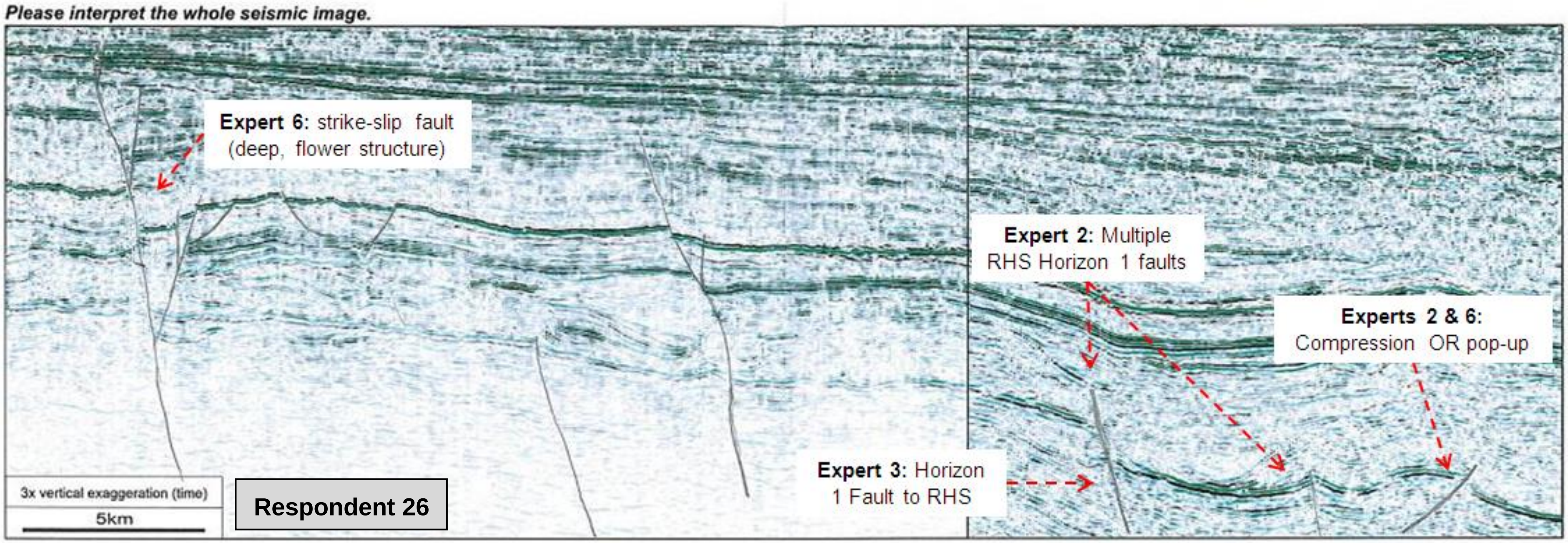
To score respondents' interpretations, I noted evidence of key features to determine the Max RE Score. Thus, only a part of respondents' interpretations were used to determine their score, e.g. whether or not respondents had noted 'carbonate reefs' and/or marked the most likely oil prospects did not influence their Max RE Score. I also did not penalise respondents if parts of their interpretation looked implausible, e.g. 'negative marking' was not used. It would be nonsensical to ignore 'implausible parts' of an interpretation in industry. However, in the context of this research, it was fair not to negatively mark respondents' interpretations to keep the scoring system as simple as possible. For example, the scoring system might become overcomplicated if negative marking had been used, and would then have its own weaknesses such as allowing negative scores. More importantly, the Max RE Score represented a measure of 'similarity' between respondents and reference experts, rather than a measure of accuracy of interpretations since the true 'answer' is unknown.

In this analysis, it only mattered that respondents had at least one instance of each key feature. Respondents' interpretations often had multiple bits of evidence for the same key feature but in these cases, respondents were only given one point. For example, a respondent might have annotated *"salt"* and *"shale"* in different parts of package 2. Both *"salt"* and *"shale"* were individually enough to get the similarity point for Experts 1, 2, 3 and 4; but if both were mentioned, the respondent would still only get one similarity point. Also, annotating or writing about a key feature was considered the same as drawing the key feature, e.g. *"extension"* was taken to mean the same as drawing a normal fault between reflectors that were offset extensionally.

4.5.1. Example interpretations with annotated key features

The following three respondents' interpretations have been chosen to illustrate how I assessed the interpretations. The respondents below have interpreted 2, 4 and 6 key features, respectively. The experts that chose each specific key feature are also noted. I provide a summary of respondents' interpretations, based on my observations. The tables on the bottom right-hand side of each page show how the 'Max RE Score' was determined for each respondent.

Figure 4.9 – Respondent 26's interpretation of the seismic image. Respondent 26 interpreted the left fault to be strike-slip, and the middle and right faults to be extensional. The left and middle faults both cut horizon 1. The individual key features interpreted by the respondent are noted inside white boxes on their interpretation. The Max RE Score attained was 2 key features.



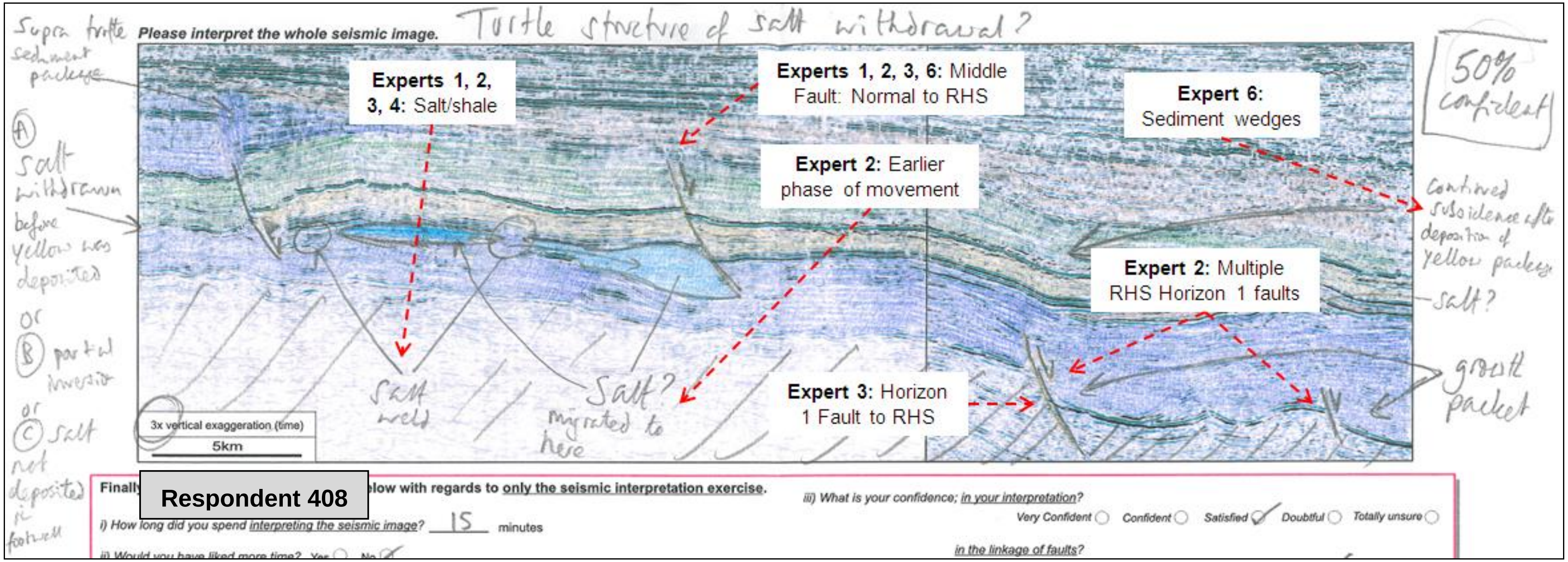
Expert	No. key features interpreted
1	0 out of 7
2	2 out of 9
3	1 out of 8
4	0 out of 6

Summary

Respondent 26 has interpreted the left fault as being strike-slip (the fault cuts into package 1 and looks like a flower structure), and has interpreted compression on the right side of horizon 1 (reverse fault offsetting horizon 1). The middle fault is a normal fault, downthrown to the right, cutting horizon 1. The Max RE Score attained was 2 key features, and the respondent's interpretation is most similar to Experts 2 and 6.

5	N/A
6	2 out of 7
Max RE Score = 2	

Figure 4.10 – Respondent 408's interpretation of the seismic image. The respondent interpreted inversion on the left fault and extension on the middle and right faults. The respondent also circled *"3x"* vertical exaggeration. The individual key features interpreted by the respondent are noted inside white boxes on their interpretation. The Max RE Score attained was 4 key features.



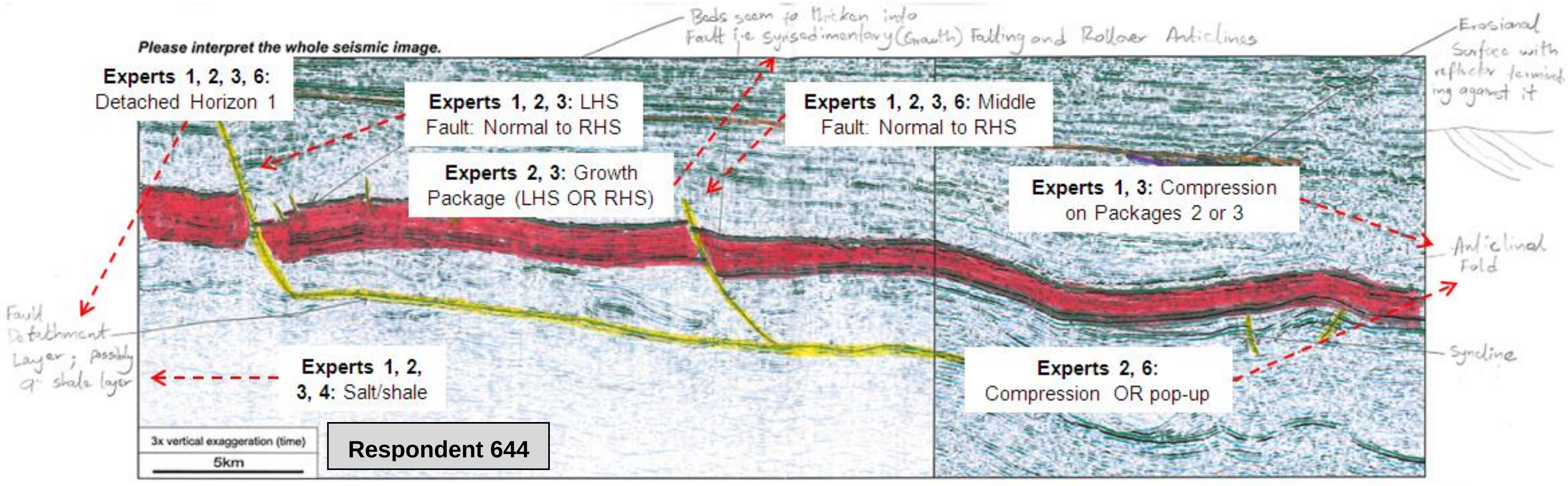
Expert	No. key features interpreted
1	2 out of 7
2	4 out of 9
3	3 out of 8
4	1 out of 6

Summary

Respondent 408 has interpreted the left fault as being inverted, compression then extension (double-headed arrow), with an earlier phase of movement (salt withdrawal) in package 2. The middle fault is a normal fault, downthrown to the right, not cutting horizon 1. Two extensional faults were interpreted, cutting the right side of horizon 1. *“Continued subsidence”* was interpreted in package 4, which implies that the sediment wedges were identified by the respondent. Multiple faults were interpreted on the right of horizon 1. The Max RE Score attained was 4 key features and the respondent’s interpretation is most similar to Expert 2.

5	N/A
6	2 out of 7
Max RE Score = 4	

Figure 4.11 – Respondent 644's interpretation of the seismic image. Respondent 644 interpreted an extensional fault on the left that soles-out on horizon 1, which they interpreted to be a shale detachment layer. Roll-over anticlines were noted. The fold on the right was identified and marked as an *"anticlinal fold"*. The Max RE Score attained was 6 key features. The individual key features interpreted by the respondent are noted inside white boxes on their interpretation.



Summary

Respondent 644 has interpreted the left fault as being listric and detaching on horizon 1, *“possibly on a shale layer”*. The middle fault is also listric and detaches onto horizon 1. Compression is interpreted on the right of horizons 2 and 3 in the form of an *“anticlinal fold”*. Growth faulting was identified on the left fault. The Max RE Score attained was 6 key features and the respondent’s interpretation is most similar to Experts 2 and 3.

Expert	**No. key features interpreted**
1	5 out of 7
2	6 out of 9
3	6 out of 8
4	1 out of 6
5	N/A
6	3 out of 7
Max RE Score = 6	

4.5.2. Comparing the example interpretations

Respondent 26 interpreted 2 key features, while respondent 408 interpreted 4 key features. Respondent 644 interpreted 6 key features and their interpretation was thus 'very similar' to Experts 1, 2, 3 and 4 (extensional faults detaching on salt/shale horizon with downdip compression). Respondent 644 attained high similarity scores for Experts 1, 2 and 3 and a lower score (one key feature) for Expert 4. It is interesting to note that respondent 26 did not explain their interpretation with labels or annotations, while respondents 408 and 644 did.

4.6. Distribution of the Max RE Score

Figure 4.12 shows the distribution of the 'maximum number of key features interpreted' (Max RE Score) for the 444 respondents. Seventy per cent of respondents either had '1', '2' or '3' key features and only 14% had five or more key features. The maximum score possible was 9 key features, but the median score was only 3 key features. The mean Max RE Score interpreted was 2.84 features. It was possible that the low median Max RE Score of 3 key features was partly due to non-response bias in the survey design, e.g. more capable geoscientists chose not to take part. However, to reduce the effect of potential bias, I actively tried to sample a wide range of geoscientists.

The variable in Figure 4.12 (Max RE Score) is used as the response variable in the main analysis (Chapter 6). The analyses determined which aspects of respondents' backgrounds (education, experience and training) were most associated to the Max RE Score variable. Knowing what variables were most associated to the Max RE Score variable allowed me to determine what 'types' of respondents were near the top end of the Max RE Score distribution, on average.

Figure 4.12 – The distribution of the Max RE Score variable for the 444 respondents. Six respondents achieved a Max RE Score of 0 key features, while one respondent achieved a Max RE Score of 8 key features. The median Max RE Score was 3 key features.

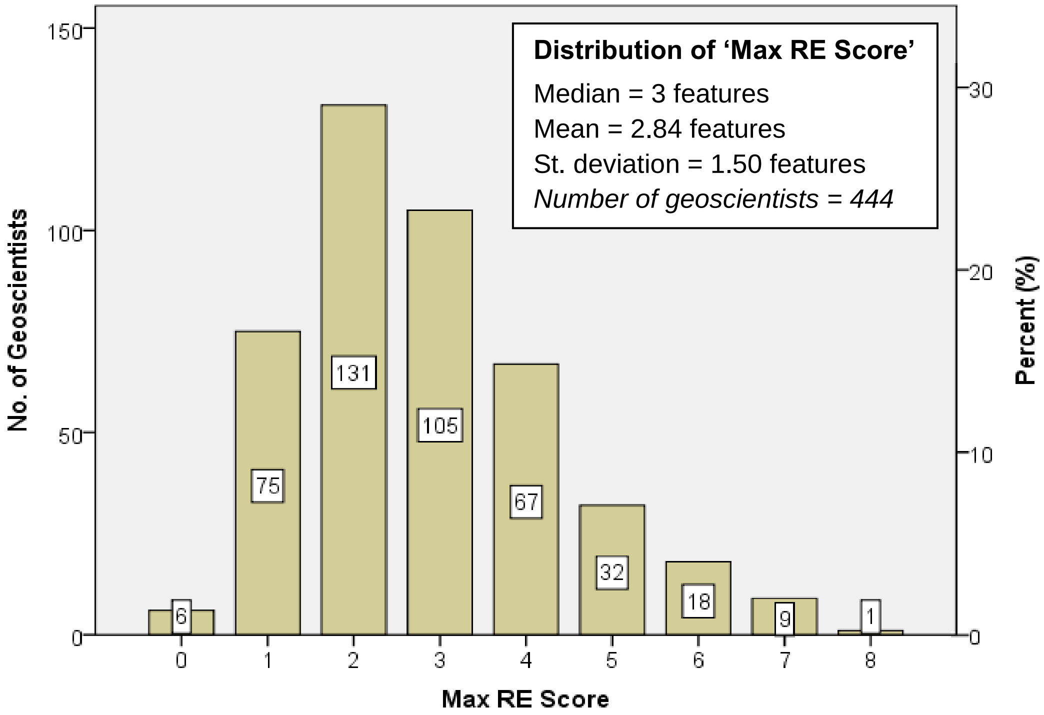


Table 4.9 presents a matrix that compares each expert against all other experts. The comparison is in terms of the reference experts' Max RE Scores for all other experts, to establish which experts picked similar and dissimilar key features. The matrix does not imply that some experts' interpretations were considered 'better' than others, just 'more similar' to each other. Columns contain the sets of key features that were chosen by the experts, and rows contain the number of key features interpreted by each expert. Because Expert 5 did not provide key features, other experts' interpretations could not be compared against them, and *"N/A"* was noted. Experts were also not given a similarity score for themselves. The grey cells indicate high similarity between experts' sets of key features (6+ key features the same). The Max RE Score is also listed for each expert in the right-hand column.

Table 4.9 – Reference experts' similarity scores for all other experts. Cells with 6+ key features are in grey to indicate high similarity between experts. Reference experts were not given a similarity score against themselves. Because Expert 5 did not provide key features, the other experts were not marked against them.

		Key features that were chosen by…						**Max RE Score**
		RE 1	**RE 2**	**RE 3**	**RE 4**	**RE 5**	**RE 6**	
Number of key features interpreted by…	**RE 1**	-	7	7	3	N/A	3	7
	RE 2	7	-	7	2	N/A	4	7
	RE 3	6	8	-	2	N/A	3	8
	RE 4	5	7	7	-	N/A	4	7
	RE 5	1	3	3	1	-	1	3
	RE 6	4	6	5	1	N/A	-	6

It can be seen from Table 4.9 that the experts generally interpreted more of Expert 1, 2 and 3's key features and less of Expert 4 and 6's key features. Expert 5 interpreted fewest of the other experts' key features, as indicated by the Max RE Score of 3, underlining the fact that Expert 5 had the most different interpretational approach to all other experts. It might also be true that if respondents used the same interpretational approach as Expert 5 did, they would be expected to score lower in the interpretation exercise. Different interpretational approaches were discussed in the methodology chapter.

4.7. Conclusions

Six reference experts with different technical backgrounds, from both industry and academia, were recruited to interpret the seismic image, allowing me to derive a scoring system for respondents' interpretations. There were differences in the reference experts' interpretations that highlighted the risk of using only one expert as other valid interpretations may not be identified. I noted differences in the interpreted tectonic style, the positions and dips of horizons and faults, and also differences in the timing of faults. Likely due to their different technical backgrounds, the reference experts also used different terminology to describe the same features.

The reference experts were asked to provide 'key geological features' that were integral to their interpretation. The chosen key features were then used to create the

‘Max RE Score’ response variable that was used in the main analysis. The median Max RE Score was 3 key features, and only 14% of respondents achieved Max RE Scores of 5 or more key features. Respondents who gained high Max RE Scores were said to have produced ‘similar’ interpretations to the reference experts.

Expert 5, whose interpretation was consistent with most of the other experts, did not provide key features as they said that their approach could not be represented by key features, as I had defined them. Expert 6 used a similar approach to most of the other experts but arrived at a tectonically different interpretation. Hence, different interpretational approaches can produce similar interpretations, and similar interpretational approaches can produce different interpretations.

This chapter documented the differences in the reference experts’ interpretations and showed how the response variable (Max RE Score) for the main analysis was derived. The next chapter explains how I defined the predictor variables (respondents’ backgrounds and the techniques that were used), which will be statistically modelled against the response variable.

Chapter 5 – Sample Information

In this chapter, I introduce the predictor variables that were created with the responses to the questionnaire and thus represent respondents' education, work environment and professional experience. I also define technique variables based on the interpretational approaches used by respondents. The predictor variables will be statistically modelled against the response variable (Max RE Score) in the main analysis (Chapter 6).

5.1. Introducing the questionnaire variables

As shown in Chapter 3, the background questionnaire captured information about respondents' education, work environment and experiential background. Table 5.1 to Table 5.4 present the variables from the questionnaires that were used in the analyses. Each table contains one group of variables: 'education and training', 'work environment', 'experience' and miscellaneous variables. The numbers of respondents in each category for each variable can be found in the data tables in Appendix 4. The right-hand column shows the variable name from the statistical coding that was used to perform the analysis (Chapter 6). I also note whether the variable was categorical or continuous.

The education variables (Table 5.1) captured whether respondents had Bachelor's, Master's or Ph.D. degrees; whether they had attended an industry graduate scheme, and whether they had taken training courses in structural geology and/or seismic interpretation.

Table 5.1 – The 'education' and 'training' variables from the background questionnaire. The response categories are noted in parenthesis. The type of variable, and the variable's name in the statistical coding, is noted.

Education and training variables	Variable type	Name in statistical coding
Q3: Completed a Bachelor's degree? (Yes, No)	Categorical	Q3_Bach
Q3: Completed a Master's degree? (Yes, No)	Categorical	Q3_Mast
Q3: Completed a Ph.D. degree? (Yes, No)	Categorical	Q3_PhD
Q15: Completed an industry graduate	Categorical	Q15_Graduate_Course

training course? (Yes, No)		
Q17: Completed a seismic interpretation course (not including University training)? (Yes, No)	Categorical	Q17_Seismic_Course
Q19: Completed a structural geology course (not including University training)? (Yes, No)	Categorical	Q19_Structure_Course

The 'work environment' variables (Table 5.2) covered academia, consultancy, service companies and oil companies. I sampled oil and gas industry professionals more heavily than other professions so that the results gained are more applicable to the industry. The type of oil and gas companies that respondents had experience in was captured: super-major, major, national and medium-small independent. In addition, I collected information on whether respondents had experience in exploration or production. I provided an 'other' category where respondents could write their profession if it did not fit into any of the categories. Post-data collection, I combined 'super-major' with 'major' oil companies, since in my experience, 'super-major' and 'major' oil companies are more similar than any of the others types of oil company that I specified and thus, geoscientists might be trained in similar ways.

Table 5.2 – The 'work environment' variables from the background questionnaire. The response categories are noted in parenthesis. The type of variable, and the variable's name in the statistical coding, is noted.

Work environment variables	**Variable type**	**Name in statistical coding**
Q7: Background is mainly in academia? (Yes, No)	Categorical	Q7_Academic
Q7: Background is mainly in consultancy? (Yes, No)	Categorical	Q7_Consultant
Q7: Background is mainly in a service company? (Yes, No)	Categorical	Q7_Service_Company
Q7: Background is mainly in super-major or major oil company? (Yes, No)	Categorical	Q7_Super_major_OR_major
Q7: Background is mainly in national oil company? (Yes, No)	Categorical	Q7_National
Q7: Background is mainly in medium, small or independent oil company? (Yes, No)	Categorical	Q7_Medium_small
Q7: Background in oil company is mainly in	Categorical	Q7_Exploration

exploration? (Yes, No)		
Q7: Background in oil company is mainly in production? (Yes, No)	Categorical	Q7_Production

I captured whether respondents had experience in structural geology (Q8) and seismic interpretation (Q9) as they were likely to impact respondents' interpretational abilities (Table 5.3). To determine what effect practise has, I captured how frequently respondents *"interpreted or used"* seismic images (Q10). Question 10 originally had the following six answer categories: 'daily', 'weekly', 'monthly', '6-monthly', 'yearly, and 'almost never'. However, post-data collection, I combined the categories into 'daily / weekly', 'monthly / 6-monthly' and 'yearly / almost never' to achieve a better distinction between the categories. For example, a respondent who interpreted seismic data a few times a week might be unsure whether they fit into the 'daily' or 'weekly' category. The new variable with three categories is likely to more accurately represent what respondents had intended, and represents 'regular users' (daily / weekly), 'occasional users' (monthly / 6-monthly) and 'very infrequent users' (yearly / almost never).

Question 13 captured respondents' 'most worked' geological setting, which will be used to test for anchoring bias in their tectonic interpretation of the seismic image. The geographical range of experience is represented by three variables; whether respondents had experience in more than one country, whether they had experience in more than one continent and the number of locations where they had investigated the geology. The number of years of experience since the completion of their highest degree was also captured. Respondents' range of technical skills in geoscience was recorded by Q11.

Table 5.3 – The 'experience' variables from the background questionnaire. The response categories are noted in parenthesis. The type of variable, and the variable's name in the statistical coding, is noted.

Experience variables	Variable type	Name in statistical coding
Q8: Level of experience in structural geology? (Specialist, Good Working Knowledge, Basic Working Knowledge, No	Categorical	Q8_Structural_Geology_Exp

Knowledge)		
Q9: Level of experience in seismic interpretation? (Specialist, Good Working Knowledge, Basic Working Knowledge, No Knowledge)	Categorical	Q9_Seismic_Exp
Q10: How often seismic images are interpreted or used? (Daily or Weekly, Monthly or 6-monthly, Yearly or Almost Never)	Categorical	Q10_How_Often_CATEG
Q13: Most worked geological setting? (Compressional tectonics, Extensional tectonics, Inversion tectonics, Salt or Shale tectonics, Strike-slip tectonics, Multiple settings)	Categorical	Q13_Most_Worked_Geo_Setting
Q21: Geology has been investigated in more than one country? (Yes, No)	Categorical	Q21_More_Than_One_Country
Q21: Geology has been investigated in more than one continent? (Yes, No)	Categorical	Q21_More_Than_One_Continent
Q5: Number of years of experience (relating to geoscience) since completion of highest degree? (number)	Continuous	Q5_Years_Experience
Q11: Number of active subject areas in geoscience in the last 24 months? (number: 0 to 13)	Continuous	Q11_Range_Disciplines
Q21: Number of geographical locations? (number)	Continuous	Q21_Number_of_Ticks

The variables in Tables 5.1, 5.2 and 5.3 were created from the questions that respondents filled in before completing the interpretation exercise. However, the three variables in Table 5.4 were created from questions that were filled in after the interpretation exercise. I captured the length of time spent on the exercise, whether respondents wanted more time, their confidence in their interpretation, and also their confidence in the linkage of their interpreted faults. Question i (Qi) is noted as being the ‘adjusted’ time as some respondents did not give a number, instead giving a range, e.g. *“10-15 minutes”*. In these cases, I took the middle value of the range and rounded up, so this example would be 13 minutes.

Table 5.4 – Miscellaneous variables from under the seismic image. The response categories are noted in parenthesis. The type of variable, and the variable's name in the statistical coding, is noted.

Miscellaneous variables	Variable type	Name in statistical coding
Qii: More time wanted to interpret seismic image? (Yes, No)	Categorical	Qii_Want_More_Time
Qiii: Confidence in interpretation? (Very Confident, Confident, Satisfied, Doubtful, Totally Unsure)	Categorical	Qiii_Interp_Confidence
Qi: Time spent interpreting the seismic image? (number)	Continuous	Qi_Adjusted_Time

5.2. About the respondents

In this section, I describe the responses to the questionnaire and summarise respondents' backgrounds. A summary of respondents' confidence in the interpretation exercise is also given.

5.2.1. Academic background and work environment

Of the 444 respondents, 85.6% had a Bachelor's degree, 54.5% had a Master's degree and 37.2% had a Ph.D. degree. The median years of experience, after the completion of the highest degree, was 10 years. Table 5.5 shows the work environment that describes respondents' backgrounds most accurately. Respondents were allowed to select multiple options to this question and it was not uncommon for them to do so. It can be seen that 57% of respondents (27.9 + 12.4 + 16.7) had experience in an oil company. Most respondents with 'experience in an oil company' had experience working in exploration (39.2%), while only 14.4% of the 444 respondents had experience working in production. Not all respondents indicated whether they worked in exploration or production, and some noted other areas such as *"research"* (7 respondents) and *"development"* (5 respondents). Only 9.8% of respondents had experience working in a service company, while 15.7% had experience in consultancy and 37% had experience in academia.

Table 5.5 – The percentage of respondents in each work environment category (Q7). Respondents were allowed to select more than one option. Grey boxes show the percentages of respondents that have worked in exploration or production in an oil company.

Q7: Work environment *"Which of the below describes your background most accurately?"*	**% of respondents with experience (number)**
Academia	37.0 (158)
Consultancy	15.7 (67)
Service company	9.8 (42)
Oil company: super-major or major	27.9 (124)
Oil company: national	12.4 (55)
Oil company: medium, small or independent	16.7 (74)
Oil company: in exploration	39.2 (174)
Oil company: in production	14.4 (64)

5.2.2. Experience in geoscience

Respondents had substantial experience in structural geology and in seismic interpretation. Sixty-five point five percent of respondents were a 'specialist' in, or had a 'good working knowledge' of, structural geology; while 59.6% were a 'specialist' in, or had a 'good working knowledge' of, seismic interpretation. Respondents also had similar levels of experience in each discipline. Table 5.6 shows that 242 respondents (27 + 120 + 95) rated their experience in both disciplines as being equal. Only 28 respondents (7 + 21) had a 'basic working knowledge' in one discipline and 'specialist' knowledge of the other discipline (the biggest difference possible). A chi-square test showed that 'structural geology experience' and 'seismic interpretation experience' were significantly associated for these respondents ($p<0.001$, Appendix 7).

Table 5.6 – Cross-tabulation of 'experience in structural geology' (Q8) with 'experience in seismic interpretation' (Q9). Respondents had similar levels of experience in both disciplines.

Number of respondents

		Q9: Level of experience in seismic interpretation?			
		Specialist	Good Working Knowledge	Basic Knowledge	Total
Q8: Level of experience in structural geology?	Specialist	27	29	21	77
	Good Working Knowledge	30	120	62	212
	Basic Knowledge	7	50	95	152
Total		64	199	178	441

5.2.3. Specialisms

Most respondents were active in multiple areas of geoscience within the last 24 months (Q11). The median 'number of areas of geoscience' where respondents were active was 4 subject areas. The four most commonly chosen areas (ranked as either '1' or '2'), out of the 444 respondents, were seismic interpretation (47.7%), structural geology (46.6%), geophysics (32.2%) and stratigraphy (27.9%). Respondents were allowed to select multiple options. Fifty-three respondents (11.7%) had experience in all subject areas listed in Question 11, within the last 24 months.

5.2.4. Geological Settings

Table 5.7 shows that respondents' most-worked geological setting (Q13), over the course of their entire career, was 'extensional tectonics' (48.2%). Nineteen point one percent of respondents had joint most-worked geological settings, i.e. they had equal experience in multiple tectonic regimes. Having a 'joint most-worked geological setting' meant that the respondents had no single 'most-worked' setting. 'Inversion tectonics' was the least-worked tectonic setting.

Table 5.7 – Respondents' most-worked geological setting (Q13). The most commonly worked geological setting was extensional tectonics. Inversion tectonics was the least-worked tectonic setting.

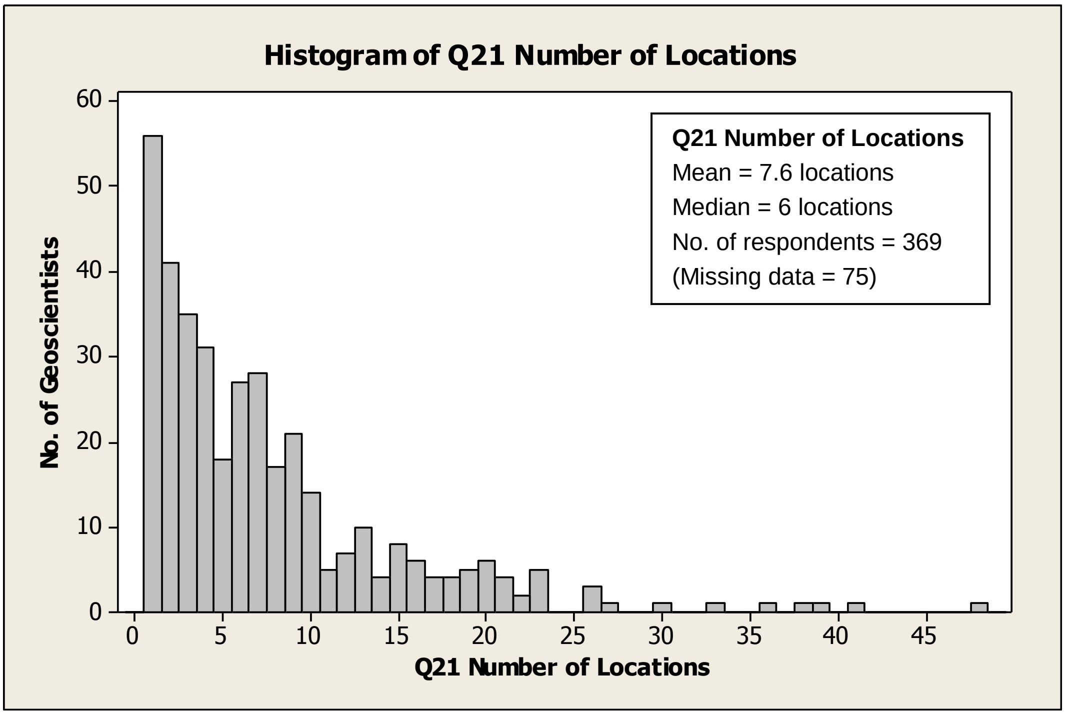
Q13: Most worked geological setting?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Compression	80	18.0	20.1	20.1
	Extension	192	43.2	48.2	68.3
	Inversion	7	1.6	1.8	70.1
	Multiple settings	76	17.1	19.1	89.2
	Salt	31	7.0	7.8	97.0
	Strike-slip	12	2.7	3.0	100.0
	Total	398	89.6	100.0	
Missing	.	46	10.4		
Total		444	100.0		

5.2.5. Worldwide locations

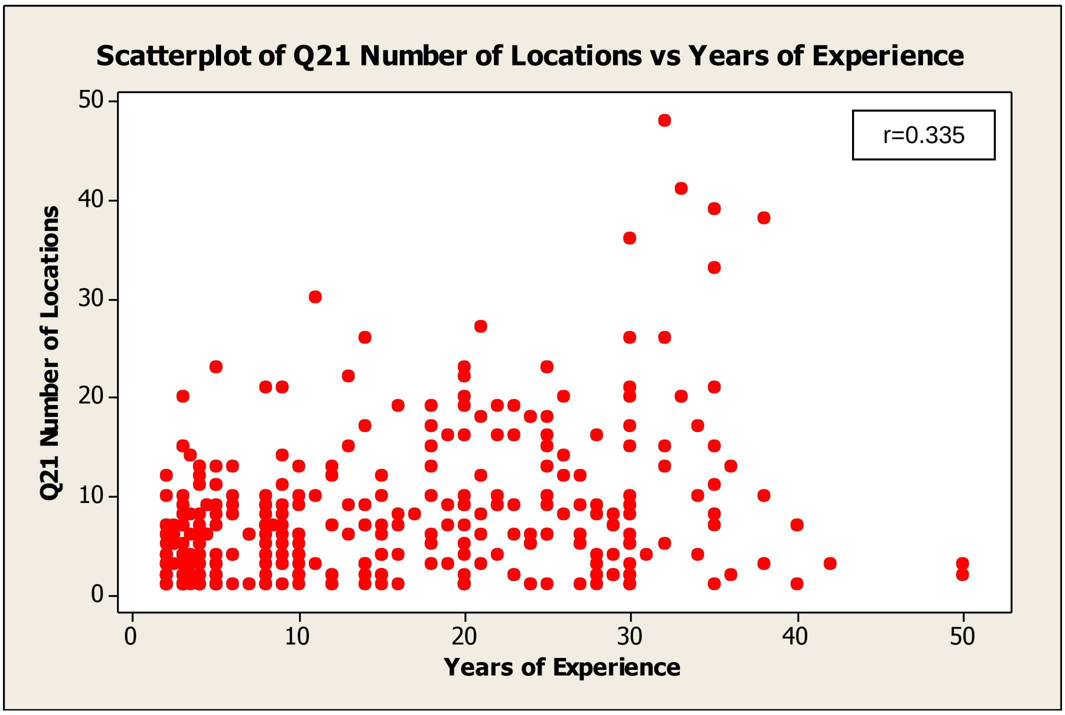
The median number of geographical locations where the geology had been investigated over respondents' entire careers (Q21) was 6 locations (Figure 5.1). Eighty one point eight percent of respondents had investigated the geology in more than one country, while 74.3% had investigated the geology in more than one continent. Seventy-five respondents missed out Q21.

Figure 5.1 – Distribution of the 'number of geographical locations where the geology had been investigated' (Q21). The data are skewed to the right, with 49% of respondents having identified 5 or fewer locations. The top 10% of the distribution had identified at least 17 locations, with the maximum number of locations being 48.



Interestingly, the number of locations where respondents had investigated the geology was not strongly correlated to their years of experience (Figure 5.2). For example, some respondents with low 'years of experience' had investigated the geology in the same number of locations as respondents with 20+ years more experience than them. The two respondents with 50 years' of experience indicated that they had only investigated the geology in 2 and 3 locations. These data do not explain why respondents 'numbers of locations' was not correlated to respondents' years of experience. However, one possible explanation is that younger geoscientists might have selected geographical locations where they had minimal experience, and some older geoscientists might have only selected the locations where they had substantial experience (years).

Figure 5.2 – Scatterplot of 'number of locations' (Q21) against 'years of experience' (Q5). There was no strong linear relationship between the variables (Pearson's correlation coefficient is 0.335).



5.2.6. Interpretational confidence

After completing the seismic interpretation exercise, respondents were asked for their 'confidence in their interpretation' (Qiii). The majority of respondents (50.9%) were 'satisfied' with their interpretation (Table 5.8). 'Interpretational confidence' was non-significant in any of the analyses (Chapter 6). Similarly, 'confident' respondents did not produce better interpretations than 'doubtful' respondents (p=0.11, Appendix 7). Hence, there was no statistical evidence of the overconfidence bias in these data. As in Chapter 3, I used words rather than probabilities to define confidence levels, which may have mitigated the effect of the overconfidence bias.

Table 5.8 – Respondents' stated confidence in their interpretations. The responses were surprisingly symmetric with 87 respondents who were 'confident' and 86 respondents who were "doubtful', and 12 respondents were 'very confident' and 12 respondents were 'totally unsure'.

Qiii: Confidence in interpretation?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Very Confident	12	2.7	3.0	3.0
	Confident	87	19.6	21.7	24.7
	Satisfied	204	45.9	50.9	75.6
	Doubtful	86	19.4	21.4	97.0
	Totally Unsure	12	2.7	3.0	100.0
	Total	401	90.3	100.0	
Missing	.	43	9.7		
Total		444	100.0		

5.3. Introducing the technique variables

Based on the interpretations produced by the respondents in the main dataset, I identified 17 separate techniques and split them into the following four groups: writing, drawing, time and concepts. The groupings were only approximate as some techniques could have potentially fitted in multiple categories. Whether or not respondents had used each of the 17 techniques in their interpretation was determined via visual inspection. The evidence that I used to make the decisions, for whether respondents had applied each technique included markings, lines, shadings and words, as well as any explanations written above or beside the seismic image. The written comments were therefore taken to be part of the interpretation. Hence, I analysed respondents' interpretations after they had completed the exercise.

On the 'post-interpretation' questionnaire that was used in the workshop experiment (Chapter 7), one workshop participant discussed the impact that 'salt' had, but did not include 'salt' on their interpretation. The participant thus, did not communicate their interpretation properly as there was no written evidence of 'salt' even though they were thinking about it. For the respondents from the main dataset, who were not asked to complete the 'post-interpretation' questionnaire, I could not have included their thoughts unless they were written down.

It is not known whether all respondents were equally keen to explain their interpretation. For example, some respondents may have avoided adding information to their interpretation to ‘keep it clean’ or maybe lacked the confidence to explain their interpretation. It was not logistically possible to interview respondents after they had interpreted the seismic image to confirm that their interpretations were written down accurately. In the discussion, I argue that thoughts and annotations should be systematically added as they improve the communication of interpretations.

Table 5.9 to Table 5.12 present examples of the four groups of technique variables. Within each table, the following information is given: the variable name, whether the variable was categorical or continuous, how the technique was defined, the percentage of the 444 respondents who used that technique, and a photograph of part of a respondent’s interpretation that illustrates the technique.

The four groups of techniques are first presented and then described underneath because of the following change of page orientation.

Table 5.9 – The two 'writing' variables. These techniques represented different forms of writing and were used by less than half of the respondents.

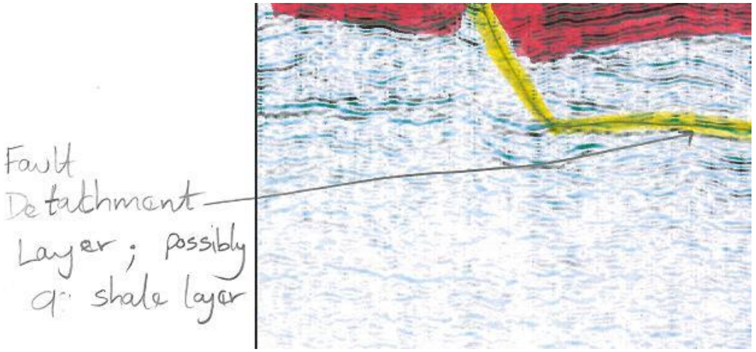
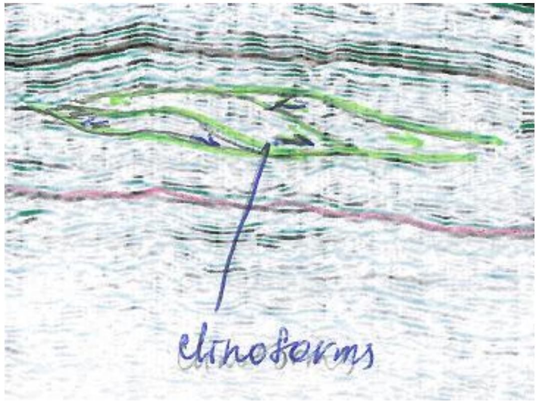
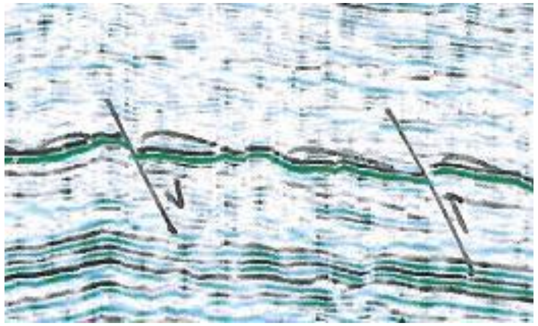
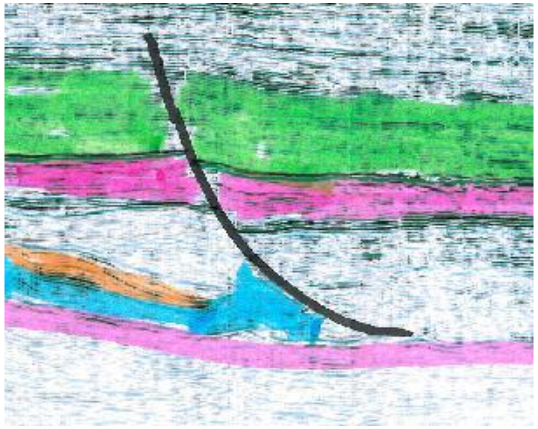
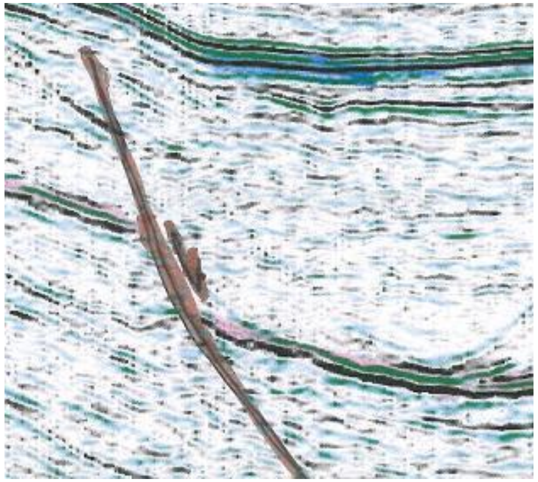
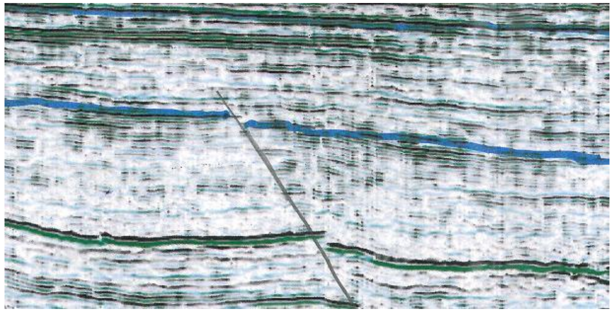
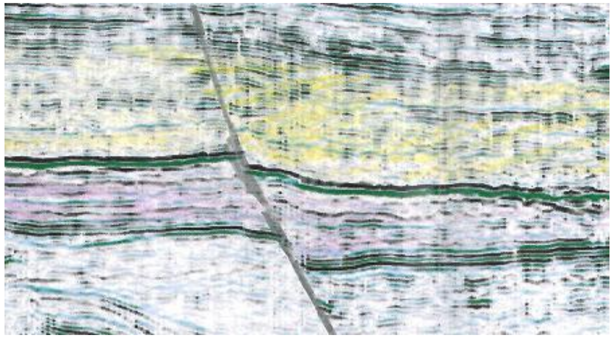
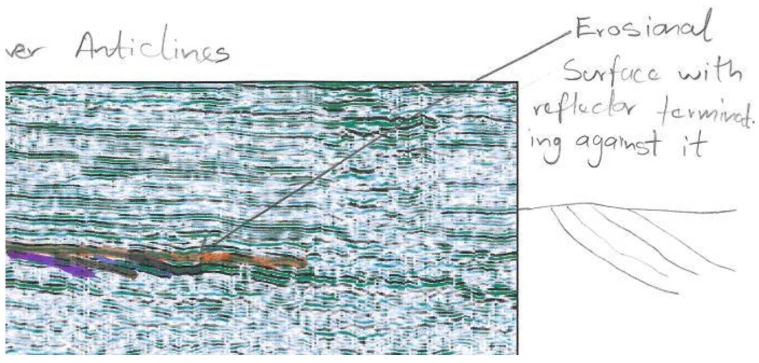
Writing variables *Name in statistical coding*	**Variable type**	**% of respondents who used technique (#)**	**Definition of technique, and description of technique in action (italics)**	**Photograph of part of a respondent's interpretation**
Descriptions or explanations added? *T1_Descriptive_Explaining*	Categorical	25.7 (114)	Phrases or sentences that described or explained part of the interpretation. *The respondent explained their observation of a fault detachment and inferred that a shale layer might be involved.*	
Annotations or labels added? *T2_Annotations_Labelling*	Categorical	41.0 (182)	A single word or short phrase labelling part of the interpretation (e.g. *"salt"*) but no evidence of reasoning. *The respondent annotated "clinoforms" on their interpretation. There was no explanation, just a label for a localised feature.*	

Table 5.10 – The seven 'drawing' variables. These techniques represented different forms of drawing, including fault and horizon interpretation. It was common for respondents to interpret fault sticks (83.1%) but less common for them to drawn a cartoon (2.7%).

Drawing variables *Name in statistical coding*	**Variable type**	**% of respondents who used technique**	**Definition of technique, and description of technique in action (Italics)**	**Photograph of part of a respondent's interpretation**
Fault sticks drawn? *T3_Fault_Sticks*	Categorical	83.1 (369)	Faults drawn with straight lines, no curvature. Does not include when faults is written about, only drawn. *The respondent has drawn small straight faults showing extensional and compressional areas.*	
Faults drawn with non-planar geometry? *T4_Faults_with_Geometry*	Categorical	66.4 (295)	Faults drawn with non-planar geometry. Does not include when faults are written about, only drawn. *The respondent has drawn a curved fault that appears to sole out on the bottom reflector (light pink).*	

Arrows drawn on faults? *T5_Arrows_For_Faults*	Categorical	31.5 (140)	Arrows drawn for faults showing the relative up/down movement of the footwall / hanging wall. Fault can be noted instead of drawn. *The respondent has drawn an arrow on the extensional fault showing that the hanging wall was down thrown.*	
Horizons drawn in? *T7_Marking_Horizons*	Categorical	79.3 (352)	Drawn in (or highlighted in any way) at least 1/3 of any horizon. Colouring in packages above and below the horizon, in different colours, also counted as this defines the boundary. *The respondent has correlated the blue horizon across an extensional fault.*	

Packages coloured or drawn in? *T8_Colouring_Marking_Packages*	Categorical	38.5 (171)	Pen or pencil used to shade, colour or fill packages. *The respondent has coloured in the packages in different colours. The top package is yellow, while the bottom package is pink.*	
Cartoon drawn? *T10_Cartoon_Drawn*	Categorical	2.7 (12)	Cartoon or sketch drawn that explains any part of the interpretation. *The respondent has identified reflector terminations and inferred them to be an erosional surface. A cartoon was drawn highlighting the reflector termination interaction.*	

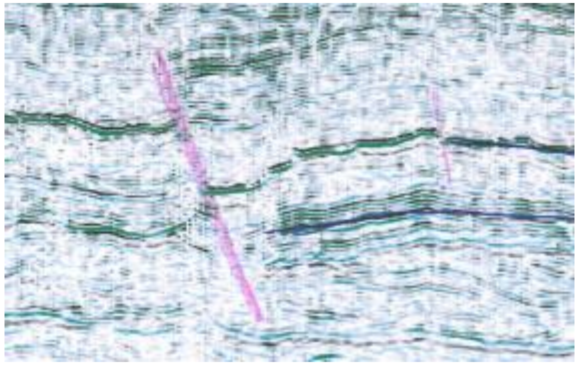
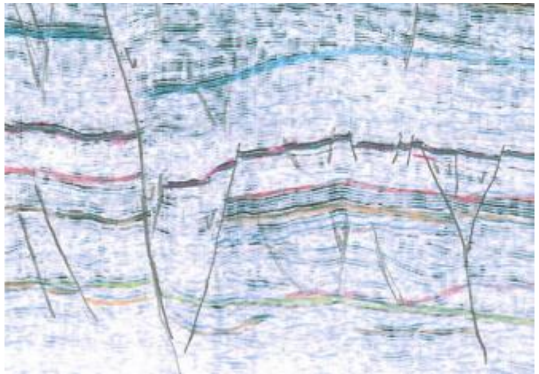
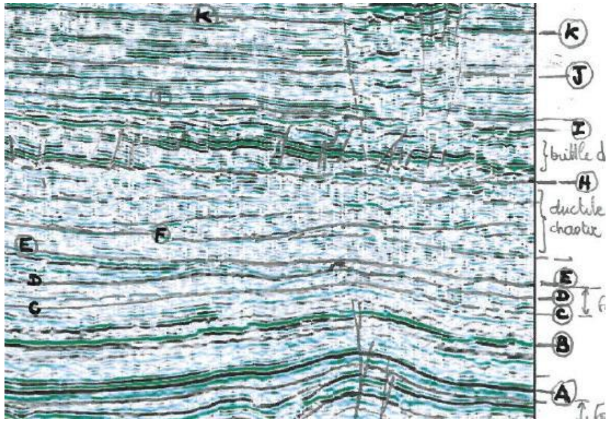
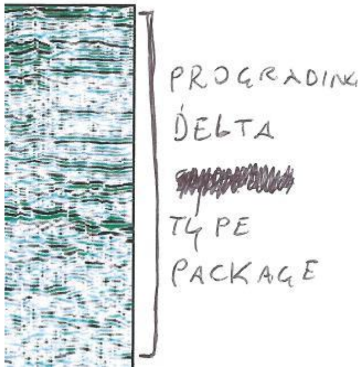
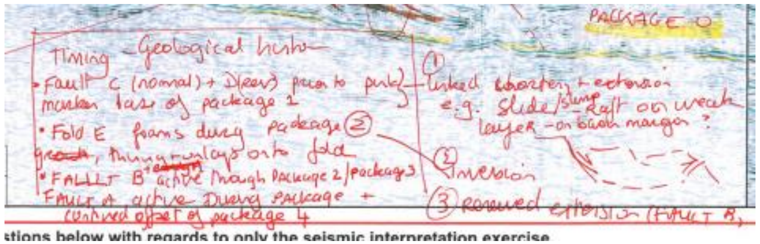
Level of detail? *Level_of_Detail*	Continuous	N/A	A count of the main structural features and the small faults interpreted over the whole image; no maximum value as respondents could have drawn as many small faults as they wanted. *The top respondent has interpreted an extensional fault and one smaller fault ('level of detail' score = 10).* *The bottom respondent has interpreted the same extensional fault but added lots of additional detail in this area, e.g. synthetic and antithetic faults ('level of detail' score= 45).*	 

Table 5.11 – The five 'time' variables. These techniques represented different forms of geological time and also geological reasoning more generally. 'Writing about geological time' was used by 22.7% of respondents, while 'considering the geological evolution' was used by 1.1%.

Time variables *Name in statistical coding*	**Variable type**	**% of respondents who used technique**	**Definition of technique, and description of technique in action (Italics)**	**Photograph of part of a respondent's interpretation**
Horizons ordered? *T9_Horizons_Ordered*	Categorical	7.7 (34)	Conditional on horizons being highlighted (T7). The highlighted horizons are ordered by some number/letter system, implying relative ages. *The respondent has ordered the reflector horizons and correlated them across the seismic image.*	
Written about geological time? *T11_Writing_About_Time*	Categorical	22.7 (101)	Local scale features specifically about geological time. Examples were noted in 'Example Words' document in Appendix 4. *The respondent has stated that the lower package was the pre-rift and the top package was the syn-rift.*	

Written about geological processes? *T12_Writing_About_Process*	Categorical	27.9 (124)	Other local scale features, e.g. clinoforms, prograding delta, harpoon structure. Examples were noted in ‘Example Words’ document in Appendix 4. *The respondent has identified the bracketed sequences to be a prograding delta package.*	
Geological evolution stated? *T15_Evolutionary_Thought*	Categorical	1.1 (5)	Changes in large-scale structure over time, e.g. evolution pictures or listing evolutionary steps. Not local scale. *The respondent has written out the geological history of the section and linked it to the interpreted seismic image.* *The respondent has also drawn a series of three evolutionary cartoons to illustrate how the structural geology evolved over time.*	

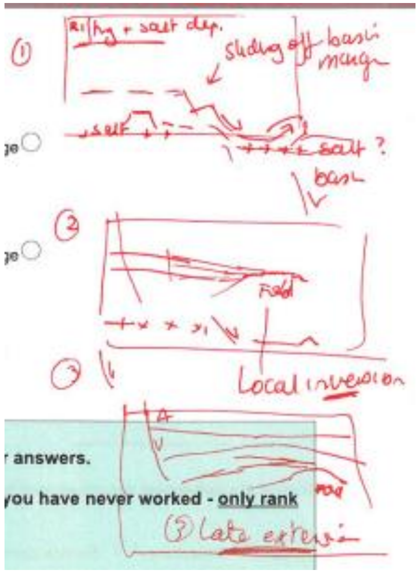
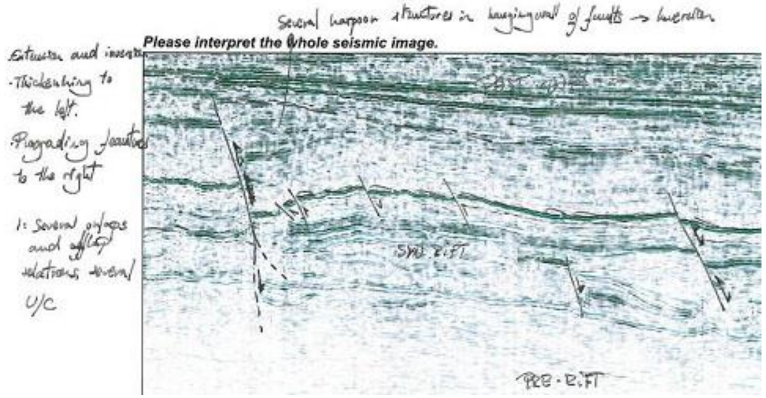
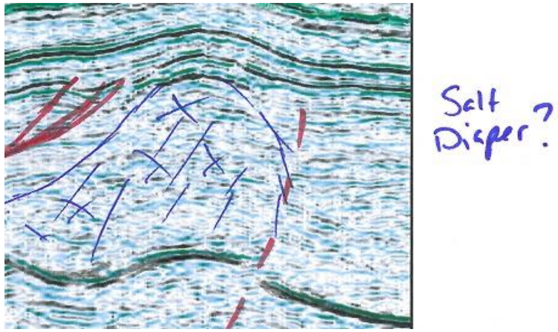
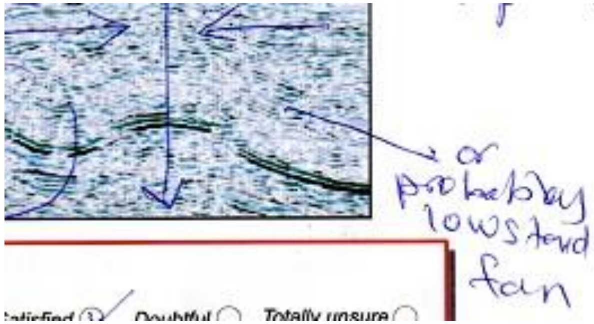
				
Justified interpretation (without use of geological evolution) stated? *T16_Justified_Interp_not_evolution*	Categorical	6.8 (30)	Extensive explanation: cartoons and/or have shown substantial reasoning. Not including evolutionary thought. *The respondent has explained their interpretation in detail (writing on left margin and at top), but not considered the geological evolution of it.*	

Table 5.12 – The three ‘concept’ techniques. These techniques represented different ways of using geological theory, such as stating a tectonic concept or using sequence stratigraphy. ‘Asking questions’ is included here because respondent ask questions are acknowledging the uncertainty in their interpretation.

Concept variables *Name in statistical coding*	**Variable type**	**% of respondents who used technique**	**Definition of technique, and description of technique in action (Italics)**	**Photograph of part of a respondent’s interpretation**
Concept explicitly stated? *T13_Concept_Explicitly_ Stated*	Categorical	8.8 (39)	At least one word written to describe tectonic concept (e.g. *“extension”*, *“compression”*). *The respondent has written “extensional basin” along the top of their interpretation.*	
Questions stated? *T14_Are_Questions_Stat ed*	Categorical	29.3 (130)	A question mark noted on interpretation (*“?”*), even on its own. Indicates uncertainty in part of the interpretation. *The respondent was unsure in their interpretation and stated “salt diaper?”*	

Sequence stratigraphy used? *T17_Sequence_Stratigraphy*	Categorical	20.0 (89)	Sequence stratigraphy arrows drawn showing the termination style of horizon, e.g. onlap, offlap, toplap. Includes mentioning *"onlap"* / *"offlap"* etc instead of drawing arrows. Also includes *"Highstand"*, *"Lowstand"*, *"Regression"* etc… *The respondent noted "probably a lowstand fan".* *The respondent has drawn onlap arrows on reflectors that terminate against another reflector. These terminations have been circled and respondent has annotated "pinch out onlaps" underneath.*	

5.4. Usage of techniques

In this section, I describe the techniques and their usage. The techniques are grouped by their type.

5.4.1. Writing

Using annotations or labelling was the most basic technique where respondents note a single word or phrase that clarifies their interpretation without making an inference, e.g. *“salt”*, rather than *“this looks like salt because…”*. Only 40.5% used annotations or labelling. The descriptive or explaining technique captured whether there was evidence for respondents using geological reasoning. Descriptive or explaining was used by 25.7% of respondents.

5.4.2. Drawing

Drawing techniques were used much more frequently than the writing techniques. Eighty-three point one percent of respondents drew fault sticks, while 66.4% drew faults with geometry. Fault sticks were defined to be short straight lines, while faults with geometry had a non-planar geometry. The ‘arrows on faults’ variable, which was used by 31.5% of respondents, captured whether arrows were drawn to indicate the downthrown side of a fault. Respondents could potentially have written about the fault rather than drawing one, but no respondents did this. Seventy-nine point three percent of respondents drew horizons and 38.5% coloured or marked packages, e.g. between horizons. Colouring in two successive vertical packages, in different colours, was counted as ‘drawing in the horizon boundary’ since the boundary would be defined. Hence, this means that those respondents who coloured in successive packages in different colours would have applied two techniques according to my classification scheme (‘colouring packages’ and ‘drawing horizons’). However, this is not necessarily beneficial in any way.

Drawing a cartoon was used by only 2.7% of respondents. The ‘level of detail’ score was the summation of the number of large faults, horizons and smaller faults (e.g. anti and synthetic) that were interpreted. Hence, respondents with high ‘level of detail’ scores went into a lot more detail in the interpretation of the seismic image.

5.4.3. Time

Horizons were ordered by only 7.7% of respondents. Writing about time was used by only 22.7% of respondents (e.g. *"syn-rift"*), and writing about processes was used by 27.9% of respondents (e.g. *"clinoforms"*). 'Writing about time' concerned 'local scale' features directly related to geological time, while 'writing about geological processes' captured other 'local scale' features (i.e. not related to geological time'). Examples of both geological time and geological processes can be found in Appendix 4. Evolutionary thought, which was thinking about geological time over a larger scale, was only used by 1.1% of respondents. Only one respondent out of 444 considered the geological evolution of the seismic image by drawing a series of cartoons, while the others wrote the evolutionary steps out. Respondents needed to show evidence of thinking about the large scale changes in structure over time.

'Justified thought', used by only 6.8% of respondents, was considered to be 'extensive explanation of the interpretation but without using evolutionary thought'. To allow me to test for differences between 'justified thought' and 'evolutionary thought', I ensured that no respondents could have both used 'evolutionary thought' and 'justified interpretation'. I did this by sorting the interpretations into three separate piles; 'evolutionary thought', 'justified interpretation' and neither. Hence, no respondents could have applied both techniques.

5.4.4. Concepts

Only 8.8% of respondents noted the overall tectonic concept that they had applied during their interpretation, e.g. *"strike-slip"*. I only included instances of the concept being stated when the concept applied to most of the seismic image and ignored annotations on a specific fault, e.g. *"extensional fault"*. Twenty-nine point three percent of the 444 respondents noted questions, which could be as short as *"?"*. Asking questions may be related to confidence, but given the unconstrained nature of seismic data, asking questions may be useful. I found that there was virtually no difference between respondents' confidences in their interpretations for those respondents who had asked questions and those who had not. Sequence stratigraphy was applied by 20% of respondents; this included marking horizons' termination points (e.g. onlap, toplap), and also using sequence stratigraphy words, such as *"high stand"*, *"transgression"* or *"maximum flooding surface"*.

5.4.5. What techniques were used most?

The techniques are ranked by usage in Table 5.13, and the frequency tables for each technique can be found in Appendix 4. ‘Fault sticks’ and ‘marking horizons’ were used most, while ‘evolutionary thought’ and ‘cartoon drawn’ were used least. Generally speaking, the ‘drawing’ techniques were used more often than the ‘time’ techniques. The writing techniques were used infrequently, e.g. only 40.5% of respondents used ‘annotations or labelling’, and only 25.7% of respondents used the ‘descriptive or explaining’ technique. Two hundred and fifty-four of the 444 respondents (57.2%) used neither writing technique; 104 respondents (23.4%) used both techniques, and 86 respondents (19.4%) used one writing technique but not the other.

Table 5.13 – The usage of 16 techniques by the 444 sample respondents. The technique of ‘level of detail’ is not included here as it was a count of the geological features and all respondents used it to some extent. The techniques that were used least required geological reasoning.

Variable Name	% of 444 respondents who used technique	Type of technique
Fault sticks drawn?	83.1 (369)	Drawing
Horizons drawn in?	79.3 (352)	Drawing
Faults drawn with non-planar geometry?	66.4 (295)	Drawing
Annotations or labels added?	41.0 (182)	Writing
Packages coloured or drawn in?	38.5 (171)	Drawing
Arrows drawn on faults?	31.5 (140)	Drawing
Questions stated?	29.3 (130)	Concept
Written about geological processes?	27.9 (124)	Time
Descriptions or explanations added?	25.7 (114)	Writing
Written about geological time?	22.7 (101)	Time
Sequence stratigraphy used?	20.0 (89)	Concept
Concept explicitly stated?	8.8 (39)	Concept
Horizons ordered?	7.7 (34)	Time
Justified interpretation (without use of geological evolution) stated?	6.8 (30)	Time
Cartoon drawn?	2.7 (12)	Drawing
Geological evolution stated?	1.1 (5)	Time

5.5. Conditionality and dependence of techniques

In most cases, the interpretational techniques were completely independent of each other. For example, 'descriptive or explaining' and 'evolutionary thought' were not dependent because evidence of the usage of evolutionary thought could be in the form of a series of cartoons instead of being written down and descriptions or explanations could be about other aspects of the interpretation, other than the geological evolution. However, four techniques were subsets of other techniques:

- 'Evolutionary thought' was a subset of 'writing about time'. It was the most restricted version of 'writing about time'. Hence, respondents could have written about time but not applied 'evolutionary thought', but if respondents had applied 'evolutionary thought' then they had also 'written about time', by definition.
- 'Arrows on faults' was a subset of the union of 'fault sticks' and 'faults with geometry'; meaning that the respondents that had used the technique of 'arrows on faults' must have also used 'fault sticks' and/or 'faults with geometry'. Respondents could have noted a fault via writing then drawn an arrow to illustrate the relative fault movement, but none did.
- 'Writing about time' and 'writing about processes' were both subsets of the union between 'annotations or labelling' and 'descriptive or explaining'. The respondents that had used 'writing about time' (or 'writing about processes') must have also used 'annotations or labelling' and/or 'descriptive or explaining'. This is because 'annotations or labelling' and/or 'descriptive or explaining' capture whether respondents wrote on their interpretation, while 'writing about time' and 'writing about processes' capture different forms of written geological reasoning.

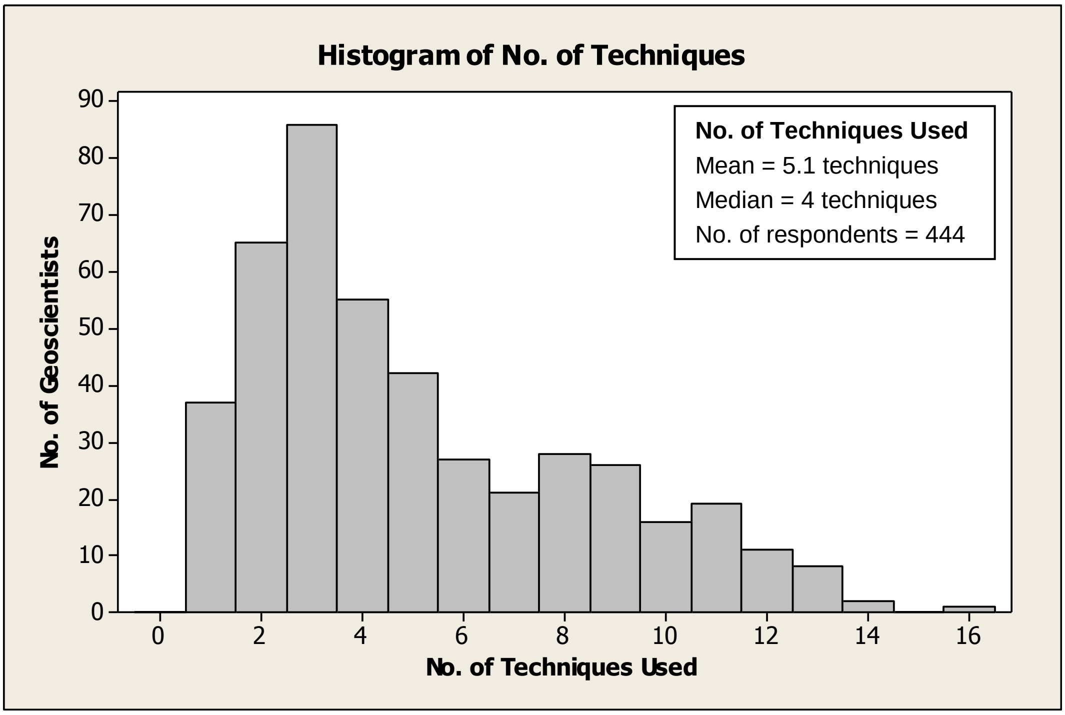
Variables that are subsets of each other should not generally be analysed together as they are dependent. However, in the context of this research, it was worth distinguishing between general uses of techniques and special cases that have geological importance. For example, there was an important distinction to be made between 'writing about time' and 'describing and explaining'. The 'writing about time' technique captured whether respondents approached the exercise with geological time in mind. The 'descriptive or explaining' technique, however, only captured the usage of logical thought, without distinguishing this from thinking about geological time. Hence, I decided to include all 17 'technique' variables in the analysis.

It was clear that some techniques were more ‘advanced’ than others. Perhaps a sufficient number of years of experience are needed before it becomes easier to use the advanced techniques? This question will be addressed by the statistical analysis in Chapter 6.

5.6. Number of techniques used

The usage of different techniques is presented above, but in this section I discuss the total number of techniques that were used by respondents. Figure 5.3 shows the distribution of the number of techniques used in the interpretation exercise. The median number of techniques used by respondents was four; 54.7% of the 444 respondents used four techniques or less, while 9.2% of respondents used 10 or more techniques.

Figure 5.3 – Distribution showing the ‘number of techniques’ used by respondents. The median number of techniques used by respondents was 4, while the maximum was 16.



The preliminary data exploration showed that the ‘number of techniques’ variable was positively associated to the ‘Max RE Score’ variable ($r=0.606$; $p<0.001$). Therefore, respondents who used more techniques also attained higher Max RE

Scores, in general. However, I wanted to go further than just showing that using more techniques was better than using fewer techniques, and determine the individual techniques that were most effective. I thus decided not to use 'number of techniques' in the main analysis as it was significantly associated to other predictor variables, including 'writing about time', 'annotations or labelling', 'descriptive or explaining', and 'arrows for faults'. For example, the median number of techniques used for respondents who 'wrote about time' was 9, compared to a median of 3 techniques for respondents who did not 'write about time'. It was important not to include the 'number of techniques' variables as it would mask the effect of the individual techniques in the analysis.

5.7. Conclusions

This chapter described how the responses to the background questionnaire were used to create the predictor variables for the main analysis. I defined the variables and then reported the number of respondents that were in each category, thus summarising respondents' backgrounds.

The majority of respondents had either a Master's or Ph.D. degree. The sample as a whole had a median of 10 years of experience after the completion of their highest degree. Fifty-seven percent of respondents had experience working in an oil company, while 37% had experience working in academia. Sixty-six percent of respondents had at least a 'good working knowledge' of structural geology; while 60% had at least a 'good working knowledge' of seismic interpretation. I established that the majority of respondents had equal levels of experience in both structural geology and seismic interpretation. The median number of geographical locations where respondents had investigated the geology was 6, but two respondents had noted more than 40 locations. However, there was no correlation between the years of experience and the number of locations. There was also no evidence of overconfidence, on average, in respondents' interpretations as most were 'satisfied' with their interpretation and few respondents were 'very confident'.

In addition to respondents' educational and experiential backgrounds, the techniques used also affect the interpretation of geological data. I defined 17 techniques based on respondents' interpretations of the seismic image to analyse their individual effect in the main analysis. The most commonly used techniques were based on drawing faults or horizons, while the more ‘advanced’ techniques that were based on geological reasoning were used least. Four techniques were special cases of other techniques, but I retained them for the analysis because of their geological significance. The median number of techniques used was four, while 9% of respondents used 10 or more techniques. I established a significant positive association between the ‘number of techniques’ used and the Max RE Score, but did not to use the ‘number of techniques’ variable in the main analysis as it would mask the effect of individual techniques.

The following chapter details the statistical analysis of the response and predictor variables, aiming to answer the following two research questions:

- What ‘type’ of interpreters produced the ‘most similar’ interpretations to the reference experts?
- What techniques were most effective?

I also determine whether respondents’ backgrounds or the interpretational techniques used were most associated to producing a valid interpretation.

Chapter 6 – Main Analysis

What 'types' of respondents gained the highest Max RE Scores? And, *what interpretational techniques made them effective?* This chapter explains how I used statistical modelling to answer these two research questions using the survey data. In previous chapters, I described how the six reference experts' interpretations were used to make the 'Max RE Score' response variable, which was a 'scale of similarity' between the respondents' interpretations and the reference experts' interpretations. The survey data were used to make the predictor variables that represented respondents' backgrounds and the interpretational techniques that were used. In this chapter, I present the data input procedure, the relevant theory and the results of the analyses.

6.1. Introducing the main analysis

The variables used in the analyses were split into the following two groups: 'background' and 'techniques'. The background variables were the variables from the questionnaire that captured respondents' education, geological experience and training, while the technique variables captured how respondents interpreted the seismic image. All variables were discussed in the 'sample information' chapter. In the 'reference experts' chapter, I introduced the key features that the reference experts picked and I then described how I made the 'Max RE Score' variable from the key features.

The 'Max RE Score' was the response variable, and the background and technique variables were the predictor variables in this analysis. The analysis allowed me to build a statistical model that best described the response variable via a combination of the predictor variables. The analysis determined which variables should be included in the model; and thus, identified the variables that were highly associated to the response variable.

In addition, I consider whether respondents' backgrounds had a greater impact on the 'Max RE Score' attained than the techniques used, indicating whether the most effective techniques can only be used by experienced geoscientists.

6.2. Data input procedure

I input respondents' background data separately from their technique data, and afterwards, independently input the key feature data. Inputting the three different sets of data separately meant that I was not influenced by my knowledge of the respondent. For example, if I had known that a specific respondent was a *"specialist"* in seismic interpretation, I might have expected them to interpret more key features than a respondent who had a *"basic working knowledge"* in seismic interpretation. I therefore covered over the bottom half of the questionnaire and could not see the outside pages (pages 1 and 4) while inputting the key feature data; I could only see respondents' interpretations. The key feature data was thus input 'blindly'.

The technique and key feature data came from respondents' interpretations and it was therefore not possible to input these data without 'seeing' the other set of data. For instance, when deciding whether respondents interpreted the 'regional slope' key feature, it was not possible not to notice whether respondents had also used 'annotations or labelling', as I had to read the annotations to be able to judge whether respondents had interpreted the key feature. It should be emphasised though, that I input the sets of data separately and followed the pre-determined guidelines strictly.

I also checked some of the variables for consistency. In these cases, I re-input the data for the 444 respondents, to ensure that the same acceptance criteria had been applied for all respondents. I used the 'data validation' feature within Microsoft Excel to ensure that only input pre-approved values would be accepted into each cell in the spreadsheet. Using data validation decreased the likelihood of me making data input mistakes. After inputting the data, I also tabulated the responses to check that I did not miss out any cells, which would not have been picked up by the data validation tool. I found few mistakes in the data input and these were corrected before the analysis.

6.3. Relevant statistical theory

I next describe the relevant statistical theory needed to investigate the research questions. I introduce the modelling technique used and explain how the p-values and odds ratios were used to determine which variables were chosen for the statistical model. I describe the reasons for analysing variables in a univariate analysis before using a multivariate analysis, and explain the variable selection procedure that was used during the multivariate analysis. I then outline the steps that I took to build the statistical model and summarise the modelling checks that were applied. Hosmer and Lemeshow (1989), Agresti (1996) and Kleinbaum (2002) should be consulted for details of logistic regression and odds ratios.

6.3.1. Modelling technique

I used the statistical modelling technique of ‘ordinal logistic regression’ since the data type of the response variable (Max RE Score) was ordered and categorical (e.g. ‘ordinal’). Ordinal logistic regression is a form of ‘generalised linear models’ (Dobson and Barnett, 2008), and builds an equation (statistical model) to maximise how well the response variable can be predicted, based on a combination of the predictor variables. The ‘best’ statistical model is one that explains the data well but is parsimonious. Statistical models with too many predictor variables can be mathematically unstable and the results can be less transferable to the underlying population as the model becomes too specific. Therefore, the influential predictor variables need to be selected from the group of all predictor variables, to maximise how well the equation explains the observed data, but without over explaining it.

6.3.2. Univariate and multivariate analyses

A ‘univariate’ analysis is when one predictor variable is analysed against the response variable on its own. On the other hand, a multivariate analysis is where the variables are analysed relatively to one another, e.g. after allowing for the ‘effect’ of other variables. I used univariate analyses to screen predictor variables before putting the significant variables into the multivariate analysis; this is standard statistical protocol. Univariate analyses are not results since they do not take into account the inter-dependencies between predictor variables, which can lead to confounding results. A fictional example of confounding results is as follows: a univariate analysis showed that there was a strong relationship between heart disease and eating apples. However, perhaps regular apple eaters were also older and exercised less than the population in general. These factors could also be strongly related to heart disease, meaning that the univariate ‘result’ (there is a strong relationship between heart disease and eating apples) would not be valid, because the other factors were not taken into account. In this research, all analyses were multivariate to ensure robustness.

6.3.3. P-values

The ‘p-value’ gives an indication of how strong the evidence is for whether the variable in question should stay in the statistical model or whether it should be rejected (a binary outcome). The ‘p-value’ is a probability that ranges from just above zero[9] to one, and measures the chance that the observed relationship in the data was due solely to chance. For example, if the p-value equals 0.02, then there is only a 2% chance that the observed relationship in the data was due to chance. The significance level, that defines how ‘strong’ the evidence needs to be, was chosen to be 5%, following standard statistical protocol. Thus, if the p-value of a variable was <0.05, the variable would be classed as ‘significant’ and not rejected, i.e. there was enough evidence that the relationship between the response and predictor variables was not due to chance. However, if the p-value is =>0.05 then the evidence is not strong enough and the observed relationship might be due to chance. The variable is then rejected from the model.

6.3.4. Variable selection procedure

I ‘screened’ the predictor variables with univariate analyses to determine which variables were associated to the response variable, i.e. each predictor variable was analysed independently to test for basic association with the response variable. Variables that were significant in the univariate analysis were referred to as ‘independently significant’. During this screening phase, I used the higher significance level of 10%, instead of the usual 5% applied at all other stages. Independently significant variables were then added to a separate multivariate model, to be analysed simultaneously. Screening predictor variables with univariate analyses is a standard statistical protocol because there is less chance of variables being falsely rejected from the model, than if all variables were added to the multivariate analysis at the start of the analysis. It is also more efficient to remove the predictor variables that are clearly not associated to the response variable.

In the multivariate analyses, I input all independently significant variables at the start and then removed the least-significant variables to build a parsimonious model that

[9] The p-value cannot actually equal zero, but is instead noted as being *“<0.001”*.

explained the response variable. The iterative variable selection procedure of ‘backwards stepwise regression’ was used to remove one variable at each step. I administered the procedure manually. P-values were first calculated for all variables in the model and then the least-significant variable was removed (i.e. the variable with the highest p-value was rejected from the model). Once this variable had been removed, the p-values of all other variables were then re-evaluated and compared. Again, the least-significant variable was removed. The selection procedure was repeated until all variables in the model were significant (i.e. had p-values less than 0.05). The p-values for variables changed at each step, since the analysis compared the variables relative to one other. However, highly significant variables were usually significant the whole way through.

6.3.5. Odds ratios

The ‘odds ratio’ is a measure to quantify how strong the relationship is between the predictor and response variables, and is different from the significance of a variable. The significance is a binary outcome (significant and non-significant), while the odds ratio is based on a continuous scale. The odds ratio indicates how much more (or less) likely respondents would be, to be in a higher category of the response variable, if they had originally been in a different category of the predictor variable.

It is important to emphasise that these odds ratios relate to when the other variables are held constant, i.e. when there are no changes in other predictor variables. Variables with the largest odds ratios therefore have the most influence on the response variable. There are three types of predictor variable: categorical with only two categories, categorical with 2+ categories, or continuous. An example of odds ratios will be given for each type of predictor variable, using a fictional ‘heart disease and eating apples’ theme. ‘Heart disease’ will be the response variable and ‘eating apples’ will be the predictor variable. The eating apples variable could potentially take three forms depending on how the data are grouped. For example, the predictor variable could be:

- Categorical with two categories: ‘eat apples’ or ‘do not eat apples’.
- Categorical with 2+ categories: ‘eat 0 apples per day’, ‘eat 1 apple per day’ or ‘eat 2+ apples per day’.
- Continuous: the weight of apples eaten per day in grams.

The ‘heart disease’ response variable will have the following categories: 1) ‘severe risk of heart disease’, 2) ‘mild risk of heart disease’, and 3) ‘no risk of heart disease’. Example odds ratios are now explained for each type of predictor variable:

- *Two categories (‘eat apples’ or ‘do not eat apples’).*
  In this scenario, an odds ratio of, say, 2.39 means that individuals in the ‘eat apples’ category of the predictor variable would be 2.39 times more likely to be in a higher category of the response variable (lower risk of heart disease) than if they had been in the other category of the predictor variable to begin with (‘do not eat apples’). In plain English, this means that individuals would be 2.39 times more likely to being in a lower risk category for heart disease if they eat apples.

- *More than two categories (‘0 apples per day’, ‘1 apple per day’ or ‘2+ apples per day’).*
  A ‘reference category’ is needed since there are more than two categories in the predictor variable. The other categories are then compared against the reference category individually, giving an odds ratio for each comparison. There are two associated odds ratios in this example since there are three categories. Generally, for a variable with n categories, there are n-1 odds ratios produced for that variable in the analysis. The reference category is usually chosen to be the first or last category of the variable. In this example, if the reference category was chosen to be ‘0 apples per day’, then the two comparisons would be: ‘2+ apples per day’ compared to ‘0 apples per day’, and ‘1 apple per day’ compared to ‘0 apples per day’. In this example, suppose that the first odds ratio (‘2+ apples per day’ compared to ‘0 apples per day’) was 1.89; and the second odds ratio (‘1 apple per day’ compared to ‘0 apples per day’) was 1.32. The first odds ratios says that individuals who eat 2+ apples per day would be 1.89 times more likely to be in a lower risk category for heart disease than if individuals ate zero apples per day. The second odds ratio says that individuals who eat 1 apple per day would be 1.35 times more likely to be in a lower risk category for heart disease than if they ate zero apples per day. Hence, it makes sense that, if ‘eating apples’ reduces the risk of heart disease, then eating 2+ apples per day has a larger effect (i.e. larger odds ratio) than ‘eating 1 apple per day’.

- *Continuous (the weight of apples eaten per day in grams).*
  For a continuous variable, the odds ratio refers to the increased or decreased likelihood per unit of the variable since there are no categories. In this example the 'units' are the grams of apple that are eaten. Each additional gram of apple eaten per day may increase individuals' likelihoods of being in a higher category of the response variable. For example, if the odds ratio was 1.02, then for each gram of apple eaten, individuals' likelihoods would increase by 2%, e.g. 8 grams of apple per day would yield an odds ratio of 1.02^8=1.17. Hence, in this case, individuals would be 1.17 times more likely to be in a higher category of the response variable (more likely to be in a lower risk category for heart disease) than individuals who eat zero grams of apple per day.

It should be emphasised that these fictional examples are only given to elucidate what the odds ratios mean in relation to the main results for this research.

6.3.6. Confidence intervals

The odds ratios are actually 'point estimates' that have their own inherent uncertainty, which is due to the data being collected by random sampling, i.e. if we collected a new sample, then the odds ratios would all be slightly different. Hence, following standard statistical protocol, 95% confidence intervals were given with the odds ratios as a measure of their precision. The 95% confidence interval is the interval that the true unknown odds ratio (which is being estimated by the observed data) would be within, 95% of the time, if the experiment was repeated many times. Thus, small confidence intervals (tight around the point estimate) relate to high precision and larger intervals indicate low precision. Also, if the confidence interval contains the number '1', then the odds ratio is non-significant (i.e. p=>0.05).

6.3.7. Sampling weights

Sampling weights arithmetically adjust for the relative over (or under) sampling of categories within a variable. For example, since I found that the '21-30' year old age category had been over-sampled and the '61+' year old age category had been under-sampled (section 3.7.2) I could have used sampling weights to adjust for this difference. The sampling weight would have been less than one for the '21-30' age category to ensure that the effect of these respondents within the analysis was lessened, since they were over-represented in the sample. On the other hand, the sampling weight would have been greater than one for the '61+' age category to enlarge the impact of these respondents within the analysis since they were under-represented in the sample. The intermediate age categories ('31-40', '41-50', '51-60') were a good match to the four geoscience organisations that were used to represent the true but unknown population of geoscientists who work in Europe and the USA (Chapter 3).

However, I decided not to use sampling weights because I did not know how precisely the four organisations approximated the underlying population. While the match was reasonable, the error gained by approximating the underlying population with the four geoscience organisations might be larger than the decrease in error gained by using sampling weights. For instance, if I had contacted different geoscience organisations, the age distributions might have looked different and hence, the sampling weights would have been different.

6.3.8. Modelling checks

After obtaining the final statistical model (results shown in Table 6.5), I added each non-significant variable back into the model individually to determine whether any variables had been falsely rejected during the selection procedure. I found that no variables had been falsely rejected.

6.3.9. Interaction effects

I also tested the final model for 'interaction effects' but found none significant. An interaction effect occurs when one predictor variable affects the relationship of another predictor variable with the response variable. For example, perhaps a wide range of geographical experience is needed before respondents are able to draw meaningful cartoons that assist their interpretations? E.g. the technique of 'drawing cartoons' may only become significantly associated to the Max RE Score variable when respondents have investigated the geology in many geographical locations, and not before. If this is true (it is not), then the technique of 'drawing cartoons' and 'years of experience' would be 'interacting', which would be represented within the statistical model as an interaction effect. However, I found no significant interaction effects. The variables in the final model were hence, those in Table 6.5.

6.3.10. Goodness of fit

To determine the 'goodness of fit' of the statistical models, e.g. how well the models represented the observed data, I used the Akaike Information Criterion (AIC) value (Akaike, 1974) and the Nagelkerke pseudo R^2 value (Nagelkerke, 1991). Both measures help explain how well the equation explains the observed data, but in different ways. The AIC value measures the 'quality' of a model with a penalty for it being overly complex, i.e. having too many predictor variables. There is no absolute scale for the AIC value; rather, competing models are compared relatively, with small AIC values relating to better models.

The Nagelkerke pseudo R^2 value is an estimate (based on the observed data) that explains how much of the variance in the response variable is explained solely by the model. The Nagelkerke R^2 value has an absolute scale that ranges from zero, a very bad fit, to one, a perfect fit. If the R^2 value is subtracted from one, this gives the percentage of variation in the response variable that is still unexplained by the statistical model, which is useful when using the model to make predictions. However, in this research, I was not interested in building a predictive model[10] and was instead looking to determine significant relationships within the dataset. The values of the AIC and R^2 measures should be taken lightly as it was more useful to compare which model fitted the data best rather than what the indicator values were.

6.4. Results of the analysis

Using the respondents' background data, the technique data and the Max RE Score achieved by respondents, I will answer the following three research questions:

- *Which respondents were effective? (Analysing the background variables on their own).*
- *Why were they effective? (Analysing the technique variables on their own).*
- *Is it their background or the techniques used? (Analysing all variables together).*

The statistical analyses from this chapter were completed in the *"IBM SPSS Statistics 18"* software package. The code and relevant output can be found in

[10] The difference between 'investigating relationships within a dataset' and 'building a predictive model for new respondents' is that building a predictive model is more challenging due to 'unknown unknowns'.

(Appendix 5). The code and output shows the order that the variables were rejected in for each analysis.

The 'background' variables were first screened via univariate analyses (Table 6.1) and then I analysed the remaining variables multivariately to determine which respondents were effective, i.e. achieved high Max RE Scores (results in Table 6.2). The background variables were analysed on their own to ensure that the results were robust. For example, if the technique variables had been added then it would be unclear whether any differences are due to the background variables or the techniques that had been used. I also considered which aspects of respondents' backgrounds ('education or training', 'work environment' and 'experience') were most associated to the Max RE Score.

The techniques were also screened via univariate analyses (Table 6.3), and independently significant technique variables were added to a separate multivariate analysis to investigate what made respondents effective (results in Table 6.4). This analysis of the techniques was completed without the background variables for the same reason as above. In addition, I determined which types of techniques ('writing', 'drawing', 'time' and 'concept') were most associated to the Max RE Score.

After I had analysed the background and technique variables in their own multivariate analyses, I conducted a separate analysis to determine whether the techniques were significant in addition to the background variables (Table 6.5). If any 'technique' variables were significant at the end of the analysis, there would be strong evidence that interpretational techniques can assist respondents in their interpretations, regardless of their background. Throughout this results section, I have clearly noted whether the analysis was univariate or multivariate, and what set of data (e.g. background variables or technique variables) was used.

6.4.1. Univariate screening of 'background' variables: which respondents were effective?

The background variables (education, experience and training) were screened in univariate analyses to test for association with the response variable. Table 6.1 contains the independently significant variables ($p<0.10$) in the left column and the non-significant variables in the right-hand column, each listed in alphabetical order.

Eleven variables were independently significant and will be added to the multivariate analysis, while 17 variables were non-significant and will not be added to the multivariate analysis. ‘Having a Bachelor’s degree’ and ‘having a Master’s degree’ were not significant, while ‘having a Ph.D. degree’ was significant.

Table 6.1 – The results from the univariate screening of the background variables. The independently significant variables are in the left column while the non-significant variables are in the right column.

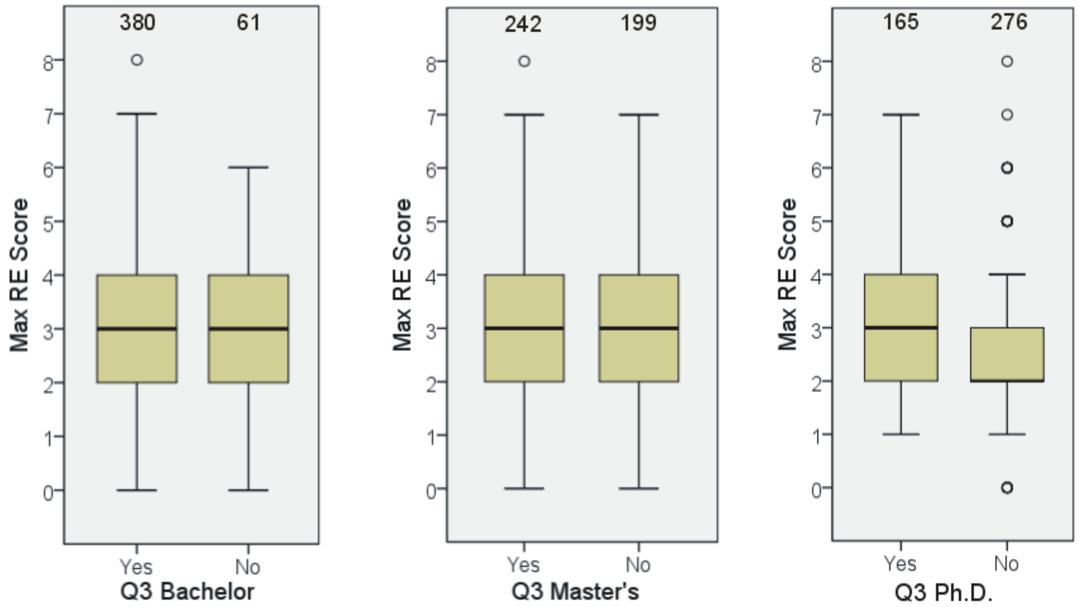
Significant variables (p<0.10) (will be added to multivariate analysis)	**Non-significant variables (p=>0.10)** (removed from analysis)
Q3: Completed a Ph.D. degree?	Q1: Gender
Q5: Number of years of experience (relating to geoscience) since completion of highest degree?	Q2: Age
Q7: Background is mainly in a service company?	Q3: Completed a Bachelor’s degree?
Q7: Background is mainly in super-major or major oil company?	Q3: Completed a Master’s degree?
Q8: Level of experience in structural geology?	Q7: Background is mainly in academia?
Q9: Level of experience in seismic interpretation?	Q7: Background is mainly in consultancy?
Q10: How often seismic images are interpreted or used?	Q7: Background in oil company is mainly in exploration?
Q19: Completed a structural geology course (not including University training)?	Q7: Background is mainly in medium, small or independent oil company?
Q21: Geology has been investigated in more than one country?	Q7: Background is mainly in national oil company?
Q21: Geology has been investigated in more than one continent?	Q7: Background in oil company is mainly in production?
Q21: Number of geographical locations?	Q11: Number of active subject areas in geoscience in the last 24 months?
	Q13: Most worked geological setting?
	Q15: Completed an industry graduate training course?
	Q17: Completed a seismic interpretation course (not including University training)?
	Qi: Time spent interpreting the seismic

	image?
	Qii: More time wanted to interpret seismic image?
	Qiii: Confidence in interpretation?

Figure 6.1 illustrates why ‘having a PhD degree’ was significant, while having a Bachelor’s or Master’s degrees were not associated to the Max RE Score. The first two boxplots (Bachelor’s and Master’s degrees) show no clear differences between the distributions of Max RE Score for respondents who had, and did not have, the respective degrees. However, it can be seen that ‘having a Ph.D. degree’ was positively associated to the Max RE Score, since the median value for the respondents in the ‘yes’ category was larger than the median value for the respondents in the ‘no’ category (3 key features opposed to 2 key features).

Respondents with a Ph.D., in general, tended to achieve higher Max RE Scores, but the circles for the ‘no’ category (representing outliers) show that it was possible to achieve a high Max RE Score without a Ph.D. degree. In fact, the single respondent who achieved 8 out of the 9 key features (the highest Max RE Score) had a Master’s degree and not a Ph.D. degree. The number at the top of the boxplots in this chapter is the number of respondents in that category. For example, 380 respondents had a Bachelor’s degree, 61 did not, and 3 did not answer the question. Non-response to questions means that not all sets of numbers sum to 444, which is the total number of respondents.

Figure 6.1 – Boxplots showing the distributions of Max RE Scores for the education variables (Q3). Bachelors and Master's degrees were not associated to the Max RE Score, while a Ph.D. degree was associated positively to the Max RE Score. The bold numbers at the top show the how many respondents were in each category.



6.4.2. Multivariate analysis of 'background' variables: which respondents were effective?

The independently significant background variables in the left column of Table 6.1 were then analysed in the multivariate analysis (results are in Table 6.2). The variables are listed in order of decreasing odds ratio, each with a 95% confidence interval (CI). Three variables were significant in the multivariate analysis ($p<0.05$), while 8 variables were non-significant. The AIC value of the statistical model was 854.8, which will be compared to the AIC values from the other models below. While, the Nagelkerke pseudo R^2 value was 0.167, which shows that the model does not explain the data well as 83.3% (1 – 0.167 = 0.833) of the variance in the Max RE Score variable is unexplained by the model.

Table 6.2 – The results from the multivariate analysis of the background variables. Variables are listed in order of decreasing odds ratios. Confidence intervals (CI) for the odds ratios are also noted. The AIC value was 854.8, and the Nagelkerke pseudo R^2 value was 0.167.

Variable name	Background type	P-value	Odds ratio (OR) with 95% CI
Q8: Level of experience in structural geology?	Experience	<0.001	
"Specialist" to "Basic Working Knowledge"			3.85 (2.17–6.83)
"Good Working Knowledge" to "Basic Working Knowledge"			1.28 (0.84–1.95)
Q10: How often seismic images are interpreted or used?	Experience	0.02	
"Daily / Weekly" to "Yearly / Almost Never"			2.14 (1.32–3.47)
"Monthly / 6-Monthly" to "Yearly / Almost Never"			1.57 (0.92–2.68)
Q21: Number of geographical locations?	Experience	<0.001	
Per location, since a continuous variable			1.06 (1.03–1.09)

The background variables that were not multivariately significant were as follows (p=>0.05):

Q3: Completed a Ph.D. degree?, Q7: Background is mainly in a service company?, Q7: Background is mainly in super-major or major oil company?, Q9: Level of experience in seismic interpretation?, Q19: Completed a structural geology course (not including University training)?, Q21: Geology has been investigated in more than one country?, Q21: Geology has been investigated in more than one continent?, Q5: Number of years of experience (relating to geoscience) since completion of highest degree?

The significant background variables were from the 'experience' category, showing that experience was generally more associated to the response variable than 'education or training' and the 'work environment' variables. In all cases, the association was positive, meaning that additional experience increased the odds of attaining a greater Max RE Score. 'Structural geology experience' had the highest odds ratio in the multivariate analysis, OR=3.85 (95% CI: 2.17–6.83), for 'specialist' compared to 'basic working knowledge'. Interestingly, the odds ratio for 'good working knowledge' compared to 'basic working knowledge' contained the number 'one' (i.e. 95% CI: 0.84–1.95) and hence, was non-significant on its own. However, since the variable contains both comparisons, the variable was significant overall, shown by $p<0.001$. It was also clear that the gains in the Max RE Score only started once respondents were ‘specialists’ in structural geology since there was not much difference between in the Max RE Score for respondents with a ‘good working knowledge’ and ‘basic working knowledge’ of structural geology. As discussed in section 3.4, the original question might have better distinguished between the medium levels of experience if I had used five categories instead of four.

‘How often you interpret or use seismic data’ had the second highest odds ratio (OR=2.14, 95% CI: 1.32–3.47) for ‘daily or weekly’ compared to ‘yearly or almost never’. ‘The number of geographical locations where the geology had been investigated’ was also significant with an odds ratio of 1.06 (95% CI: 1.03–1.09), meaning that respondents were 1.06 times more likely to be in a higher category of the response variable for each additional location. Clearly, for this result to occur, respondents who had a wide range of geographical experience attained higher Max RE Scores than those respondents, on average, who a smaller range of experience. There is hence, evidence to support the old adage that *“the best geologist is the one who has seen the most rocks”*.

Figure 6.2 shows boxplots of the Max RE Score for the two significant categorical variables. I observed a very strong positive association between Max RE Score and structural geology experience, which also explains why the variable had the largest odds ratio in the multivariate analysis. For example, the median Max RE Score for ‘specialists’ in structural geology was 4 key features, while the median Max RE Score for respondents with a ‘basic working knowledge’ of structural geology was 2 key features. ‘How often you interpret or use seismic images’ was also associated to the response variable since ‘yearly or almost never’ respondents, on average, attained smaller Max RE Scores than the other two categories (‘daily or weekly’ and ‘monthly or 6-monthly’). The median Max RE Score for ‘daily or weekly’ and ‘monthly or 6-monthly’ respondents was 3 key features, while for ‘yearly or almost never’ the median Max RE Score was 2 key features.

Figure 6.2 – Boxplots showing the distributions of Max RE Score for ‘structural geology experience’ (Q8) and ‘how often you interpret or use seismic images’ (Q10). Structural geology experience was strongly associated to the Max RE Score, while ‘how often you interpret or use seismic images’ was not.

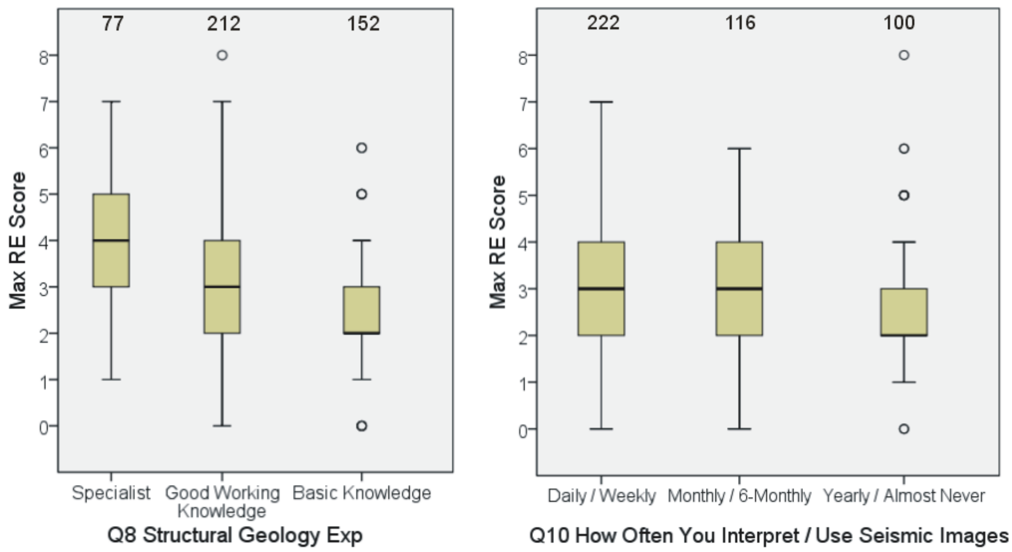
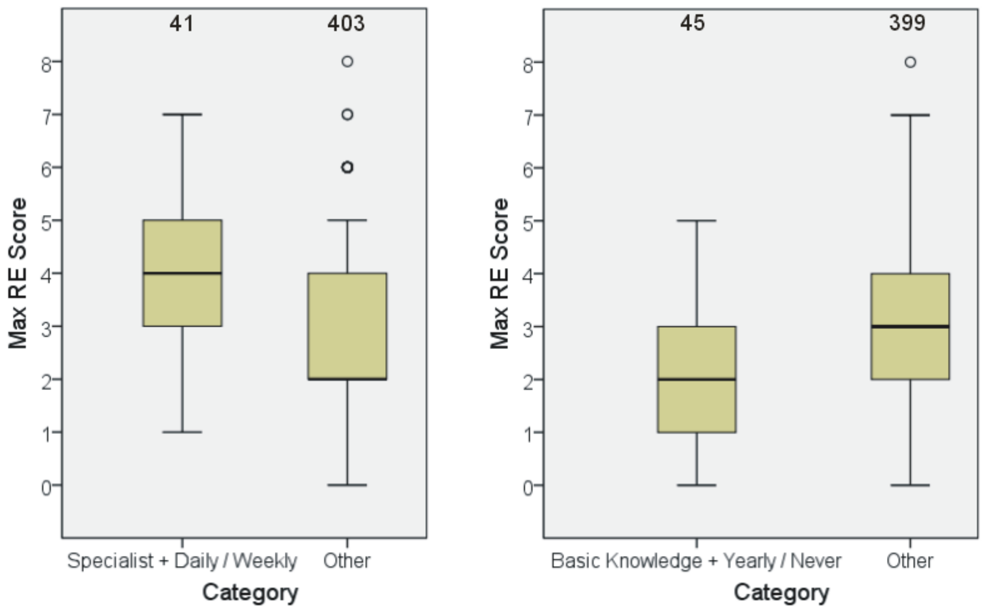


Figure 6.3 shows the effect of being a ‘specialist in structural geology’ AND ‘interpreting or using seismic images daily or weekly’. Therefore, the respondents in both of these categories were compared against all other respondents. The median Max RE Score for these respondents was 4 key features, compared to the median Max RE Score of 2 key features for all other respondents. Also, the right-hand boxplot compares the respondents who had a ‘basic working knowledge of structural geology’ AND who ‘interpreted or used seismic images yearly or almost never’ against all other respondents. The median Max RE Score for these less experienced respondents was 2 key features, with a median Max RE Score of 3 key features for all other respondents. The addition of ‘interpreting seismic images daily or weekly’ to being a ‘specialist in structural geology’ did not increase the median Max RE Score attained, since the median Max RE Score for ‘specialists in structural geology’ was also 4 key features (Figure 6.2).

Figure 6.3 – The boxplots on the left show the distribution of Max RE Scores for respondents who were 'specialists in structural geology' AND who 'interpreted or used seismic images either daily or weekly'. The boxplots on the right show the distribution of Max RE Scores for respondents who had a 'basic working knowledge of structural geology' AND who 'interpreted seismic images either yearly or never'.



From the boxplots and multivariate analysis it was clear that, on average, there was a strong positive relationship between respondents' level of experience and their Max RE Score. The respondents that achieved high Max RE Scores were therefore the respondents that were experienced in terms of their structural geology ability, how frequently they interpreted seismic images and their range of geographical experience. However, outliers occur in the boxplots (denoted by circles), suggesting that some respondents achieved high Max RE Scores with less experience than other respondents.

It is not yet clear whether the experienced respondents interpreted more key features, on average, because they were more experienced, or whether the experienced respondents just used more effective techniques than the less experienced respondents. The next section considers the technique variables independently, to determine which were most effective.

6.4.3. Univariate screening of 'technique' variables: why were they effective?

The technique variables were screened in univariate analyses to test for a basic level of association before being added to a multivariate analysis. Table 6.3 contains the independently significant variables in the left column and the non-significant variables in the right-hand column, each listed in alphabetical order. Sixteen variables were significant in the univariate analysis ($p<0.10$), while one variable was non-significant. Only one technique variable was removed at the screening phase, while 17 background variables were removed in their screening phase, implying that the technique data were more associated to the response variable.

As noted in Chapter 5, the 'technique' variables could not include respondents' thoughts as my analysis of respondents' interpretations took place long after their interpretations had been completed. Hence, since the 'technique' variables were based on what I could see on the seismic image, the techniques were, to some extent, 'related' to the Max RE Score. However, the Max RE Score was based on the amalgamation of key features, chosen by the reference experts. Hence, the impact of the Max RE Score and techniques being related was minimal. For example, even if respondents used a technique and unknowingly interpreted an additional key feature in doing so, it would only inflate their Max RE Score by one.

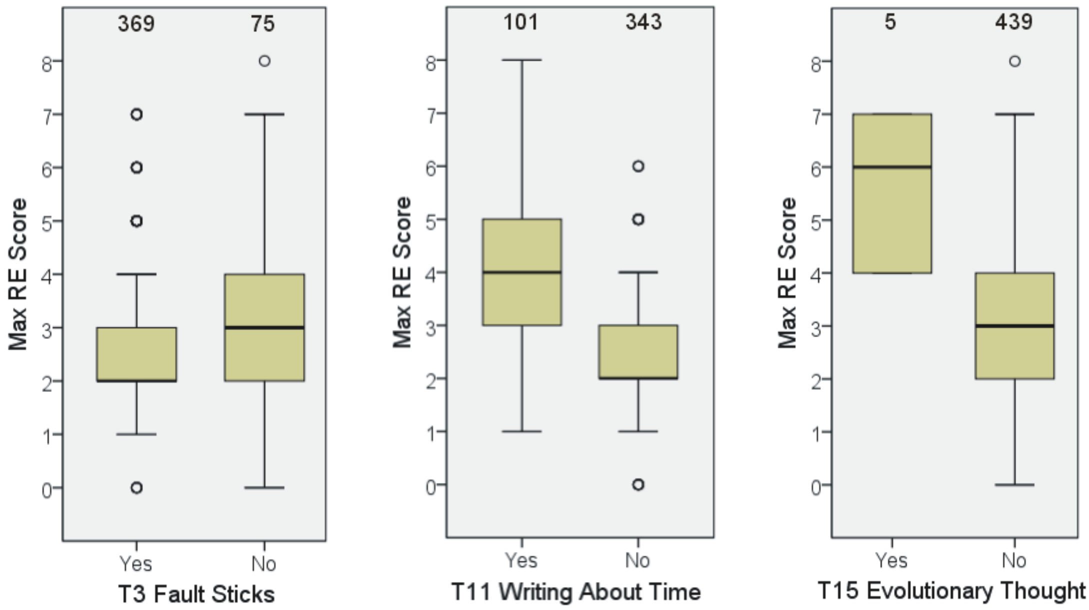
The non-significant variable in Table 6.3 was 'level of detail', showing that the 'level of detail' variable was not significantly associated to respondents' Max RE Scores. Although the Max RE Score was capped at 9 key features, this did not force the 'level of detail' variable to be non-significant since the reference experts could have picked 'multiple horizons' or 'multiple faults' as a single key feature. Expert 2 picked multiple geological features to be one key feature by choosing: 'multiple faults on right of horizon 1' as one of their key features (section 4.4.1). The reference experts, in general, focussed on large-scale features that influenced the whole seismic image, which might explain why the 'level of detail' was not significantly associated to the Max RE Score.

Table 6.3 – The results of the univariate screening of the technique variables. The independently significant technique variables are in the left column while the non-significant variables are in the right column.

Significant variables (p<0.10) (will be added to multivariate analysis)	**Non-significant variables (p=>0.10)** (removed from analysis)
Descriptions or explanations added?	Level of detail?
Annotations or labels added?	
Fault sticks drawn?	
Faults drawn with non-planar geometry?	
Arrows drawn on faults?	
Horizons drawn in?	
Packages coloured or drawn in?	
Horizons ordered?	
Cartoon drawn?	
Written about geological time?	
Written about geological processes?	
Concept explicitly stated?	
Questions stated?	
Geological evolution stated?	
Justified interpretation (without use of geological evolution) stated?	
Sequence stratigraphy used?	

Figure 6.4 presents boxplots of the Max RE Score for three technique variables, chosen to illustrate the differences in Max RE Score between the respondents who used, and did not use, each of the techniques. The use of ‘fault sticks’ was associated to lower Max RE Scores. ‘Fault sticks’ were faults that were drawn with straight lines and no curvature. Using ‘writing about time’ and ‘evolutionary thought’ both resulted in greater Max RE Scores, in general. For example, the median Max RE Score for ‘writing about time’ was key 4 features, while the median for evolutionary thought was 6 key features. The respondents who used these two techniques interpreted more key features than respondents who did not use them.

Figure 6.4 – Boxplots showing the distributions of Max RE Score for three techniques ('fault sticks', 'writing about time' and 'evolutionary thought'). The use of 'fault sticks' was negatively associated to Max RE Scores, while 'writing about time' and 'evolutionary thought' were positively associated to Max RE Scores.



6.4.4. Multivariate analysis of 'technique' variables: why were they effective?

The independently significant technique variables from Table 6.3 were then analysed in a multivariate analysis, the results are shown in Table 6.4. Six variables were significant in the multivariate analysis ($p<0.05$), while 10 variables were non-significant. For this statistical model, the AIC value was 430.3, which was much smaller than the AIC value for the first model (854.8), indicating a much improved goodness of fit. The Nagelkerke pseudo R^2 value also increased to 0.357 showing that more of the variation in the Max RE Score was explained by this model than by the first model.

Table 6.4 – The results from the multivariate analysis of the technique variables. Variables are listed in order of decreasing odds ratios. Confidence intervals (CI) for the odds ratios are also noted. The AIC value was 430.3, and the Nagelkerke pseudo R^2 value was 0.357.

Variable name	Technique type	P-value	Odds ratio (OR) with 95% CI
Written about geological time? *"Yes" to "No"*	Time	<0.001	4.06 (2.35–7.00)
Concept explicitly stated? *"Yes" to "No"*	Concept	0.003	2.77 (1.41–5.45)
Written about geological processes? *"Yes" to "No"*	Time	<0.001	2.51 (1.49–4.22)
Faults drawn with non-planar geometry? *"Yes" to "No"*	Drawing	<0.001	2.02 (1.39–2.92)
Arrows drawn on faults? *"Yes" to "No"*	Drawing	0.006	1.77 (1.18–2.66)
Sequence stratigraphy used? *"Yes" to "No"*	Concept	0.036	1.62 (1.03–2.54)

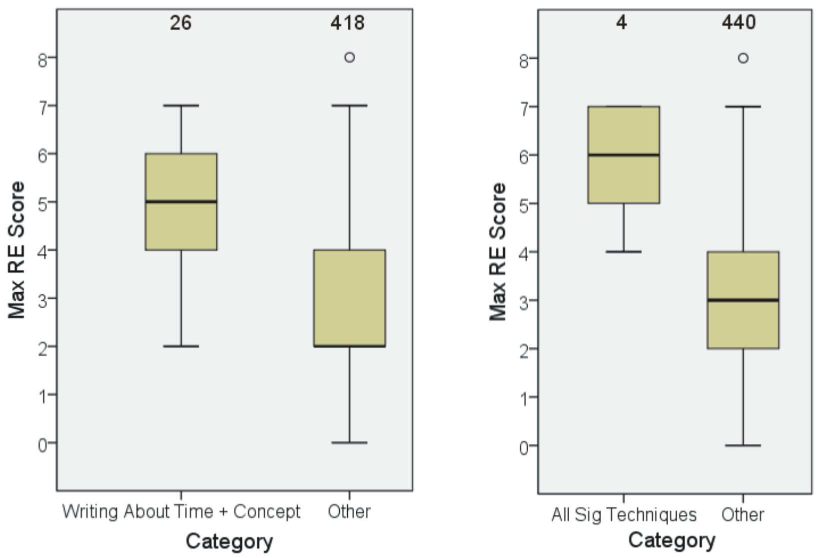
The technique variables that were not multivariately significant were as follows (p=>0.05):

Descriptions or explanations added?, Annotations or labels added?, Fault sticks drawn?, Horizons drawn in?, Packages coloured or drawn in?, Horizons ordered?, Cartoon drawn?, Questions stated?, Geological evolution stated?, Justified interpretation (without use of geological evolution) stated?

'Writing about time' had the highest odds ratio of 4.05 (95% CI: 2.35–6.99) in the multivariate analysis, showing that it was the most influential variable within the model. The variable that had the second-largest odds ratio was 'concept explicitly stated'. However, 'writing about time' had a much larger odds ratio than 'concept explicitly stated', 4.05 compared to 2.77 (95% CI: 1.41–5.46). There was no dominant 'type' of technique as the 'concept', 'drawing' and 'time' variables were each represented by two variables. 'Evolutionary thought' was one of the most significant variables in the univariate analysis but became non-significant in the multivariate analysis, possibly because only five geoscientists had used it, too small a number to improve the fit of the model. The difference between 'writing about time' and ‘evolutionary thought’ was that evolutionary thought only included changes in the structural geology over the whole section, while writing about time included local features based on geological time. As in Chapter 5, ‘evolutionary thought’ was a subset of ‘writing about time’, meaning that the respondents who had used evolutionary thought had also necessarily written about time.

The boxplots in the left side of Figure 6.5 show the effect of using the ‘writing about time’ and ‘concept explicitly stated’ techniques together against all other respondents on the Max RE Score. The median Max RE Score for respondents who used both techniques was 5 key features, well above the median for all other respondents, which were 2 key features. The boxplots on the right show the effect of using all six multivariately significant techniques from Table 6.4, which only 4 out of the 444 respondents did. The use of all six significant techniques was strongly associated to the response variable in a positive manner. The median for these four respondents was 6 key features, which was much larger than the median of 3 key features, for all other respondents. Overall, this showed that respondents who used these combinations of techniques in their interpretation tended to attain higher Max RE Scores than the other respondents.

Figure 6.5 – The boxplots on the left show the distribution of Max RE Scores for those respondents who used the techniques of ‘writing about time’ AND ‘concept stated’. The boxplots on the right show the distribution of Max RE Scores for those respondents who used all six of the significant techniques. Both combinations of techniques were positively, and very strongly, associated to the Max RE Score variable.



Overall, I found that the 'experience' variables had the largest effect when considering only respondents' backgrounds. I observed that fewer technique variables were removed at the screening phase than background variables (1 variable compared to 17 variables). 'Writing about time' was the most influential technique variable, having an odds ratio of 4.05. The odds ratio of 4.05 is larger than the largest 'background' odds ratio, which was 3.85, for 'structural geology specialists' compared against respondents with a 'basic working knowledge'. So, will the techniques remain significant once respondents' backgrounds are taken into account or are the techniques a by-product of experience?

6.4.5. Multivariate analysis of 'background' and 'technique' variables: is it respondents' experience or the techniques used?

This section determines whether the techniques remain significant when respondents' backgrounds are also considered, to determine what has most impact on producing a similar interpretation to the reference experts: education and professional background or the techniques used. I included all independently significant variables (left-hand columns of Table 6.1 and Table 6.3) in a separate multivariate analysis. As before, I used the backwards stepwise procedure to remove one variable at each step. The final model contained the variables in Table 6.5. Nine variables were significant in this multivariate analysis, while 18 variables were non-significant.

The AIC value was 1004.1, meaning that this model fits the data worse than the first two models (854.8 and 430.3), possibly because of the greater number of predictor variables in the equation, which are penalised by the AIC measure. However, in this model the Nagelkerke pseudo R^2 value increased to 0.439, which was the largest of the three models. Thus, the statistical model that contained only the technique variables explained the data most efficiently, i.e. the data is explained well by few predictor variables. While this model, with respondents' backgrounds and techniques, is best for making predictions since it explains more of the variance in the Max RE Score.

Table 6.5 – The results from the multivariate analysis of both the background and technique variables. Variables are listed in order of decreasing odds ratios. Confidence intervals (CI)

for the odds ratios are also noted. The AIC value was 1004.1, and the Nagelkerke pseudo R^2 value was 0.439.

Variable name	Background / technique type	P-value	Odds ratio (OR) with 95% CI
Written about geological time? *"Yes" to "No"*	Time	<0.001	4.46 (2.48–8.00)
Cartoon drawn? *"Yes" to "No"*	Drawing	0.022	3.76 (1.23–11.49)
Q8: Level of experience in structural geology? *"Specialist" to "Basic Working Knowledge"* *"Good Working Knowledge" to "Basic Working Knowledge"*	Experience	<0.001	3.25 (1.80–5.87) 1.20 (0.78–1.85)
Written about geological processes? *"Yes" to "No"*	Time	<0.001	2.70 (1.55–4.72)
Concept explicitly stated? *"Yes" to "No"*	Concept	0.017	2.34 (1.17–4.69)
Q10: How often seismic images are interpreted or used? *"Daily / Weekly" to "Yearly / Never"* *"Monthly / 6-Monthly" to "Yearly / Never"*	Experience	0.004	2.33 (1.38–3.95) 2.24 (1.27–3.95)
Arrows drawn on faults? *"Yes" to "No"*	Drawing	0.008	1.83 (1.17–2.87)
Q7: Background is mainly in super-major or major oil company? *"Yes" to "No"*	Work Environment	0.008	1.81 (1.17–2.79)
Q21: Number of geographical locations? *Per location, since a continuous variable*	Experience	0.022	1.04 (1.005–1.07)

The variables that were not multivariately significant were as follows (p=>0.05):

Q3: Completed a Ph.D. degree?, Q5: Number of years of experience (relating to geoscience) since completion of highest degree?, Q7: Background is mainly in a service company?, Q9: Level of experience in seismic interpretation?, Q19: Completed a structural geology course (not including University training)?, Q21: Geology has been investigated in more than one country?, Q21: Geology has been investigated in more than one continent?, Descriptions or explanations added?, Annotations or labels added?, Fault sticks drawn?, Faults drawn with non-planar geometry?, Horizons drawn in?, Packages coloured or drawn in?, Horizons ordered?, Questions stated?, Geological evolution stated?, Justified interpretation (without use of geological evolution) stated?, Sequence stratigraphy used?

The technique variables had larger odds ratios, in general, than the background variables in Table 6.5. The variable with the largest odds ratio overall was ‘writing about time’ (OR=4.46, 95% CI: 2.48–8.00), while the second highest odds ratio belonged to ‘cartoon drawn’ (OR=3.76, 95% CI: 1.23–11.49). The confidence interval for the ‘cartoon drawn’ variable was relatively large compared to the other variables, indicating uncertainty in the true value of the odds ratio.

During a descriptive comparison of the techniques (Chapter 5), I noted that sufficient ‘years of experience’ might be needed before it becomes easier to apply the ‘advanced’ techniques. A strong positive association between ‘evolutionary thought’ and ‘years of experience’ was evident upon inspection of the data. For instance, the five respondents who used ‘evolutionary thought’ had a median of 18 years of experience, while the main dataset had a median experience of 10 years. However, due to the fact that so few respondents had used the ‘evolutionary thought’ technique, the association was not statistically significant (p=0.886, Appendix 7). There was no association between ‘years of experience’ and ‘cartoon drawn’ (p=0.985, Appendix 7); and no association between ‘years of experience’ and ‘writing about time’ (p=0.684, Appendix 7). There was also no significant interaction effect in the main analysis, which would suggest that a strong relationship existed between ‘years of experience’ and using the ‘advanced’ techniques. Hence, there was no direct evidence to suggest that the advanced techniques could only be used by geoscientists with substantial numbers of years of experience, implying that lower ‘years of experience’ did not have to be a barrier to applying the most effective techniques.

The fact that the techniques of ‘writing about time’ and ‘drawing a cartoon’ were still significant, in addition to the background variables, showed that these techniques assisted respondents in gaining a high Max RE Score, regardless of their education and professional backgrounds. Hence, inexperienced geoscientists could produce valid interpretations if they use these particular techniques. However, an ‘experience threshold’ might exist where geoscientists become able to apply the techniques. The third largest odds ratio belonged to ‘structural geology experience’, which was also the most significant ‘background’ variable. Again, the ‘experience’ variables had larger odds ratios than the ‘education or training’ variables and the ‘work environment’ variable.

‘Cartoon drawn’ and ‘background in a super-major or major oil company’ were both significant in this analysis (Table 6.5) even though they were non-significant in the separate multivariate analyses (Table 6.2 and Table 6.4). ‘The number of geographical locations where the geology had been investigated’ remained significant with an odds ratio of 1.04 (95% CI: 1.005–1.07) per location. For example, if a respondent had investigated the geology in 20 geographical locations, the associated odds ratio would be 2.19 ($1.04^{20}=2.19$). In all cases, the associations were positive, meaning that additional experience or techniques used increased the odds of attained a higher Max RE Score.

6.5. Conclusions

The multivariate analyses in this chapter showed that:

Background variables

- The ‘experience’ background variables were generally more associated to the response variable than the ‘education or training’ and ‘work environment’ variables.
- Within the background variables, ‘structural geology experience’ was the most associated to the response variable (OR=3.85).
- ‘Structural geology experience’ was more associated to the response variable than ‘seismic interpretation experience’. Although, the frequency with which seismic images were interpreted or used was also significant.

Technique variables

- The 'time' technique variables were generally more associated to the response variable than the 'writing', 'drawing' or 'concept' techniques.
- Within the technique variables, 'writing about time' was most associated to the response variable (OR=4.05).

All variables

- The technique variables were generally more associated to the response variable than the background variables (larger odds ratios in final model).
- 'Writing about time' had the largest odds ratio overall (OR=4.46), compared to the background and technique variables in the final model.

These data suggest that the respondents who attained the highest Max RE Scores were those who were experienced in structural geology, who interpreted seismic images often and who had a wide range of experience (investigated the geology in many geographical locations). In addition, the two techniques that were most effective were 'writing about time' and 'drawing a cartoon'. The final model (Table 6.5) contained a mixture of both background variables and techniques.

The two 'goodness of fit' indicators gave different models as being 'best'. These indicators were not important as I was more interested in determining the significant relationships within the data. The statistical model containing only the technique variables was the most efficient description of the data (AIC=430.3), while the model containing the background and technique variables described more of the variance in the Max RE Score (pseudo R^2=0.439). The value of 0.439 shows that more than 40% of the variation in the Max RE Score can be directly explained by the final model. However, this means that 56.1% (1 – 0.439 = 0.561) was still unexplained[11]. The final model is therefore not effective at predicting the Max RE Score for new geoscientists who might interpret the seismic image.

To improve the predictive power, more data might be useful. These results show that experience is generally more important than education and the work environment of respondents, and that certain interpretational techniques were

[11] It is common for statistical models of real phenomenon to have moderate R^2 values, e.g. 40-60%, in my experience.

especially effective. Hence, collecting further detail on respondents' experience and investigating the significant techniques would be beneficial. For example, a further survey could include specific questions about respondents' experience, or respondents could be interviewed after the survey to identify what techniques respondents thought they were using. However, predictive power is fundamentally limited due to 'unknown unknowns', which by definition cannot be eliminated with further data collection. The relationships within these data still exist, are meaningful, and require further investigation.

It was interesting that 'evolutionary thought' was non-significant in the multivariate analysis even when the median Max RE Score for respondents who had used the technique was 6 key features. Other respondents may have used evolutionary thought in their mind, but only five respondents left clear evidence of their thinking. As such, the variable was not significantly related to the Max RE Score. The most effective technique was 'writing about time'. However, 'considering the geological evolution' was a subset of 'writing about time', which confirms that evolutionary thinking is effective.

Hence, there is strong evidence in these data that respondents who wrote about geological time gained higher Max RE Scores. However, it is not yet clear whether this relationship was causal, i.e. is writing about geological time just a feature of a good interpretation, or does it make participants interpret data better? The workshop experiment, described in the following chapter, tests for evidence of causality: *does considering the geological evolution of the seismic image cause respondents to interpret more key features in their interpretation?*

Chapter 7 – Industry Workshops: is considering the geological evolution effective?

While strong evidence exists that evolutionary thinking is associated to gaining a high Max RE Score, it is not clear whether evolutionary thinking causes geoscientists to produce better interpretations or is just part of a ‘good’ interpretation. The workshop experiment was conducted separately from the main survey, to determine how effective the technique of ‘considering the geological evolution’ really was. In this chapter, I outline the methodology used in this experiment, how the participants were recruited and what the results were. Furthermore, qualitative data were elicited from the workshop participants concerning how they approached the interpretation exercise and what they thought common practices in industry were.

7.1. Rationale for the workshop experiment

One of the main results gained in the analysis of the main dataset was that respondents who used the interpretational technique of ‘writing about time’ were almost 5 times more likely to produce a ‘more similar’ interpretation to the reference experts, when the other factors were held constant. For example, given two respondents with identical backgrounds, where one respondent wrote about time and the other did not; then the respondent who used the technique of ‘writing about time’ would be 5 times more likely to attain a higher Max RE Score than the other respondent who did not use the technique. The relationship between ‘writing about time’ and the Max RE Score variable was found to be highly significant ($p<0.001$). While this is very strong evidence, I wanted to test for evidence of causality rather than just association, and designed an experiment to do this via a workshop format.

7.2. Workshop methodology

The workshop experiment tested the effectiveness of ‘considering the geological evolution’ of the interpretation of a 2D seismic image, i.e. does ‘considering the geological evolution’ increase or decrease the Max RE Score attained by geoscientists? In addition, geoscientists’ opinions about the given seismic interpretation exercise and their interpretation workflows in general, were captured.

Group discussions were used to probe relevant issues. The following section explains how I set up the workshop experiment.

In the workshop experiment, one group of geoscientists was asked to interpret the seismic image in terms of geological evolution (and therefore, in terms of geological time), while the other group was just told to *“interpret the whole seismic image”*. I decided not to use the phrase ‘writing about time’ as it was less clear than ‘geological evolution’. The workshop experiment was effectively a ‘controlled trial’. Controlled trials are widely used in medical and pharmaceutical studies to determine whether a new treatment or drug is effective (e.g. Matthews, 2006). In this experiment, I tested whether the geoscientists who ‘considered the geological evolution’ attained a higher Max RE Score, on average, than the control group.

The group of geoscientists that were told to ‘interpret the whole seismic image’ will be referred to as the ‘control’ group, whilst the other group who were told to ‘interpret the whole seismic image by focussing on the geological evolution’ will be referred to as the ‘evolution’ group. The exact instructions that were given to each group were as follows:

- Control group: *“Please interpret whole seismic image”.*
- Evolution group: *“1. Interpret the whole seismic image. Please focus your interpretation on the geological evolution of the section. 2. Summarise the geological evolution below”.*

The geoscientists involved in the workshop experiment were completely separate from the main dataset. However, I added the control group into the survey respondents to increase its sample size; this was defined to be the ‘main dataset’ of 444 respondents. In the workshop experiment, the control and evolution groups were also independent of each other, i.e. participants did not know what the other group’s instructions were at the time of completing the seismic interpretation exercise, and each participant within the control and evolution groups completed the exercise on their own. All workshop participants completed the seismic interpretation exercise under identical conditions. The conditions I considered to be important were: ‘the amount of time’ and ‘the environment in which the exercise was completed’. The participants were told that they were attending a ‘seismic uncertainty workshop’ and given a brief introductory outline of the research project, but with no specific information on the study hypothesis. Hence, since the conditions were the same, if I observed a difference in the distributions of Max RE Scores for the control and evolution groups then it would be due to the different approaches used.

7.2.1. Recruitment of participants

I advertised that I was looking for workshop participants after my presentation at the DEVEX 2011 conference and on the GEO-TECTONICS email list (JISCMail). According to their website[12]:

> *"DEVEX is a well-established Aberdeen-based conference that caters for subsurface geoscience, engineering and drilling professionals who are actively involved in field exploitation and aims to foster cross-discipline co-operation and dialogue between the different disciplines".*

And, according to their website[13], the GEO-TECTONICS email list is:

> *"A forum for the discussion of all aspects of tectonics, structural geology and related disciplines. The list operates under the auspices of the Tectonic Studies Group of the Geological Society of London but is open to all".*

I decided to recruit participants from multiple companies to ensure diversity in the sample and to gain as large a sample as possible. A larger sample also makes the results more transferable to the underlying geoscientist population, e.g. we can be more confident that the results apply to the geoscientist population, in general.

Ultimately, four companies agreed to take part in the experiment: two oil and gas service companies and two independent oil and gas companies (one of these specialised in production only, while the other specialised in both exploration and production). I hosted the workshop three times in Aberdeen and once in London during September 2011. Senior managers were the point of contact within each company. I provided the senior managers with general information about the content of the workshop beforehand, but did not say what hypothesis I was testing. Companies were sent a promotional flyer in the weeks leading up to the workshop to

[12] http://www.devex-conference.org/ (last accessed: 22nd August 2012)

[13] https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=geo-tectonics (last accessed: 22nd August 2012)

attract as much participation as possible. In all cases, participation was voluntary and no payment was given.

Across the four companies, a total of 54 participants took part. However, 5 participants were deemed to be 'inexperienced' based on the filtering criteria and were excluded post-workshop from the analysis, leaving 49 participants to be analysed. The filtering criteria were: '<21 years old', 'no university degree', '<2 years of experience after the completion of their highest degree', 'no experience in seismic interpretation' or 'no experience in structural geology'. Of the five excluded geoscientists, two had no experience in seismic interpretation and three had less than 2 years of experience after the completion of their highest degree. Table 7.1 summarises the participation at the four workshops. The largest number of participants came from company B (21), while the other companies contributed 11, 12 and 10 participants respectively.

Table 7.1 – An approximate company description and the number of participants in each workshop. One participant from each of companies A, B and C, and two participants from company D were excluded post-workshop due to insufficient experience.

Company	Approximate company description	Total number of participants	Number of eligible participants used in the analysis	Helper(s) used?
A	Service company for oil and gas industry	11	10	Yes (Lunn)
B	Service company for oil and gas industry	21	20	Yes (Shipton and Bond)
C	Independent oil and gas (production)	12	11	No
D	Independent oil and gas (exploration and production)	10	8	Yes (independent Ph.D. student in petroleum geoscience)

Participants were split into the control and evolution groups, at each workshop, by the senior manager. At the point when the senior manager was allocating people into the two groups, they did not know what hypothesis I was testing. Participants were told that the groups had been selected randomly beforehand. The groupings were made by the senior manager to ensure equal distributions of experience between the groups. I could not do this myself as I did not know the participants taking part. It is not known precisely how each senior manager approached the task of splitting the geoscientists into these groups, but they were told to keep the distributions of experience approximately equal and to take no other factors into account.

After the data collection, the control and evolution groups from each workshop were analysed together since I was not interested in differences between the companies, and because I said at the outset that I would not compare individual companies. The samples of participants from each company were also too small to make meaningful comparisons. I checked whether the control and evolution groups had approximately equal distributions of experience and discovered that the control group was slightly more experienced than the evolution group. The results in Chapter 5 showed that experience was positively correlated to the Max RE Score. Hence, I would expect the control group to attain slightly higher Max RE Scores, on average, assuming that the different interpretational approaches had no effect.

7.2.2. Ethics approval for workshops

I successfully gained ethics approval from the University of Strathclyde, which was needed since the workshop involved human participation. The completed ethics form can be found in the Appendix 6. At the start of the introduction to the workshop, I displayed the following points in a PowerPoint slide:

- *“Participation is voluntary; you are free to leave at any time without giving a reason.*
- *The paper questionnaires will be kept for research purposes.*
- *Responses copied from the paper questionnaires will be stored and analysed electronically.*
- *Researchers from the Universities of Strathclyde and Aberdeen, and from Midland Valley Exploration Ltd (the industry sponsor), will have access to the paper questionnaires and the associated electronic data.*
- *The data will be kept for as long as the research is ongoing and possibly for many years afterwards.*
- *If agreeable with all, an audio recording will be taken of the group discussion.*
- *The data will be used in Euan Macrae’s Ph.D. thesis and publications”.*

All participants agreed to be audio recorded during the group discussion, and no participants left before the end of the workshop.

7.2.3. Outline of the workshops

Table 7.2 gives an overview of the different parts of the workshop, while the instructions that I and the helpers used are detailed underneath.

Table 7.2 – Overview of the different parts of the workshop with a time allocation for each. The workshop was meant to take 1 hour and 45 minutes but in practice, had to fit into the time available on the day.

Part of workshop	Time allocated to each part (minutes)
Overview of the workshop	5
Background questionnaire (education and geological experience)	10
Seismic interpretation exercise (Freyja seismic image)	35
Post-interpretation questionnaire (about the seismic interpretation exercise and seismic interpretation in general.	20
Group discussion	15
Presentation of previous results with question and answer session	20

In practice, the lengths of the workshop varied between 1.5 and 2 hours. I had to adjust the lengths of my presentation and the group discussion to ensure that all participants had equal amounts of time to complete the interpretation exercise. If time allowed, I increased the length of the group discussion and the question and answer session. I and the helpers used the following instructions for the workshops:

1. Split participants into two groups with at least 1 metre of separation. Participants should not be able to see the other group's instructions.

   (Explain that participants were randomly assigned to their groups and that the reasons behind having two groups will be explained at the end of the workshop. Make sure that participants don't know how they were grouped).

2. Show the introduction and ethics slides (**5 mins**).

   (Adjust lighting in room if necessary).

3. Instruct participants that there should be *"no talking"* from now on as the exercise is individual (and not a team exercise); but that participants can put their hands up if they have any questions.

4. Hand-out the first questionnaire (background questionnaire).

5. Wait until the first questionnaire has been completed by all participants (**10 mins**).

(Remind participants to complete all questions).

6. Hand-out seismic interpretation exercises to each group (Freyja seismic image).

(Inform participants that each set of worksheets has been given an anonymous ID number and it is very important that all their sheets have the same ID).

7. Wait until the seismic interpretation exercise has been completed by all participants (**35 mins**).

(Remind participants that it is an individual exercise and there should be no talking).

(Check that questionnaires and interpretations have the same IDs)

8. Hand-out the post-interpretation questionnaire.

(Remind participants not to change their seismic interpretation from this point onwards).

9. Wait until the post-interpretation questionnaire has been completed by all participants (**20 mins**).

10. Ask participants to fold their A4 questionnaires into their A3 seismic image, and to check IDs.

11. Say to participants:

"The group discussion is next. I would like to say before we start that we are interested only in your methodology and the thought processes which were used during the seismic interpretation exercise; rather than the actual interpretations. There are a number of valid interpretations and approaches so it is not important whether your interpretation is similar or different to anyone else's. Please feel free to speak freely".

12. Ask participants whether it is okay to audio record the group discussion.

(Explain that the recording is only useful to me because I can't note down everything that has been said during the discussion).

If agreed with all, then ask participants to get as close as possible to the front.

13. Have the group discussion (**15 mins**).

    (Refer to the questions in the post-interpretation questionnaire).

14. Thank participants for their time, collect in the worksheets and ensure that each set has the same ID. Collect in pens and pencils.
15. Give presentation of previous results and answer questions (**20 mins**).
16. Thank participants again.

It should be emphasised that I did not reveal what hypothesis was being tested until the group discussion. The background questionnaire asked the same questions that the respondents in the main dataset answered. The interpretation exercise was printed with the two different sets of instructions beforehand (one set of instructions for each group). The post-interpretation questionnaire concerned the 'interpretation of the seismic image' and 'seismic interpretation in general'. The groups received slightly edited versions of the post-interpretation questionnaire to reflect their differing interpretation exercises.

See Appendix 6 for the following five worksheets used in each workshop:

- Workshop background questionnaire (same for both groups).
- Interpretation exercise for evolution group.
- Interpretation exercise for control group.
- Post-interpretation questionnaire for evolution group.
- Post-interpretation questionnaire for control group.

Each set of worksheets had a unique participant ID noted on it to ensure that each set belonged to the same geoscientist. I prepared the IDs while participants were filling in the background questionnaire and then checked that worksheets were handed out to the correct individual as I went through the workshop. All worksheets (background questionnaire, seismic interpretation exercise and post-interpretation questionnaire) were collected in before the group discussion started. Also, in three of the four workshops, the helpers assisted with worksheet distribution and observed the participants during the interpretation exercise. On two occasions the helper(s) were my supervisors, while on one occasion the helper was a Ph.D. student in petroleum geoscience from a UK University.

7.3. Data preparation and checks

The following section explains the checks that were applied to the data to ensure the results were robust.

7.3.1. Were the control and evolution groups equally experienced?

It was important to ensure that the geoscientists in the control and evolution groups had similar distributions of experience otherwise any results might have been due to the difference in experience rather than the interpretational approach used. For the control and evolution groups, I therefore compared the demographic data (age and gender), ‘years of experience after highest degree’, ‘experience in structural geology’, ‘how often you interpret seismic images’ and the ‘number of locations where the geology has been investigated’. The values obtained for the main dataset, of 420 respondents, are added for reference. The mean values for the gender proportion and age category proportions, for the four geoscience organisations (see section 3.7.2), are also added for reference. The number of geoscientists in each category is noted with the ‘valid percentage’ beside it. The valid percentage does not include the missing data (e.g. non-response to questions) in the grand total of the denominator, which allows a fair comparison across variables that have different numbers of missing data.

Table 7.3 presents the proportions of males and females in the control and evolution groups, main dataset and geoscience organisations. The gender proportions were within sampling error of the proportions for the main dataset and the mean of the geoscience organisations. In all cases, the percentage of male geoscientists was about 80%.

Table 7.3 – Gender proportions for the workshop groups and main dataset. The mean proportions for the four geoscience organisations are also given as a comparison. The workshop groups were an excellent match to each other, and were a reasonable match to the main dataset and four geoscience organisations.

Q1: Gender proportions	Control group	Evolution group	Main dataset	Mean % of geoscience organisations
Female	20.8% (5)	24.0% (6)	21.3% (87)	18.9
Male	79.2% (19)	76.0% (19)	78.7% (322)	81.1
Missing	0	0	11	
Total	100% (24)	100% (25)	100% (420)	

Table 7.4 shows the age distribution for the control and workshop groups, main dataset and the mean of the geoscience organisations. The control group was slightly older than the evolution group, and both groups were older than the main dataset, on average. The control and evolution groups, and main dataset, were both younger than the geoscience organisations.

Table 7.4 – Age distributions for the control and evolutions groups, and for the main dataset. The mode age category is highlighted in grey. The mean proportion from each age category in the four geoscience organisations is given for comparison.

Q2: Age category	Control group	Evolution group	Main dataset	Mean % of geoscience organisations
<21	0.0% (0)	0.0% (0)	0.0% (0)	1.2%
21-30	4.2% (1)	12.0% (3)	31.0% (129)	21.0%
31-40	16.7% (4)	36.0% (9)	24.5% (102)	19.8%
41-50	41.7% (10)	36.0% (9)	23.1% (96)	19.6%
51-60	37.5% (9)	16.0% (4)	17.5% (73)	22.7%
61+	0.0% (0)	0.0% (0)	3.8% (16)	15.8%
Missing	0	0	4	
Total	100% (24)	100% (25)	100% (420)	

Figure 7.1 and Table 7.5 show geoscientists' years of experience after the completion of their highest degree. The median experience of the control group was

6 years greater than the evolution group. Both groups had higher median experience than the main dataset. The distributions of the 'years of experience' for the workshop groups also seemed to be less skewed than for the main dataset, possibly because the main dataset over-sampled the 21-30 year old geoscientists, giving it its skewed distribution.

Figure 7.1 – Distribution of 'years of experience' (Q5) for the two workshop groups and the main dataset. The data are shown as percentages to allow meaningful comparisons to be made as the main dataset has more than an order of magnitude more geoscientists.

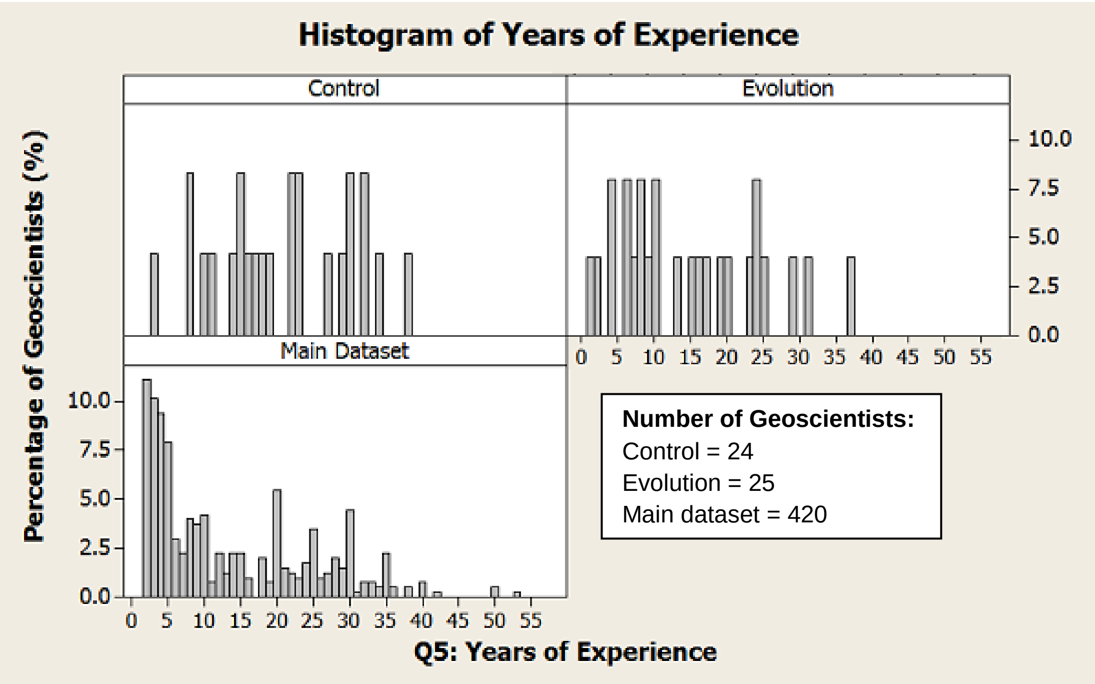


Table 7.5 – Summary data for Figure 7.1. It can be seen that the control group has, on average, 6.5 years more experience than the evolution group (median).

Q5: Years of experience	Number of geoscientists	Median (years)	Standard deviation (years)	Mean (years)
Control	24	20.5	9.4	20.7
Evolution	25	14.0	9.9	15.7
Main dataset	403	9.0	11.1	13.3

Table 7.6 shows that, on average, the control group was more experienced in structural geology than the evolution group as they had more specialists (4 specialists compared to 0). The evolution group was also less experienced in structural geology than the main dataset.

Table 7.6 – Distribution of experience in structural geology. The mode age category is highlighted in grey. The control group is more experienced in structural geology than the evolution group. The evolution group is also less experienced in structural geology than the main dataset.

Q8: Experience in structural geology	Control group	Evolution group	Main dataset
No knowledge	0.0% (0)	0.0% (0)	0.0% (0)
Basic working knowledge	41.7% (10)	48.0% (12)	34.1% (142)
Good working knowledge	41.7% (10)	52.0% (13)	48.4% (202)
Specialist	16.7% (4)	0.0% (0)	17.5% (73)
Missing	0	0	3
Total	100% (24)	100% (25)	100% (420)

Table 7.7 presents the responses to the question *"how often do you interpret or use seismic images?"* for the control and evolution groups, and the main dataset. It can be seen that the control group interpreted seismic images more frequently than the evolution group. Both workshop groups interpreted seismic images more frequently than the main dataset.

Table 7.7 – Distribution of experience in the frequency of seismic data interpretation. The mode age category is highlighted in grey. The three groups, in general, interpreted or used seismic images daily. The control and evolution groups interpreted or used seismic images more frequently than the main dataset.

Q10: How often you interpret seismic images	Control group	Evolution group	Main dataset
Yearly / Almost never	8.3% (2)	8.0% (2)	23.7% (98)
Monthly / 6-monthly	12.5% (3)	24.0% (6)	27.3% (113)
Daily / Weekly	79.2% (19)	68.0% (17)	49.0% (203)
Missing	0	0	6

Total	100% (24)	100% (25)	100% (420)

Figure 7.2 shows the distribution of the responses to Q21, the ‘number of locations where the geology has been investigated’ for the control and evolution groups, and the main dataset. The main dataset investigated the geology at a fewer number of locations than both workshop groups. The control and workshop groups had similar numbers of locations noted. Table 7.8 shows the summary data for Figure 7.2; the mean and median values were similar for the workshop groups. The respondents in the main dataset, on average, had worked at fewer geographical locations than the workshop participants.

Figure 7.2 – Histogram of the ‘number of locations where the geology had been investigated’ (Q21). The data are shown as percentages to allow meaningful comparison to be made as the main dataset had an order of magnitude more geoscientists than the workshop groups.

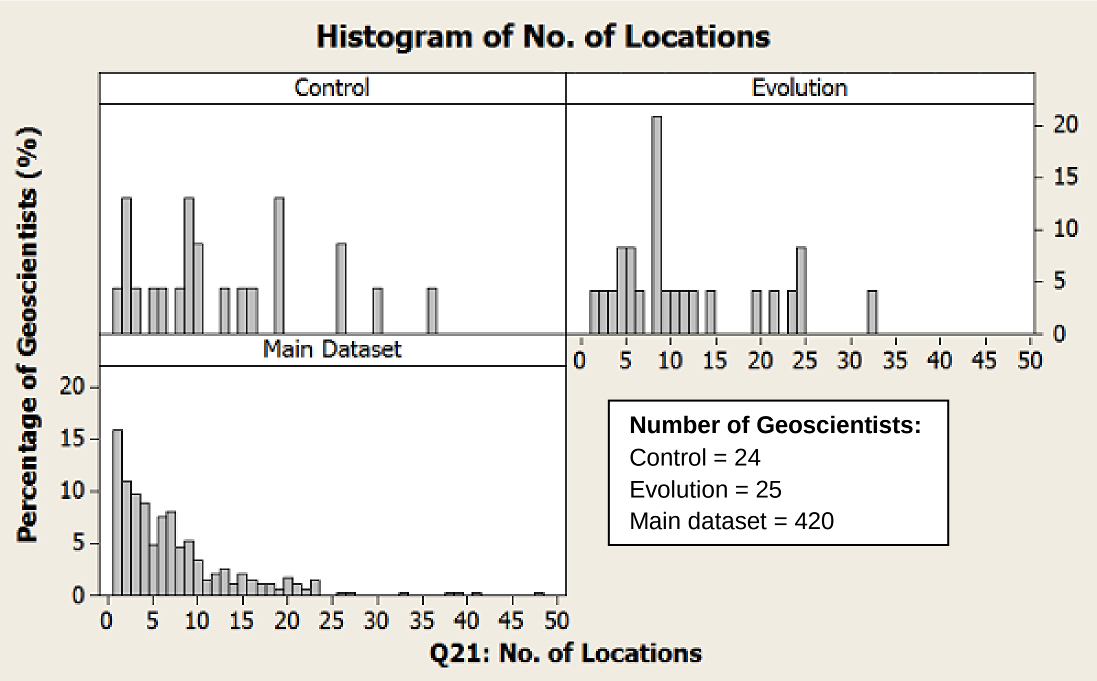


Table 7.8 – Summary information for Figure 7.2. The control and evolution groups had similar mean and median numbers of locations but both were larger for the workshop groups than for the main dataset.

No. of Locations	Number of geoscientists	Median (locations)	St deviation (locations)	Mean (locations)
Control	24	10	9.7	12.8
Evolution	25	9	8.3	12.2
Main dataset	346	5	6.9	7.2

The workshop participants, and each workshop group, were a representative sample of the underlying population. The comparison of the experience variables (‘years of experience’, ‘experience in structural geology’, ‘how often seismic images were interpreted or used’ and ‘number of geographical locations where the geology was investigated’) show that the control participants were more experienced than the evolution participants. Specifically, the control participants had more ‘years of experience’, more ‘experience in structural geology’ and ‘interpreted seismic images more often’. The only equal level of experience was in the number of geographical locations where the geology was studied. The workshop participants, as a whole, were found to be more experienced than the main dataset. The fact that the control group was more experienced than the evolution group arose randomly from the sampling and segregation of geoscientists.

In Chapter 5, I found that increasing the experience for each of the four experience variables (‘years of experience after highest degree’, ‘experience in structural geology’, ‘how often you interpret seismic images’ and the ‘number of locations where the geology has been investigated’) increased the likelihood that geoscientists would attain a higher Max RE Score. Hence, if the evolution group’s instructions had no effect, I would expect the control group to attain a higher Max RE Score, on average, than the evolution group as they were more experienced.

7.3.2. Prior knowledge of the interpretation exercise

Participants may have attended one of my conference presentations on this research and thus have more information than other participants, which potentially could have given them an advantage. However, during these presentations I did not reveal what the reference experts' interpretations were. I did briefly show four different respondent interpretations of the seismic image at the DEVEX 2011 conference (Aberdeen, 11-12th May 2011), but did not say which were more similar to the reference experts' interpretations. The following questions were presented underneath the seismic image to determine whether any participants had either seen the seismic image before or interpreted it, and the responses are shown in Table 7.9.

Have you seen this seismic image before?

Yes ☐ No ☐ Not Sure ☐ Where? ___________________

Have you interpreted this seismic image before?

Yes ☐ No ☐ Not Sure ☐ Where? ___________________

Did you see Euan's presentation at the DEVEX conference (11-12th May 2011 in Aberdeen)?

Yes ☐ No ☐ Not Sure ☐

Table 7.9 – The responses to the questions used to determine whether participants had seen or interpreted the seismic image before. The responses show that no participants had an unfair advantage.

	Have you seen this seismic image before?	Have you interpreted this seismic image before?	Did you see Euan's presentation at the DEVEX conference?
Number of participants who answered *"Yes"*	6.1% (3)	0.0% (0)	8.3% (4)
Number of participants who answered *"No"*	83.7% (41)	95.9% (47)	91.7% (44)
Number of participants who answered *"Not sure"*	10.2% (5)	4.1% (2)	0.0% (0)
Missing	0	0	1
Total	49	49	49

Of the three participants who had seen the seismic image before, all saw it at my DEVEX conference presentation in May 2011. One of them noted: *"don't remember the official interpretation"*. I confirmed after I had input the key features that these three participants did not interpret more key features than expected given their different levels of experience. Three of the five participants were *"not sure"* if they had seen the seismic image before. These three participants said that they thought that they had seen it in the: *"North Sea, Central Graben"*, *"Central North Sea / Norway?"* and "*Central Graben, UKCS"*. Interestingly, the three participants thought of geographical locations rather than where they had seen the seismic image before. It is important to note that no participants had interpreted the seismic image before and thus, were not respondents in survey sample. One of the four participants who saw my DEVEX presentation did not recognise the seismic image.

In summary, as no participants had interpreted the seismic image before, and as the three participants who had seen it before did not interpret more key features than expected, I deemed the workshop experiment fair as no participants had an unfair advantage.

7.3.3. Possibility of a 'stereotype threat'

In the second workshop, the senior manager mistakenly stated that they were moving individuals between groups to ensure that the groups had equal levels of experience. The geoscientists who had less experience and who were asked to move groups may, therefore, have been affected by a 'stereotype threat', where being labelled as *"inexperienced"* might have negatively affected their performance during the seismic interpretation exercise. See Steele and Aronson (1995) for examples of stereotype threats. The following was noted during the workshop by an observer (Shipton):

> *"Experience was highlighted as a factor by our contact, who made a point of telling people that they were being moved groups to even up the experience levels (therefore a risk of stereotype threat? Need to check against other groups)".*

To check whether stereotype threat had affected any of the participants, I compared the Max RE Scores attained by the geoscientists that may have been affected to what would be expected by individuals with similar experience profiles, based on results from main dataset. I decided to make '10 years of experience' the cut-off point, after which point geoscientists would be unlikely to be affected by being labelled as *"inexperienced"*. Ten years was reasonably high as geoscientists in industry are usually promoted from a 'junior' position after 5 years, based on my own experience from talking to industry personnel. If there was any influence of the stereotype threat then it would affect the least-experienced participants most.

There were 5 participants in the second workshop with less than 10 years of experience, who therefore might have been negatively affected. Of these five participants, four participants attained a Max RE Score similar to what would be expected, based on their answers to the background questionnaire. The fifth participant, who belonged to the control group, only gained a Max RE Score of 1 key feature, while they were expected to attain 2 or 3 key features. Hence, if the stereotype threat exists, the impact would be limited to attaining a Max RE Score of 1 or 2 key features less. The effect is negligible since the control group was more experienced than the evolution group.

In the other three workshops, geoscientists were split into the control and evolution groups without knowing that some of them had been moved groups to equalise the experience between the groups.

7.3.4. Inputting the data

I input the questionnaire and key feature data in the same manner as for the survey sample, where words and phrases were included as part of the interpretation. Some evolution participants mentioned key features in their explanation underneath the seismic image, but not on the seismic image itself. However, evolution participants might have annotated the seismic image in more detail if they had not been asked to explain their evolution. From the participants' perspectives, it was not essential to annotate the seismic image since explaining the geological evolution covered much of the same information. Only 5 of the 24 participants in the control group used the extra space to expand upon their interpretations, and one of these 5 participants also explained their geological evolution. Twelve of the 24 control participants did not apply any writing techniques and instead only drew faults, correlated horizons and coloured-in sediment packages.

7.4. Quantitative results

The workshop experiment was designed to be fair, with the control and evolution groups interpreting the seismic image independently. The control group had slightly more experience than the evolution group, and no participants had prior knowledge of the hypothesis that I was testing. There was also no evidence to suggest that some of participants in workshop 2 were affected by a potential stereotype threat.

The statistical analysis of the main dataset strongly suggested that 'considering the geological evolution' and attaining a high 'Max RE Score' were related to each other. This separate workshop was designed to determine whether there was evidence to suggest that instructing participants to *'focus on and state the geological evolution'* caused them to interpret more key features, i.e. did evolution participants attain a higher Max RE Score, on average? Figure 7.3 shows the results and Table 7.10 presents summary information for the distributions. The Max RE Score distribution of the main dataset is given for reference.

Figure 7.3 – The distribution of Max RE Scores attained by each workshop group and by the main dataset. The evolution group attained a higher Max RE Score than the control group, and the main dataset, on average. Table 7.10 contains summary data for the distributions.

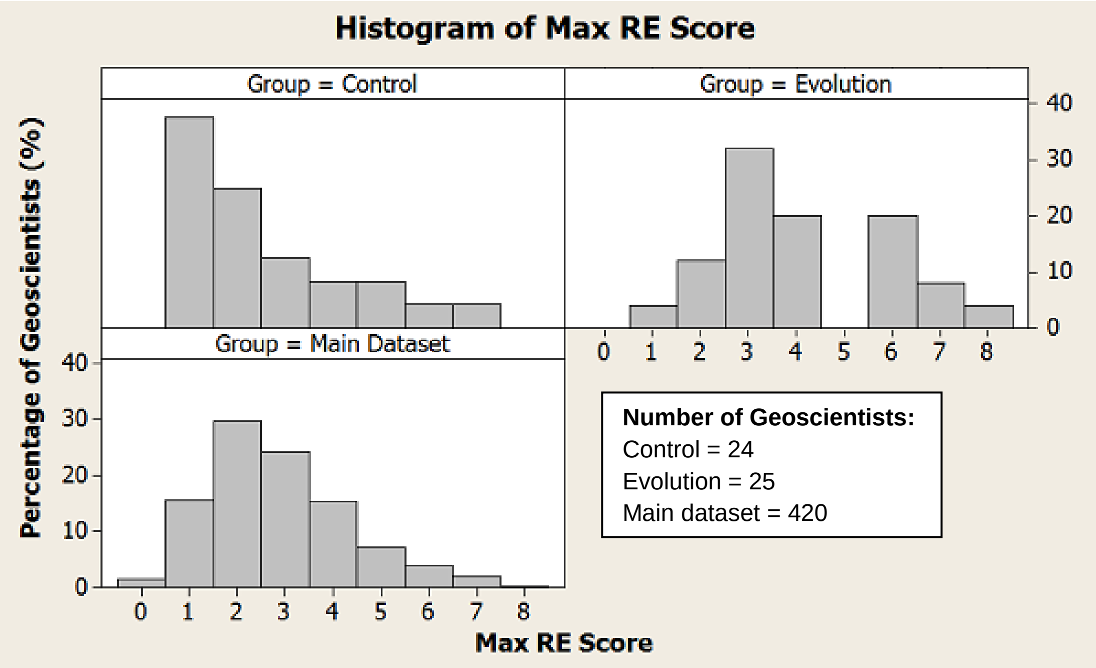


Table 7.10 – Summary information for Figure 7.3. The mean Max RE Score for the evolution group was much higher than for the control group and main dataset, suggesting that 'considering and stating the geological evolution' is an extremely effective technique.

	No. of geoscientists	Total no. of key features	Mean no. of key features	St. deviation	Median no. of key features
Control group	24	61	2.54	1.77	2
Evolution group	25	103	4.12	1.86	4
Main dataset	420	1198	2.85	1.49	3

Thus, the evolution group, though less experienced, achieved a much higher mean 'Max RE Score' than the control group; this result is highly significant (p=0.001, Appendix 7). The median number of key features interpreted by the control group was 2, while the median number for the evolution group was 4. It can be seen that no participants in the evolution group interpreted 5 key features, this is due to chance and in a larger dataset it would be likely that some of the evolution participants would attain a Max RE Score of 5 key features. The control group achieved a slightly lower mean Max RE Score than the main dataset, but the difference is within sampling error.

I used a 'two-sample Poisson rates' statistical test to determine whether the observed difference, between the Max RE Score for the control and evolution groups, occurred due to sampling chance or whether a relationship existed. I used a '1-sided' test as I was interested in whether the evolution group interpreted more key features than the control group. A 'two-sided' test would have been used if I was interested in a positive or negative difference, with no prior expectations. Based on the analysis of the main dataset my prior expectation was that the evolution group should attain a higher Max RE Score, on average, and these data confirm it.

The p-value for the observed difference was 0.001, which was highly significant (Table 7.11). There was thus, only a 0.1% chance that the observed difference occurred by chance alone. Additionally, the evolution group attained, on average, Max RE Scores 62% higher (4.12 / 2.542 = 1.62) than control participants. It was highly likely that a relationship exists between 'considering the geological evolution' of the seismic image and attaining a higher Max RE Score. (A 2-sided test gives the p-value to be 0.002, which is still highly significant). Within the underlying population of geoscientists, the estimated increase yielded by forcing geoscientists to 'consider the geological evolution' is 1.58 key features.

Table 7.11 – A two-sample Poisson Rates' statistical test comparing the mean Max RE Scores attained by the control and evolution groups. The evolution group attained a significantly higher Max RE Score than the control group (p=0.001). The estimated increase gained by 'considering the evolution' is 1.58 key features.

Group	Total occurrences	Number	Rate of occurrence (mean)
Evolution	103	25	4.120
Control	61	24	2.542
Difference = rate(Evolution) – rate(Control) = 4.120 – 2.545			
Estimate for difference: 1.578			
95% lower bound for difference: 0.723			
Test for difference = 0 (vs > 0): Z = 3.03, **P-Value = 0.001**			

In summary, the workshop experiment showed that the evolution group attained a significantly higher Max RE Score than the control group even though they were less experienced. The only other difference between the groups was that the evolution group was forced to consider and state the geological evolution of their interpretation. These data imply heavily that considering and stating their geological evolution improved participants' interpretations of the seismic image.

7.5. Qualitative data

The following two research questions are also important because knowing more about the interpretational approaches used by participants during the workshop exercise, and what they thought the common practices in industry were, will allow me to suggest useful and realistic recommendations. The results gained via the analysis of the main dataset and the quantitative workshop data imply strongly that certain interpretational techniques produce better results, regardless of interpreters' education, work environment and experience. This section does not compare individuals or companies and is based on the participants' views and my interpretation of those views. The research questions are as follows:

- *How did participants approach the workshop's interpretation exercise?*
- *What did participants think 'common practice in seismic interpretation' was?*

To investigate these two questions, I collected the following qualitative data. The format of the data is noted in parenthesis:

- Written responses to the 'post-interpretation' questionnaire (frequency tables and quotes).
- Observations made by myself and/or helpers (summary notes).
- Group discussions (summary notes).

A 'workshop report' was completed after each workshop, which included the observation notes and a summary of the group discussion. Participants were told that I was investigating the technique of 'considering the geological evolution' during the group discussion, which took place after the post-interpretation questionnaire. Hence, the fact that I was investigating the effectiveness of the technique of 'considering the geological evolution' may have influenced participants' answers during the group discussion, but could not have affected their responses to the post-interpretation questionnaire. I observed that some participants seemed to be slightly defensive in their answers during the group discussion, based on body language and tone of voice, perhaps trying to recall whether or not they had used 'geological evolution' in their interpretations. Each type of qualitative data are now explained in more detail.

7.5.1. Post-interpretation questionnaire (completed in a 20 minute slot)

The questionnaire contained two parts that investigated the research questions regarding 'how participants approached the interpretation exercise' and their views about 'common practices in seismic interpretation' (see Appendix 6 for the questionnaire and responses). To show how participants responded, frequency tables are used for some of the questions, while direct quotes are used for others. The differences in the number of participants in some of the questions were due to non-response to questions.

I chose not to use answer categories to capture participants' responses to allow participants freedom in how they responded. I instead left space underneath each question for participants' answers. However, this meant that participants' answers had to be grouped into categories to become statistically useful. For example, for the question:

> *"How often do you use a single line to build a template model (cartoon) to aid your 3D interpretation?"*

I grouped participants' responses into the following four categories: 'always', 'often', 'sometimes' and 'never'. As an example, workshop participant 1 answered this question by stating that: *"not often – need a broader context i.e. more than 1 line showing similar features"*; and hence, for this question, I put participant 1 into the *"sometimes"* category. I created appropriate answer categories for the other questions in the post-interpretation questionnaire and grouped participants' responses in the same way. It was generally straightforward to decide what category to choose for the responses.

7.5.2. Group discussion (completed in a 15 minute slot)

I consulted Greenbaum (1998) for details on how to set up and moderate the group discussions effectively. For example, Greenbaum (1998) advised that the moderator should prepare key questions beforehand to ensure that the desired information is collected, and advocated that 'open' questions should be used to avoid 'leading' the group. The group discussions were audio recorded and later summarised by me (see Appendix 6 for the summaries of the group discussions). In workshops 1, 3 and 4, the majority of the participants contributed to the group discussion and seemed at ease doing so. However, in workshop 2, most of the talking was done by about half of the participants (21 participants in total); maybe because the larger group inhibited some of the geoscientists from expressing their opinions. It was not logistically possible to record what statements were said by which geoscientists during any of the workshops.

As Greenbaum (1998) noted, it is important for the moderator not to lead the group to a particular answer by employing closed questions or by stating their opinions prior to asking a question. However, participants were told what hypothesis I was testing and that *"'considering the geological evolution' was the most effective technique in the main survey"* near the start of the group discussions. Thus, it is likely that some of the participants might have been more favourable to the technique of 'considering the geological evolution' because of this. I decided to reveal the workshop hypothesis at an early stage to gain participants' thoughts on how the technique could be applied effectively to industry since I only had 15 minutes available for the group discussion. It should also be noted that the questions about 'common practice in seismic interpretation' were fact-based and are hence unlikely to be affected by participants knowing the hypothesis I was testing.

Due to the noted time restrictions, I decided to prompt participants by asking key questions that investigated specific themes as this information was important to collect, making the group discussion a 'controlled' discussion. The responses, and my interpretation of them, are grouped under the themes. Seven key themes were selected and are stated below.

7.5.3. Observations

Observers were asked to note down anything 'interesting' during the workshop. There was no further explanation of what the observers were to look for, to avoid biasing the observers.

7.5.4. Seven key themes

The research questions were not asked directly. Instead, I based the questions for the post-interpretation questionnaire and group discussion on the seven themes below:

- Interpretational approaches.
- Team vs. individual work.
- Printing interpretations out.
- Interpretation without regional context.
- Multiple interpretations of a dataset.
- Geological evolution.
- Time pressure.

7.6. Qualitative results

I will present the responses to the two research questions separately, each under the seven themes above. The themes are noted under each research question, with the exception of 'team vs. individual work', 'printing interpretations out' and 'interpretation without regional context', which were not mentioned for the first research question ('participants' approaches in the interpretation exercise'). A summary has been provided after each question for clarity.

7.6.1. How did participants approach the workshop's interpretation exercise?

Most workshop participants said that they used a 'data-driven' approach during the interpretation exercise

A 'data-driven' approach was defined in the post-interpretation questionnaire to be:

> *"Interpreting the seismic image by identifying as many geological features as possible, which then builds up to a geological model".*

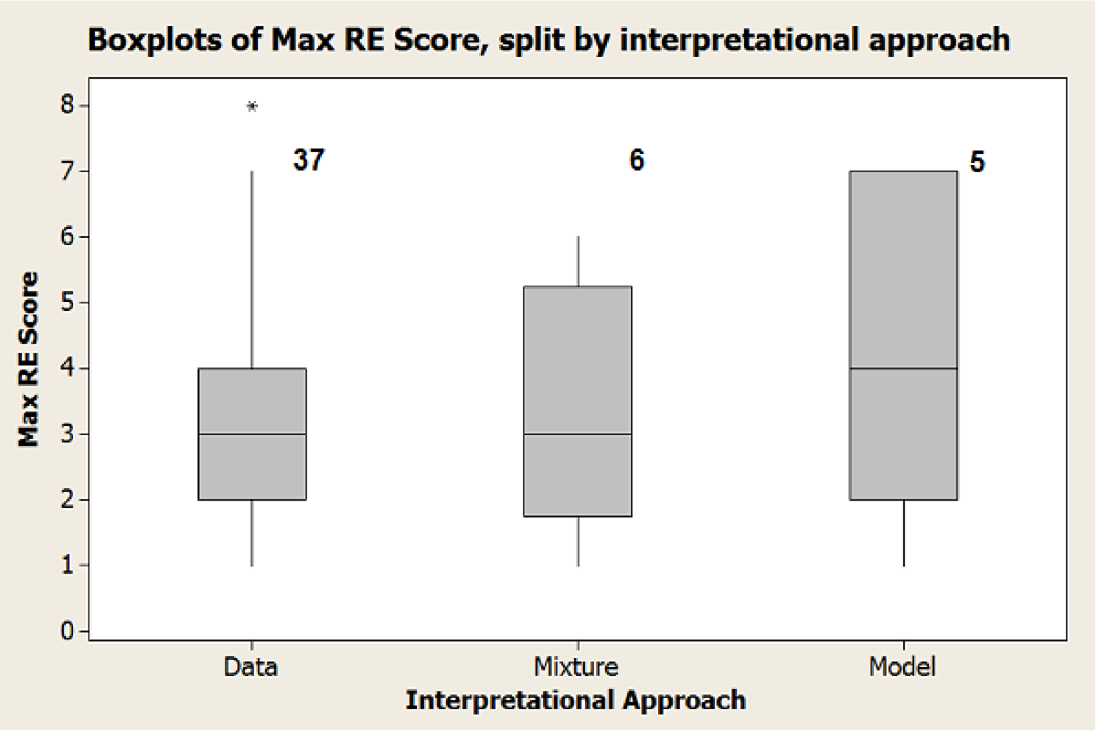
Conversely, a 'model-driven' approach was defined to be:

> *"Approaching the interpretation exercise with geological models in mind, and then deciding which fits the data best"*.

The results of the post-interpretation questionnaire showed that 77.1% of participants (37 out of 48) said that they approached the interpretation exercise with a 'data-driven' approach, 10.4% (5 out of 48) approached the exercise with a 'model-driven' approach, and 12.5% (6 out of 48) were both 'data and model-driven'. One participant noted that it was *"frustrating to have only 1 line"* and another that *"you can't truly consider strike-slip in a 2D line as it is a 3D problem"*.

It might be the case that 'model-driven' participants attained higher Max RE Scores than data-driven participants because model-driven participants were actively using geological knowledge. However, when tested, I found that model-driven participants, on average, attained a Max RE Score of 1 key feature more than data-driven participants (Figure 7.4). This result was not statistically significant (p=0.251, Appendix 7), perhaps because the sample has only five participants. Alternatively, the approach that geoscientists thought that they applied might not affect the likelihood of producing a valid interpretation in a direct way.

Figure 7.4 – Boxplots showing the distributions of Max RE Score attained by workshop participants, split by how they said they approached the interpretation exercise. Number of participants in each group is noted in bold. One participant did not complete the question.



Also, gaining a high 'level of detail' score was not necessarily the same as being 'data-driven' since participants who said that they approached the exercise 'data-driven' had, on average, the same 'level of detail' score as the 'model-driven' participants (a median of 24 features compared to a median of 26 features). However, it was clear from the main analysis (Chapter 6) that those respondents who focussed on interpreting many small features (e.g. who gained a high 'level of detail' score) did not generally attain high Max RE Scores.

Given that 'model-driven' participants did not attain higher Max RE Scores, on average, and that 'data-driven' participants did not gain higher 'level of detail' scores, participants may have found it difficult to judge which approach described them most accurately. On the other hand, it is possible that no correlation exists between participants' stated approach (as I had defined them) and the Max RE Score variable.

Multiple interpretations were not attempted by most participants

Only 38.3% of participants (18 out of 47) said that they tried multiple interpretations for the seismic image in the interpretation exercise. One participant wrote that they compared *“pure extension”* to *“early compression with later extension”*, which is a different tectonic concept (extension compared to inversion). On the other hand, one participant noted *“channels on the right side”* as their alternative interpretation, which is not necessarily a different tectonic concept but a different interpretation of a local feature.

The geological evolution was not stated by most of the participants

The evolution group were not asked whether they considered the geological evolution as it was part of their instructions. However, still only 56% of participants (14 out of 25) considered the evolution, while 9 participants explained or described their interpretation instead. Two participants missed out the question.

Twenty out of the 24 control participants said that they checked the geological evolution of their interpretation but only 1 participant left evidence of doing so. Thus, 19 of the 20 participants chose not leave evidence of evolutionary thinking or perhaps used a different definition than I had. For example, I defined geological evolution to be *“changes in large-scale structure over time”* and required clear evidence (e.g. a list of evolutionary steps or a series of diagrams); while some participants were presumably not as restrictive in their definition. Consider the following description given by a participant about how they worked out the geological evolution of their interpretation:

> *“[I] started off interpreting bright reflectors to determine where the faults might be, then interpreted stratigraphic markers, all the time thinking about their significance in relation to the geological model”*.

I interpret this quote to mean that the participant was more focused on local features than *“large-scale”* structures; therefore, this would be classed as ‘writing about time’ under my classification scheme.

Some of the control participants who did not consider the geological evolution gave the following reasons for not doing so: *“lack of time / size of paper”*, *“it will always look different the next time round”*, *“wasn’t sure what was required, e.g. full* *interpretation, simple or structural?"*, *"I couldn't think of anything sensible"* and *"ran out of time and lack of geological knowledge"*. The reasons given were diverse and imply that participants were unsure what to do and where to start. However, the fact that the evolution group considered the geological evolution in less than 30 minutes refutes the control group's reasons that related to 'not having enough time'. One participant said that they did not consider the evolution because it would *"look different the next time"*, which I interpret to mean that the participant did not believe that 'considering the geological evolution' would assist their interpretation.

Sixty-five point one percent of participants (28 out of 43) said in the post-interpretation questionnaire that they found it challenging to consider the geological evolution, and 44 out of 44 participants agreed with the statement that 'considering the geological evolution was beneficial to getting a valid interpretation'. One participant said in the group discussion that considering the evolution *"made me realise that it was not as simple as I thought it was"*. These responses were similar to the Ph.D. student's observations in workshop 4, where they observed the participants during the seismic interpretation exercise. The following notes were taken by the Ph.D. student:

- Control group (5 participants):
  "This group started colouring the seismic section immediately after starting. They looked quite confident at the beginning (in the first 5 minutes), then they started re-inspecting/checking and modifying bits of their interpretation. Another interesting observation was that the members of this group were the first to start interpreting and the first to finish. The first three participants to finish were from this group, whereas the 4th and 5th to finish were from the evolution group".

- Evolution group (5 participants):
  "The main observation I noticed, at the start of the interpretation exercise, was that the group took some time to think before starting to colour. Then, a few minutes after starting their interpretation, they stopped and looked back at the seismic section. They did this quite frequently and looked puzzled. Some members of the group seem to have been thinking deeply. During the last five minutes they looked like they were still unsure, or not convinced, of what they had interpreted".

I interpret the Ph.D. student's observations to mean that the evolution group found their exercise more challenging than the control group because they were forced to think about whether their interpretation was geologically valid by being required to explicitly state the evolutionary steps. The Ph.D. student said that the control group completed the interpretation quicker and seemed more confident than the evolution group, perhaps the overconfidence bias in action.

Interestingly, the Ph.D. student observed that:

> *"…the [evolution] group took some time to think before starting to colour. Then, a few minutes after starting their interpretation, they stopped and looked back at the seismic section".*

This approach sounds like a 'combined' data and model-driven approach; where the data are observed, then possible geological model are considered, and then the data are re-checked to confirm that the geological model does not contradict it. However, all four evolution participants in workshop 4 stated in the post-interpretation questionnaire that they used a data-driven approach.

Not all participants read the instructions fully

About half of the participants in workshop 2 admitted to not reading the instructions fully, perhaps due to the time pressure in their normal working environment.

7.6.2. Summary

The participants indicated that they mainly used a 'data-driven' approach during the exercise. Most participants did not consider multiple interpretations, or the geological evolution, of their interpretation. Nineteen of the 20 control participants that said they had considered the geological evolution left no evidence on their interpretation of having done so, and only 14 out of 25 evolution participants stated the geological evolution of their interpretation even though they were instructed to do so. I interpret the low 'usage' rates for the technique of 'considering the geological evolution' to be because participants did not define the technique in the same way I had.

The Ph.D. student's comparisons of the control and evolution groups identified differences in how the two groups approached the interpretation exercise. In general, the control group began their interpretation quickly and finished before the evolution group. The evolution group took longer to start their interpretations and appeared to be unsure of what they had interpreted. In general, all participants found it challenging to consider the geological evolution but thought that it was beneficial to getting a valid interpretation. The evolution group appeared to use the combined approach of interpreting the data, thinking of the implications with regard to the geological model and then checking the data again even though they all stated that they used a data-driven approach.

7.6.3. What did participants think 'common practice in seismic interpretation' was?

Participants seemed to be aware of the effects of anchoring bias

Participants seemed to be aware that anchoring bias, in the form of pre-conceived ideas, may influence their interpretation. For example, one participant said that *"once you have an interpretative model in mind it is hard to consider other scenarios"*. Similarly, another participant said:

> *"The first thing I looked for was, 'are the faults normal?' And that takes you down a path, doesn't it?"*

I interpret this quote to mean that 'expecting to see normal faults' could affect how the rest of the seismic image is interpreted. However, based on the following quote, one participant appeared to disagree that anchoring bias could have a large impact:

> *"[Given different interpreters' backgrounds] you might interpret the evolution in a different way, but I don't think you will interpret the seismic image differently to what you can see"*.

Based on this quote, it looks like the participant thought that the interpretation of 'what you can see' would be unambiguous, e.g. the interpretation of the observed data would be straightforward. The discussion about to what extent anchoring bias can influence interpretations was not conclusive, but the evidence suggests that participants had thought about it.

Seismic interpretation was mostly individual work, but ideas were discussed with colleagues

One participant said that they do *"a combination of individual and group work"*, and another said *"individuals"*. A different participant said that:

> *"We don't collaborate all the way through a project. After a meeting everyone goes away to develop their ideas before meeting up again".*

Since, in this case, geoscientists interpret data separately and then discuss their ideas together, herding bias might be mitigated and multiple interpretations of the dataset might be encouraged. However, it is not known how interpretation decisions are made and implemented through a project, or whether the individuals work on different parts of an interpretation between meetings. Hence, it is not clear to what extent herding bias is mitigated and multiple interpretations encouraged, based on these data.

Most participants said that they rarely or never printed out their interpretations, but still thought that they were constrained by working on a computer monitor

Seismic images are normally longer horizontally than vertically, which means that large monitors are needed to view them easily:

> E.g. *"In the days when we worked on paper, you could see a lot more in one 'view'. Monitors are not large enough to reproduce that"*.

The participants thought that it was easier to be collaborative around a printed interpretation than a computer monitor where only one person controls the computer (*"hard to be collaborative around a screen, easier to be collaborative around a print-out"* and *"you're at the whim of the person who has their hand on the mouse"*). However, some participants have not printed out seismic images for years, and one participant said *"always on a computer"*.

Participants also noted that for a complex area or a larger project, multiple printed lines can be displayed on the wall, giving an in and out-of-plane perspective. However, whether interpretations are printed out also depends on how well the area was known, and whether past work has been completed, e.g. *"is there well control?"* Some of the participants had facilities to look at seismic data on very large screens or in data rooms, often with multiple interpreters, e.g. *“more pairs of eyes are better than one”*.

The analysis of the main dataset showed that interpretational techniques were effective. However, it is not clear how easily the techniques can be applied to seismic interpretation without working on a printed seismic image. For example, ‘writing about time’ is likely to be easier to apply when using pen and paper than when industry standard interpretation software is used, which may not have the option to add annotations. Some of the participants also had facilities to look at seismic data in data rooms, which also allows different viewpoints to be considered, but risks herding bias.

About half the participants said that they interpreted seismic data without a regional context in their normal workflow

Forty-two point nine percent of participants (21 out of 49), when describing their normal seismic interpretation workflow in the post-interpretation questionnaire, wrote *“regional context”* (or equivalent phrase) first. Likewise, participants commented that: *“if we have other data, or know where the data are from, we always start with that”*, *“sometimes one has to, but not by choice”* and *“only if in wild-card basins or tests like this, but would try to relate to analogues”*. Hence, these comments implied that some participants prefer to use the regional information where possible. However, 53.2% of participants also said that they considered seismic data without the regional context ‘at some point during their workflow’. When asked how long into the workflow they usually wait before referring to the regional context, participants wrote *“rapidly – after brief interpretation”*, *“not long”* and *“as long as possible”*, showing variability in the responses. Interestingly, the participants also noted that the regional context may provide alternative ideas.

One participant noted that it is useful to know the depositional environment of the section to *“know what to expect”*, and another recognised that there were *“dangers with using pre-conceived ideas”*. These two quotes are contradictory because ‘knowing what to expect’ is a typical case of anchoring bias, e.g. the interpreter may find it hard to imagine other scenarios once they have decided what they think they should see. While, the second interpreter seemed to understand that pre-conceived ideas can negatively affect the interpretation.

Most participants said that they considered multiple interpretations in their normal seismic interpretation workflow

Sixty-one point nine percent of participants (26 out of 42) said that they routinely try out different interpretations of seismic images in their workflow, 19% (8) of participants sometimes do, and 19% (8) do not. Participants' comments on the post-interpretation questionnaire included: *"we tend to anchor on a preferred case and then it's hard to move away"*, *"if there are features that don't fit"*, *"when applicable (i.e. data has low confidence) but not routinely"* and *"only for undrilled prospects; try early on before spending too long on a wrong model"*. It was interesting that one of the participants knew that they anchored on a *"preferred case"* in their interpretations. The other participants said that they only considered multiple interpretations when they *needed to*, but not always. Hence, participants seemed to imply that they were reactive to the situation rather than proactive, e.g. instead of considering multiple interpretations in all interpretations, participants only did when they had to.

One participant said that:

> *"Most of the time you can very quickly eliminate many of the possibilities and focus on the best ones, even just by tying a couple of wells".*

Another said that it was *"difficult to think up other scenarios"*. Participants seemed to think that considering multiple interpretations was an effective technique in only some instances, and most participants said that they routinely tried different interpretations. In particular, one participant said that *"in a separate project, I've considered multiple interpretations recently for a contentious area"*. However, as noted above, some participants defined 'considering multiple interpretations' to be experimenting with localised features, e.g. *"sometimes we experiment with parts of the interpretation"*, rather than using distinct underlying tectonic models.

Another participant said that 'considering multiple interpretations' of the data is not appropriate in all investigations. For instance, the participant noted that their company mostly does oil field-related interpretation, often with hundreds of wells, and hence does not have the same uncertainty in the overall tectonic setting as an exploration focussed company might. The majority of participants in workshop 3 (12 participants) said that they were prompted by their line manager to consider multiple interpretations. The question was not asked at other workshops.

Most participants said that they considered the geological evolution in their normal seismic interpretation workflow

Eighty-four point one percent (37 out of 44) said that they regularly check the evolution of their interpretation in their workflow. When asked at what stage they considered the evolution in their workflows, some participants wrote *"throughout"* (or equivalent phrase), implying that it was not a separate step but implicit in their interpretations, while others wrote *"early on"*. One participant said that:

> *"[The] geological evolution is integral to the interpretation – that is what an interpretation is. You don't separate them"*.

Another participant asked rhetorically:

> *"When you talk about an interpretation, do you mean how you put wiggles on the image [faults, horizons] or do you mean the evolution of the geology?"*

The majority of participants in workshop 2 (21 participants) said that they were prompted by their line manager to consider multiple interpretations. The question was not asked at other workshops.

Participants said that they were often under considerable time pressure at work

Some participants said that time pressure affected how long they could spend using interpretation techniques such as 'multiple interpretations' and 'geological evolution'. For example, one participant said *"time pressure prevents you from considering multiple models in depth"*.

7.6.4. Summary

Seismic interpretation was mainly an individual exercise, seismic images were not often printed out, and most participants collaborated with their colleagues at some point during the workflow. Participants thought that collaboration was constrained by working on a computer because only one person had control and since it was difficult to look at full sized seismic images on the monitor. There was no agreement to whether considering seismic data without a regional context was beneficial.

Managers seemed to accept that considering multiple interpretations and the geological evolution of an interpretation was worth doing since they prompted their staff to use the techniques.

It was noted that in some investigations of seismic data, considering multiple interpretations of the data was not appropriate. There was an ambiguity in what ‘considering multiple interpretations’ and ‘considering the geological evolution’ actually meant. Most participants seemed to apply these techniques at a ‘local’ scale (e.g. specific faults) rather than on the entire dataset. It is likely that focussing on the local scale is easier and quicker than applying it to the full seismic image. No participants said that they were prompted by software to consider multiple interpretations or the geological evolution of their interpretation.

In addition, I have identified four challenges in relation to the current practices in seismic interpretation, based on my analysis of participants’ views.

7.6.5. Four challenges in seismic interpretation

- Few of the control participants (and respondents from the main dataset) explained their interpretation in any way at all, e.g. there were often no annotations, descriptions or explanations.
- Participants said that they rarely printed out seismic images even though they recognised the benefits and thought that working on a computer could be restrictive to collaboration.
- The techniques of ‘multiple interpretations’ and ‘geological evolution’ were not well defined since participants thought they had used the techniques but actually did not.
- There was little incentive for participants to consider multiple interpretations of the data, and to consider the geological evolution of their interpretations on their own, since mangers had to prompt them to use the techniques.

7.7. Conclusions

The analysis of the main dataset showed that the technique of ‘writing about time’ was the most significant variable, and those respondents who used the technique tended to attain higher Max RE Scores than other respondents, on average. This workshop experiment tested that finding directly by employing a ‘controlled trial’ experiment where the only difference between the two groups of participants was the instructions that dictated their interpretational approach. The workshop experiment was conducted separately from the main survey and involved different geoscientists. The control group was told to *"please interpret the whole seismic image"*, while the evolution group was told to '*focus on and state the geological evolution*' of their interpretation of the seismic image. There were 24 participants in the control group who were slightly more experienced than the 25 participants in evolution group.

The analysis found that the evolution group attained Max RE Scores that were 62% higher (p=0.001), on average, than the control group. Because the experiment was controlled, the observed increase in Max RE Scores was likely to be caused by the different interpretational approaches, rather than by another experimental factor. Thus, there is extremely strong evidence in these data that the technique of 'considering the geological evolution' directly improves the interpretation of the Freyja seismic image. However, these data provide no direct evidence that the increases in Max RE Score will be transferable to the interpretation of other seismic images. However, as none of the participants had interpreted the seismic image before, the technique of 'considering the geological evolution' is likely to improve all interpretations of unknown seismic images. Thus, geoscientists who use the technique in their interpretation will challenge themselves to produce interpretations that more likely to be geologically valid. The process of working out the geological evolution will also give geoscientists a better understanding of which alternative interpretations are valid.

However, the qualitative data revealed there was ambiguity regarding what 'considering multiple interpretations' and 'considering the geological evolution' actually meant since participants said that they used the techniques during the interpretation exercise, but left no evidence of having used them. It is therefore likely that the participants defined the techniques differently than I had. For the techniques to be implemented effectively by all geoscientists, they will need to be well defined to ensure there is no confusion in their definition.

Based on the analysis of the Ph.D. student helper's observations in workshop 4, the evolution group appeared to use a 'mixed' interpretational approach, e.g.

interpreting the data, thinking about the geological model, and then checking the data again. However, all four of the evolution participants in workshop 4 stated that they used a ‘data-driven’ approach instead. Hence, investigating geoscientists’ interpretational approaches by asking them to categorise themselves into either a ‘data’ or ‘model’ driven approach was not effective because participants found it difficult to know what approach, out of the options I had defined, they had used. In fact, it is likely that participants would have needed to use a combination of the approaches during the interpretation, to some extent, as observed by the Ph.D. student.

The group discussion also identified what participants thought ‘common practice in seismic interpretation’ was. Using this information, I identified four challenges in seismic interpretation practices. In the next chapter, I discuss the how my results affect the current understanding of how uncertainty in geoscience interpretation should be mitigated and suggest recommendations for addressing the four challenges.

Chapter 8 – Discussion

What makes a good interpretation? And, *what makes a good interpreter?* This chapter will aim to answer these two questions, based on the results and using the information gained from the qualitative workshop data. The three sets of data that have been analysed in this thesis are the reference experts' interpretations, the sample of 444 respondents and the workshop experiment data. Finally, I will discuss how geoscientists should be trained to interpret geological data in light of these findings.

8.1. What makes a good interpretation?

In this section, I discuss the components of a 'good' interpretation based on the results and on the literature review. A good interpretation is not necessarily precise but involves creative thinking and communication.

8.1.1. Interpretation without a regional context

Bentley and Smith (2008) noted that geoscientists anchor on 'base case' models in reservoir modelling, while Delfiner (2008) noted that analogues acted as anchors in prospect evaluation. It is therefore possible for geoscientists to anchor on the regional context during an interpretation, which may lead them to 'narrow down' the number of possibilities prematurely since the resulting interpretation will be restricted by the anchor. Likewise, to avoid biasing the interpretation, Bond *et al.* (2008) advised *"removing regional and tectonic context"* if the aim is to provide a range of interpretations, but did not formally test this hypothesis.

To test for anchoring bias in the main dataset, I compared the interpreted tectonic setting of the left fault to respondents' 'most-worked' tectonic regime (a possible anchor). However, the analysis showed no evidence of anchoring bias as in Chapter 3 (Appendix 2). For example, of the 31 respondents in the main dataset who ranked 'salt tectonics' as their most-worked tectonic regime, only 5 interpreted salt or shale in the seismic image. If respondents' interpretations had been anchored on their most-worked tectonic regime, I would have expected this number to be higher.

Other anchors may have existed that were not tested for, such as the 'areas of geoscience' where respondents had been most active (Q11). I decided not to test whether the areas of geoscience acted as an anchor because it was not possible to define how each of the areas would affect respondents' interpretations. Hence, my analysis for a specific type of anchoring bias was a fair test, but the result, that respondents' interpretations of the left fault were not anchored by their 'most-worked' tectonic regime, cannot be generalised since other anchors might exist that were not tested for. However, this research was primarily designed to identify the most effective 'type' of geoscientist and the most effective interpretational techniques rather than test for anchoring bias.

Creativity in geoscience interpretation is encouraged and confirmation bias is mitigated when the regional context does not precondition interpreters' expectations. Interpreters should thus approach the interpretation of geological data with an 'open mind'. Hence, providing the regional context encourages both anchoring and confirmation biases. It was not possible to test for confirmation bias in this research.

To mitigate anchoring and confirmation biases, regional knowledge should be used to qualify interpretations instead of being available at the start of the interpretation process. For instance, imagine that two competing interpretations of a seismic dataset exist; one where salt tectonics influences the basin, and the other where salt tectonics does not influence the basin. Regional information (that the seismic data were collected from a known salt basin) is then brought into the workflow. This new information would then qualify the salt-based interpretation since the interpretation would represent the geology better, making the non-salt interpretation less likely. The non-salt interpretation could still be possible; for example, if the part of the basin under analysis did not have salt. However, it would still be beneficial to have considered both interpretations at the start of the interpretation process since a wider range would have been considered.

8.1.2. Multiple interpretations of a dataset

The technique of 'considering multiple interpretations' of geological data is now well documented in geoscience literature (Bentley and Smith, 2008; Bond *et al.*, 2008; Rowbotham *et al.*, 2010; Chellingsworth *et al.*, 2011), but it is unclear how often the technique is used in the oil and gas industry. From informal discussions with geoscientists and the workshop group discussions, I found that the main barrier to considering multiple interpretations systematically was the additional time taken. I also found that some geoscientists saw considering multiple interpretations of a dataset to be doing the same work several times over. However, I challenge this view since considering multiple interpretations is also a structured way of exploring the dataset. For example, at the start of the interpretation of 3D seismic data, geoscientists will often look at numerous 2D sections to understand the 3D volume and considering multiple interpretations of the 2D sections is a natural extension of this.

There was no evidence that the reference experts considered multiple interpretations of the seismic image, and Expert 3 noted explicitly (pers. comm.) that they did not consider any alternative interpretations. Hence, if only one of the experts had been recruited to interpret the Freyja seismic image then the uncertainty in the interpretation of the seismic image would be underestimated, e.g. if the seismic image was interpreted to be extensional, then the strike-slip interpretation might be missed, and vice versa. If multiple interpretations are considered systematically, the risk of choosing an inappropriate concept is lowered.

I did not capture whether the survey respondents considered multiple interpretations of the seismic image due to sampling restraints, i.e. because respondents were only given one copy of the seismic image to interpret. As above, the main focus of the survey was to identify the most effective 'type' of geoscientist and the most effective interpretational techniques rather than capturing the range of interpretations of a dataset, which has been investigated by Bond *et al.* (2007).

Past research (Bentley and Smith, 2008; Bond *et al.*, 2008; Rowbotham *et al.*, 2010; Chellingsworth *et al.*, 2011) indicates that best practice in geoscience interpretation should include multiple interpretations of a dataset. Hence, interpreters should try out different conceptual models even if they do not 'look right'; they should be creative and explore the uncertainty space. In some cases, models that initially do not 'look right' might still be valid. Considering multiple interpretations throughout the interpretation workflow yields the additional benefit that, if the geology is not as expected during a drilling operation, an alternative interpretation can quickly be re-assessed (Bond *et al.* (2008). Saving time is desired since drilling technology and/or production facilities can be expensive. Considering multiple interpretations also mitigates the effect of anchoring bias (Bentley and Smith, 2008) and confirmation bias (Chamberlin, 1890).

As evidenced above, 'considering multiple interpretations' mitigates cognitive biases and mitigates the risk of choosing an inappropriate concept, but is not always used because of the additional time taken. Therefore, a change in culture is needed to emphasise the benefits of considering multiple interpretations. For example, the interpretation of geological data should not require a consensus view to be reached; decision-makers should have access to multiple conceptual interpretations instead of a single 'best-guess' interpretation. The chosen conceptual model is fundamental to the interpretation and eventual outcome (Bond *et al.*, 2008) and hence, it is important to mitigate the risk of using an inappropriate concept. Consider the following example:

> *Suppose that there are two different interpretations that honour a given dataset, (e.g. the interpretations do not contradict the observed data). A team of geoscientists do not agree which interpretation is best and the decision-makers need to determine the most likely reservoir location. No other information is currently available. The traditional solution is for the team to compare the two interpretations and attempt to reach a consensus view. However, there is no guarantee that a consensus will be reached and be independently supported by all geoscientists (e.g. Polson and Curtis, 2010).*
>
> *Thus, a more informative solution is to present both interpretations to decision-makers and outline the uncertainties in each, allowing decision-makers to evaluate the interpretations within the wider project goals. The decision-makers then have three main options; choose the most likely reservoir location under interpretation 1, choose the most likely reservoir location under interpretation 2, or choose the reservoir location that would be most likely under both interpretations (assuming one exists), e.g. a reservoir location that exists under both interpretations. The three reservoir locations should then be risk-assessed within the context of the wider project* (Bond *et al.*, 2008)*, e.g. financial targets, predicted volume of oil in place, available expertise or the available drilling technology. The risk assessment would aim*

to identify the 'least risk' solution based on both interpretations rather than on a single interpretation, which might be invalid.

This example shows that multiple interpretations should be embraced and that a final risk assessment using all contextual information may identify the 'minimum risk' reservoir location within the project constraints, which could be based on multiple interpretations rather than a single interpretation. The decision-makers therefore, have more information and a better appreciation of the uncertainty associated with the interpretation of the dataset. Similarly, Ma (2011) stated that how much uncertainties should be mitigated depends on the needs of decision analysis and the 'cost of information', which both relate to the wider project objectives. Further research is needed to determine how the cost of information could be integrated into an uncertainty workflow.

8.1.3. Geological evolution of the interpretations

Cross-section balancing is one of the main techniques used to validate 2D structural interpretations, and restoration is a requirement of a 'valid balanced' cross-section. Restoration is the process of sequentially undeforming a cross-section, returning the beds to their pre-deformational geometries. In agreement with Bond *et al.* (2012), techniques that made geoscientists consider the geological evolution were most effective in this research. I found that the technique of 'writing about time' was most significant in terms of gaining high Max RE Scores, while Bond *et al.* (2012)'s technique of 'geological evolution' was most significant in their analysis in terms of identifying the original tectonic setting of the synthetic seismic image.

Bond *et al.* (2012)'s odds ratio was 89.8, while my odds ratio was 4.5. However, odds ratios should be compared relative to other variables from the same analysis rather than between analyses. In this case, the difference in the size of the odds ratios is due to what the odds ratios actually represent in each analysis. For example, Bond *et al.* (2012)'s odds ratio was a comparison between those respondents who used the technique of 'geological evolution' and those respondents who had used none of the defined techniques. However, the odds ratio from my analysis was simply the comparison between those respondents who used the technique of 'writing about time' and those who did not use that technique. The different samples, questionnaire questions and defined techniques would also impact the values of the computed odds ratios.

In this research, ‘writing about time’ was the most significant variable; more significant than respondents’ previous experience and the other interpretational techniques that I had defined. To test this result, and for evidence of causality, I conducted a separate workshop experiment that utilised a ‘control’ group to determine how effective geological thinking really was. Half the participants were asked to ‘interpret the whole seismic image’, while the other half were asked to ‘focus on and summarise’ the geological evolution of their interpretation. The analysis showed that those participants who were instructed to consider the geological evolution attained Max RE Scores that were 62% higher, on average, than the control group ($p=0.001$). Hence, even though the control group was more experienced, the evolution group still produced statistically better interpretations in terms of gaining higher Max RE Scores, thereby establishing a causal link between ‘considering the geological evolution’ and gaining high Max RE Scores.

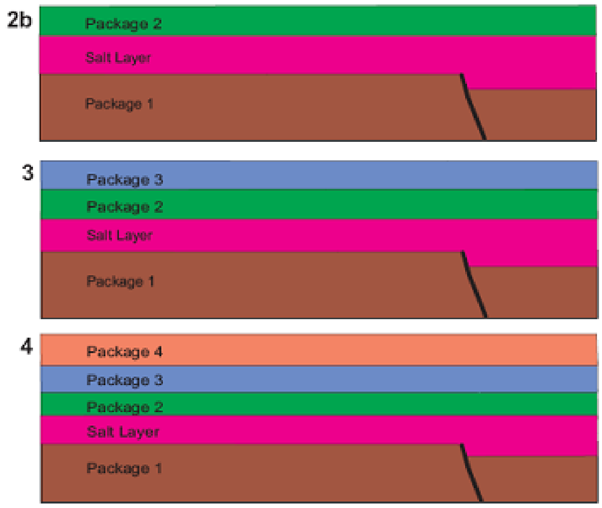
In the workshops’ group discussions, all participants agreed that ‘considering the geological evolution was beneficial to getting a valid interpretation’, but most also said that they found it ‘challenging’ to do. Thus, ‘considering the geological evolution’ might also reduce the overconfidence bias since participants found it challenging and are forced to consider their assumptions when deriving their interpretations’ evolutionary steps. Even though ‘considering the geological evolution’ was thought to be time consuming, managers still prompted their staff to use the technique.

Reasoning-based techniques (e.g. ‘writing about time’, ‘drawing a cartoon’ and ‘writing about processes’) were found to be more effective (higher odds ratios) than ‘drawing faults’ and ‘highlighting horizons’ (section 6.4.4). Hence, these results support the notion that ‘seismic interpretation’ is implicitly an investigation of the geological evolution, rather than a simple interpretation of faults and horizons. Furthermore, most of the reference experts explained their geological evolution as part of their interpretation without being asked to. Considering the evolution is something that all interpreters should strive towards and it should be explicitly included in training.

The following example shows that similar ‘looking’ interpretations can have different geological evolutions, profoundly affecting the decision-making process. The example shows that understanding the geological evolution of competing interpretations can help determine whether a reservoir could be charged with hydrocarbons, even when the interpretations look similar in terms of the interpreted tectonic style, horizons and faults. A figurative oil charge window has thus been chosen that highlights the impact of the different timings of faults on the petroleum system:

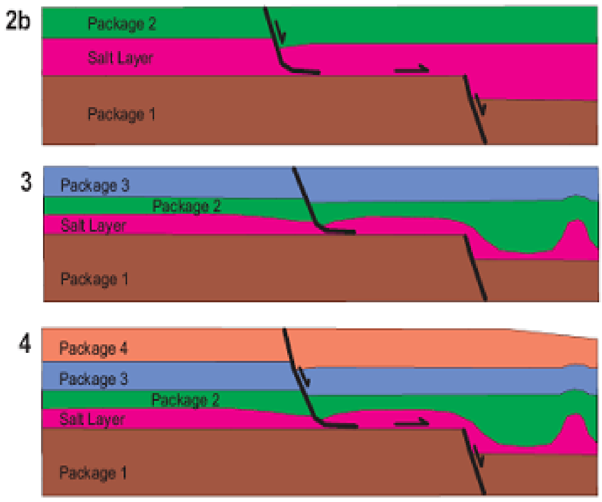
> *Experts 1 and 4 both interpreted listric faults detaching on salt with downdip compression. However, the associated geological evolutions that I derived were different. Expert 1 interpreted fault activity and folding at the time of package 5, while Expert 4 interpreted much earlier fault activity and folding, at the time of package 3. In this example, the oil charge window is assumed to be between steps 2b and 4.*
>
> *Figure 8.1 shows steps 2b to 4 of the evolution of Expert 1’s interpretation. The fold on the right, which is a trapping configuration, would not have existed before the oil charge window (steps 2b to 4). Therefore, the reservoir is unlikely to be charged with hydrocarbons since the hydrocarbons could have migrated out of the area in the absence of a trap.*

Figure 8.1 – Steps 2b to 4 of the geological evolution of Expert 1's interpretation. The middle fault was not active during this time. Hence the fold on the right, which acts as a trapping configuration, would not have formed before the oil charge window. Therefore, it is unlikely that the reservoir would be charged with hydrocarbons.



However, in the geological evolution of Expert 4's interpretation (Figure 8.2), the fold on the right formed much earlier, during the oil charge window. Under this geological evolution, it is possible that the reservoir could be charged with hydrocarbons since the trapping configuration was present.

Figure 8.2 – Steps 2b to 4 of the geological evolution of Expert 4's interpretation. The middle fault was active during this time. Hence the fold on the right, which acts as a trapping configuration, would have formed during the oil charge window. Therefore, it is possible that the reservoir would be charged with hydrocarbons.



This theoretical example illustrates that considering the geological evolution helps to determine whether a reservoir could be charged with hydrocarbons. If the geological evolution of the interpretations were not considered then the differences between the interpretations might be missed. In this case, a simple correlation of the horizons and faults was not sufficient since the reference experts' interpretations looked similar but had different geological evolutions.

Considering the geological evolution of an interpretation is therefore vital, in some cases, to understand a dataset and to produce a valid interpretation of it. Techniques based on geological thinking were found to be the most significant variables in Bond *et al.* (2012)'s research and in this research. Furthermore, this research established a causal link between 'considering the geological evolution' and gaining high Max RE Scores. One additional benefit of the technique is that the overconfidence bias is mitigated.

8.1.4. How best to define 'multiple interpretations' and 'considering the geological evolution'?

There was confusion in how 'multiple interpretations' and 'considering the geological evolution' were defined, e.g. some respondents said that they had used the techniques but left no evidence of doing so in their interpretation, according to my definitions. It is therefore likely that some respondents used different definitions than I did (perhaps less restrictive). For example, in the workshop's group discussions, some participants implied that they considered different geometries for the faults to be equivalent to 'considering multiple interpretations', whereas my definition was based on considering different geological concepts. Trying out different geometries for a fault does not necessarily change the geological concept used, e.g. steeply and shallowly dipping normal faults are both extensional.

The techniques of 'considering multiple interpretations' and 'considering the geological evolution' are also, to an extent, related to each other. For example, to consider multiple interpretations of a dataset, the interpreter implicitly needs to consider the geological evolution of each alternative to determine which are possible. Hence, the usage of either technique ('considering multiple interpretations' and 'considering the geological evolution') is expected to assist the application of the other technique since the interpreter will have a better understanding of the data. I define the two techniques in Table 8.1.

Table 8.1 – The proposed definitions for the two interpretational techniques of 'considering multiple interpretations' and 'considering the geological evolution'.

Technique	Proposed definition
Considering multiple interpretations	Multiple conceptual models of a dataset, as in Bond *et al.* (2008)
Considering the geological evolution	Drawing of a sequence of evolutionary diagrams (cartoons) that illustrate the main evolutionary stages of the geology, for each interpretation.

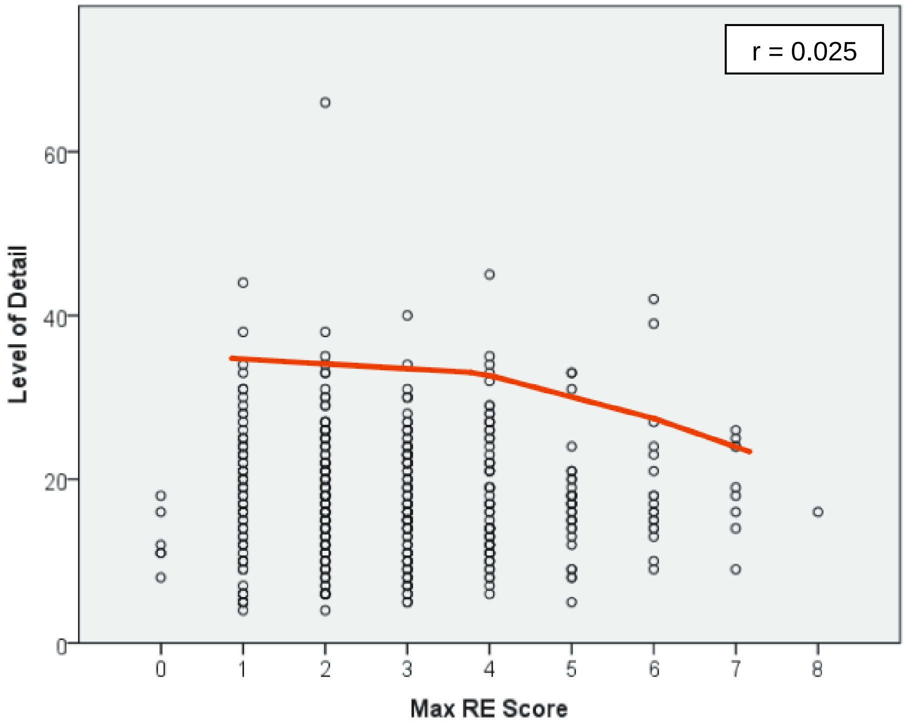
8.1.5. Is the additional time needed to apply the techniques worth it?

The qualitative data collected in the workshop experiment suggests that the main barrier to using the techniques of 'considering multiple interpretations' and 'considering the geological evolution' was the additional time that it takes. Thus, there is a balance between the effectiveness of a technique and the time (or cost) taken to apply it. For example, 'considering the geological evolution' will only benefit the interpretation if the interpreter has time to apply it. The workshops showed that it was possible to apply the technique within 35 minutes, which was the allotted time. In fact, the median reported time for the ‘evolution’ participants was 22.5 minutes, and using this time, they attained significantly higher Max RE Scores, on average, than the control group, whose median time was 30 minutes.

8.1.6. How much detail is enough?

This research has shown that more detailed interpretations did not necessarily gain higher Max RE Scores. Figure 8.3 shows a scatterplot of Max RE Scores against the ‘level of detail’ variable. Recall that the ‘level of detail’ variable is the summation of the features (main faults, smaller faults and the interpreted horizons) in respondents’ interpretations. It is clear from Figure 8.3 that there was no linear relationship between the variables. The Pearson correlation coefficient was 0.025. The ‘level of detail’ variable was also rejected from the statistical model at the screening phase, signifying that it was not associated to the Max RE Score. However, there does appear to be a decreasing upper bound on the ‘level of detail’ variable. The red line on Figure 8.3 is an approximate 90th percentile bound that was added by eye.

Figure 8.3 – A scatter plot of Max RE Scores against the 'level of detail' scores. The variables were not linearly correlated (Pearson correlation = 0.025). However, the 90th percentile upper bound (red line, added by eye) shows that high scoring respondents generally used less detail.



High scoring respondents (Max RE Scores of 5 to 8 key features) appear to have been less detailed in their interpretations. Hence, there seems to be a limit to how much detail is beneficial in an interpretation, e.g. will interpreting small faults really benefit the overall interpretation? This depends on the application of the interpretation. For example, in the oil and gas industry, small faults are likely to be studied in more detail during the appraisal and production phases, rather than in the exploration phase since they may prevent local fluid flow but have little impact on basin dynamics.

Discussing the use of modelling software in the interpretation process, Chellingsworth *et al.* (2011) recommended avoiding unnecessary complexity since it can cause overconfidence and become an anchoring point:

> *"There is a tendency to push the data beyond their limits such that the link to the hard input data becomes tenuous or is lost altogether. Our over-*

confidence in complex software solutions can drive us to ignore simpler concepts that can describe the sub-surface equally well".

It is reasonable to suggest that interpreting many small features in the seismic image was adding unnecessary complexity that could encourage geoscientists to be overconfident. This hypothesis was confirmed. A positive association was found between the interpreters' level of confidence and their 'level of detail' score, the association was found to be significant in a univariate analysis. Furthermore, 'very confident' respondents had a median 'level of detail' score of 21, while 'totally unsure' participants had a median 'level of detail' score of 15.5. Hence, these data support Chellingsworth *et al.* (2011)'s assertion that unnecessary complexity can induce overconfidence.

Perhaps the geoscientists who attained high 'level of detail' scores assumed that they were expected to interpret as much of the seismic image as possible, given that the exercise's instruction was *"please interpret the whole seismic image"*. Since respondents only had a single seismic image, the exercise was exploration based. Some respondents did not appear to realise that they were not expected to interpret many small features, but perhaps did so because they were used to producing detailed interpretations. In industry, the 'level of detail' required in an interpretation is limited by the available time. In this research, if the reference experts had chosen small features to be their key features then more detailed interpretations would have gained higher Max RE Scores.

8.1.7. Knowledge management during interpretations

Geological data are collected and analysed at great financial expense. Therefore, preserving details of the interpretations is essential. The practice of recording information that may be valuable is known as 'knowledge management'. Knowledge management is important where the reasoning for decisions might be needed in the future. Examples include: mature oil fields, sites where wastes are geologically disposed, or when the original interpreter has moved to a different company. Consider the help that additional information would be in the following hypothetical situation:

An oil field now suffers from low production rates. The original interpretation of the data is needed to assist the implementation of enhanced oil recovery *to boost production. However, no interpretation notes exist and the original interpreter has left the company.*

In this example, recorded interpretation notes would be useful in understanding the rationale behind the original interpretation and could potentially improve production rates since the decision-makers would be better informed.

During the scoring of respondents' interpretations, I found that written notes made them clearer. There are various ways to communicate interpretations including annotations, descriptions or writing about the geological evolution. Table 8.2 shows the percentages of respondents who used each of the techniques that were related to written communication. Fifty-seven point two percent of respondents used no writing techniques (i.e. relying on faults and horizons alone to explain their interpretation) and only 1.1% of respondents 'considered the geological evolution', which was the most informative technique since the evolutionary steps were written down or drawn.

The percentage of respondents using each technique was inversely related to the 'communicative power' of the technique (right-hand column), which I define to be a measure of how easily an interpretation would be understood without the original interpreter. For example, using the technique of 'descriptions or explanations' would explain the interpretation better than using 'annotations or labelling' (i.e. 'annotations or labelling' has less communicative power than 'descriptions or explanations'). The arrow in the right-hand column illustrates the increasing communicative power of each proceeding technique. The least desirable situation is when no writing techniques have been used, and the most desirable situation is when the geological evolution had been explained.

Table 8.2 – The percentages of respondents who used none of the writing techniques and each of the four written techniques. The arrow in the right-hand column shows the increasing ‘communicative power’ of each proceeding technique. The communicative power is inversely related to the percentage of respondents who used each technique.

Name of technique	Percentage of respondents who used the technique (number of respondents)	Communicative power
No writing techniques	57.2 (254)	
Annotations or labelling	40.5 (180)	
Descriptions or explanations	25.7 (114)	
Writing about time	22.7 (101)	
Considering the geological evolution	1.1 (5)	

It is not clear why so many respondents (57.2%) used no writing techniques. Possibly, some did not think they needed to, were not sure what to write or perhaps do not normally describe their interpretations and did not consider doing so.

Automatic prompts could be used in interpretation workflows. For example, interpreters could be prompted to consider the geological evolution of their interpretation and be required to record their reasoning, which would be added to a database that could be accessed by future interpreters or colleagues in distant locations. Interpreters should aim to communicate their work as clearly as possible to prevent errors in communication. Hence, written descriptions (or drawings) are an integral part of the communication of an interpretation.

8.2. What makes a good interpreter?

In this section, I discuss what makes a ‘good’ interpreter using the results and the literature review. I show that a good interpreter requires more than just experience.

8.2.1. Interpreters’ backgrounds

Certain aspects of respondents’ background were significantly associated to producing a valid interpretation. Geoscientists who did best were generally ‘specialists’ in structural geology, individuals who ‘interpreted or used seismic images’ on a regular basis, had their main work environment in a ‘super-major or major oil company’, and had a wide range of geographical experience.

Respondents' education was less important than their experience and work environment. Structural geology experience was also shown to be more influential than seismic interpretation experience. Although other interpretational approaches were used by some respondents, the most effective approach for this exercise was a structural interpretation since most of the reference experts chose some structural features as their key features.

Working in a 'super-major or major' oil company might have been an advantage, compared to experience in 'academia', 'a consultancy', 'a service company', 'a national oil company' or 'a medium-small independent oil company', because respondents may have interpreted seismic data more often or had access to a wider range of seismic data. Other reasons might include more training or a different experience demographic; it is not clear based on these data why geoscientists from 'super-major or major' oil companies, on average, gained higher Max RE Scores.

Rankey and Mitchell (2003) concluded that:

> *"Seismic interpretations likely are based on previous experiences, preconceived notions, types of data available, data quality, and geologic understanding".*

However, in conflict with Rankey and Mitchell (2003), and in agreement with Bond *et al.* (2007), my results show that certain interpretational techniques might play a larger role than prior experience in terms of producing a valid interpretation, i.e. higher odds ratios in section 6.4.5.

8.2.2. Choice of techniques

Technique variables were significant in addition to experience variables, showing that using certain techniques positively impacts the interpretation of 2D seismic data, regardless of geoscientists' education, work environment and professional experience. Therefore, inexperienced geoscientists can become more effective if they apply the effective techniques. However, future work is needed to establish whether an 'experience threshold' for the usage of these techniques exists. Based on these data, the most effective techniques were found to be: 'writing about time', 'drawing a cartoon', 'writing about processes', 'explicitly stating the concept' and 'drawing arrows on faults'. However, the restoration technique 'writing about time' was by far the most effective (highest odds ratio).

It is not clear why the 'drawing arrows on faults' technique was significant in the final analysis or how beneficial it is during an interpretation. The other significant techniques ease communication with decision-makers and stakeholders since their usage ensures more information on the interpretation. One workshop participant described the process of adding arrows to faults as being an *"artistic flourish"* and did not believe it would necessarily improve interpretations as fault movements can also be communicated through coloured horizons. However, these data disagree as the technique was significant. One possible benefit of the technique might be that it emphasises which side of a fault was downthrown and verifies that the interpreter understood how the discordance formed rather than just identifying it.

The Ph.D. student helper in workshop 4 noted that the participants in the 'evolution' group looked *"puzzled"* about their interpretations. Also, 65% of the workshop participants (28 out of 43) said that they found it 'challenging' to consider the geological evolution and 18.6% (8 out of 43) found it 'moderately challenging'. In the group discussion, one participant said that:

> *"[Considering the geological evolution] made me realise that it was not as simple as I thought it was"*.

The survey results and the workshop experiment established a strong causal link between 'considering the geological evolution' and gaining a high Max RE Score. Hence, although the technique takes additional time to apply, and can be challenging to use, these data contend that it is worth investing in. As noted by a workshop participant, it is not useful to consider the geological evolution of every interpretation, e.g. if an area has been already been explored and the regional structure is known. However, for my interpretation exercise, no regional context was provided and hence, the technique would have been useful.

8.3. How should geoscientists be trained?

The 'best' background profiles for geoscientists and the most effective interpretational techniques were identified. I discussed what makes a good interpretation and also what makes a good interpreter. In this section, I will discuss how geoscientists should be trained in order to maximise the likelihood that they will produce valid interpretations of geological data.

The most sustainable approach to training geoscientists is likely to employ a combination of specific training courses as well as prompts by managers and software. Inexperienced geoscientists should be trained from the start of their career to use the techniques that were identified to be effective, while experienced geoscientists should be prompted to use them. For example, Bond *et al.* (2011) stated that *"geological reasoning skills"* were important for novice geoscientists to develop. The aim is not to make interpretation process more complicated by adding additional steps, but to create geoscientists who use the effective techniques implicitly within their interpretation.

8.3.1. Training and experience

This research indicates that being a 'specialist' in structural geology increases the likelihood of producing a valid interpretation. Hence, to undertake a similar type of interpretation as in this exercise, geoscientists should be trained in structural geology. Additionally, training in structural geology is likely to allow geoscientists to apply the technique of 'considering the geological evolution' easier.

The results showed that interpreting seismic images often and having a wide geographical experience was beneficial to producing a valid interpretation. Thus, to extend geoscientists' experience, training should use a wide range of analogues and fieldtrips to maximise exposure to a wider range of geology. For example, geoscientists could access a range of seismic images via the 'Virtual Seismic Atlas'[14] project, and view multiple interpretations of them. Technology could be used to supplement traditional fieldtrips and reduce the need for travel. However, it is generally easier to train geoscientists to use effective techniques than it is for the geoscientists to gain experience.

In agreement with Chellingsworth *et al.* (2011), geoscientists should avoid using unnecessary complexity in their interpretations since it can lead to overconfidence. Other possible solutions might also be missed since the complexity is likely to become an anchoring point. Training should emphasise that simpler interpretations

[14] http://www.seismicatlas.org/ (last accessed: 8th December 2012).

may be able to test the underlying assumptions of interpretations better than complicated interpretations.

8.3.2. Communication and collaboration

In addition to gaining technical skills and a wide range of experience, effective communication and collaboration should be stressed during training. This research found that 57.2% of the 444 respondents did not annotate or explain their interpretation in any way at all, and only 1.1% explained their interpretation by ‘considering the geological evolution’, which had the greatest ‘communicative power’ out of the techniques that I defined. It is likely that the five respondents used the technique of ‘considering the geological evolution’ to validate their interpretation, but in doing so also communicated their interpretation effectively.

Although the exercise was an individual exercise, the fact that so few respondents explained their interpretation implies that its communication was not needed or deemed to be important. One method to systematically improve the communication of interpretations would be to require geoscientists to describe their interpretation in a knowledge management system, which could then be used by future interpreters to understand the rationale for decisions. There would then be no ambiguity in ‘knowing what the interpreter meant’. The knowledge management system does not have to be time consuming, but would become an aid during the interpretation process since geoscientists would be required to note down their assumptions and can later check them. Geoscientists should be trained to communicate their interpretations better, e.g. adding annotations and drawing a series of cartoons to illustrate the geological evolution, and using a knowledge management system is one way to achieve this. Further research would be required to determine how a knowledge management system might work.

The workshop participants rarely printed out seismic images even though they recognised the benefits of doing so (e.g. ‘more can be seen in one view’) and found working on a computer to be restrictive to collaboration. The following example highlights why effective collaboration is important in the oil and gas industry:

> *To estimate the STOIIP (stock tank oil initially in place) volume for a reservoir, experts in geology, geophysics, petrophysics and reservoir engineering are needed. The experts need to individually estimate one or*

more reservoir attributes that feed into the calculation. Attributes include the gross rock volume, the proportion of rock that could be reservoir quality, the average porosity, the average water saturation and the expansion factor for petroleum that is to be brought to the surface. However, the estimate of each reservoir attribute can affect the values of other attributes, e.g. if the geophysicist changes their interpreted horizons, which define the gross rock volume, then the geologist will have to re-evaluate their estimate of the proportion of rock that could be reservoir quality. Hence, effective collaboration between different disciplines is important.

From informal discussions with industry geoscientists, I found that the interpretation process within some companies is linear, e.g. one expert passes their interpretation on to the next expert. The reason for this linear interpretation process is to minimise cost. However, I contend that allowing all types of geoscientists to work together at the start of a project is worthwhile. Although, working in groups can result in herding bias, geoscientists with different types of experience can also complement each other by allowing different perspectives to be considered, emphasising a multi-disciplinary approach. If sufficient time is available beforehand, an expert elicitation process could be used to mitigate herding bias, assuming that the individuals are working on the same part of the project. A traditional expert elicitation process would not be applicable if individuals were working on different parts of a project.

8.4. Conclusions

The interpretation of geological data should require that multiple interpretations are actively pursued and then geologically validated. The techniques of ‘considering multiple interpretations’ and ‘considering the geological evolution’ take time to apply, but are an investment. The uncertainty in understanding the dataset will be reduced, potentially yielding economical returns since the resultant interpretations are more likely to be valid. The risk of drilling a dry well or missing an opportunity will be decreased. The additional benefit of using the two techniques is that alternative interpretations can be employed quickly if the geology found while drilling is unexpected, and the interpretations will be communicated effectively. ‘Considering multiple interpretations’ mitigates anchoring and confirmation biases, ‘while considering the geological evolution’ mitigates the overconfidence bias.

In the first instance, the interpretation should be completed with no regional context to stimulate creativity. A knowledge management system should be used to capture the rationale for decisions throughout the interpretation process. The resulting knowledge management database will be useful in the future or to colleagues in distant locations. The 'level of detail' needed in an interpretation depends on the situation, e.g. the interpretation of production data are likely to be on a smaller scale than the data interpreted for exploration. Geoscientists should avoid unnecessary complexity and instead use the time to describe their interpretations to ease their communication. A well-communicated interpretation can be more important than a detailed interpretation. Explaining the interpretation is likely to test the underlying assumptions, thereby mitigating associated risks, while adding detail (i.e. complexity) may anchor the interpretation.

A good interpreter is not necessarily experienced, but uses effective techniques. However, the use of effective techniques will become easier with experience. The priority should therefore be to ensure that experienced geoscientists are prompted to apply the effective techniques, while inexperienced geoscientists are trained to use the techniques. Training should focus on developing geological reasoning skills by interpreting seismic images often and providing a wide range of geology. Technology could be used to supplement traditional fieldtrips. A good interpreter should also be experienced in structural geology and have a wide range of geographical experience.

The next chapter presents an interpretation workflow that accounts for, and mitigates, the risk arising from uncertainty in geoscience interpretation.

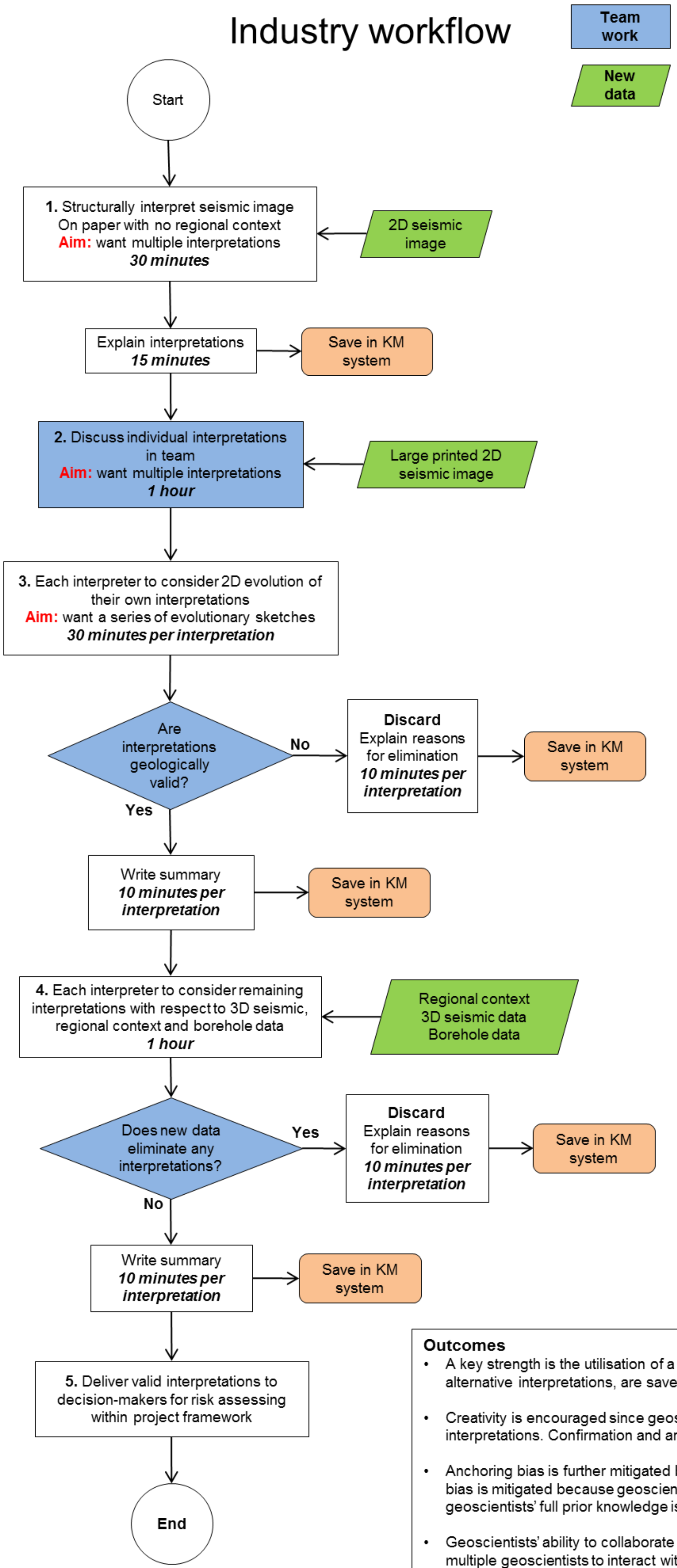
Chapter 9 – Industry Workflow

An industry workflow is now presented that mitigates the risk arising from the uncertainty in geoscience interpretation. The workflow is based upon the key results from this thesis including the review of relevant literature from geoscience and psychology. The workflow includes individual and team stages, increases geoscientists' ability to collaborate in a team, and incrementally brings more data into the workflow. Intermediate interpretations and geoscientists' evolving assumptions and thoughts are systematically captured in a knowledge management system. The workflow mitigates anchoring, overconfidence, confirmation and herding biases, while allowing multiple valid interpretations to be passed to decision-makers for risk-assessing within the project context.

As noted in the qualitative data from the workshop (Chapter 7), the main objection to 'considering multiple interpretations' is the additional time that it takes. However, this workflow only takes about four hours per geoscientist. The workflow is particularly suited to oil and gas exploration, where there is often the most uncertainty; e.g. appraisal, development and production phases in the oil and gas industry can only take place when a reservoir has been discovered. The workflow is thus, ideal for 'project framing' sessions to encourage creativity before the more specific work begins.

Each stage in the following workflow is explained and the evidence for it is noted. For convenience, the workflow is also available as an A3 handout at the back of the thesis.

Industry workflow



Instructions
The workflow needs to be completed by 3+ geoscientists simultaneously, e.g. multiple interpreters with a range of prior knowledge (Bond et al., 2008).

A moderator should lead the workshop and enforce time limits. The workflow should be explained to geoscientists prior to starting.

Time limits are short to minimise cost. Each geoscientist spends about 4 hours completing the workflow.

Geoscientists work individually, apart from at the blue boxes to mitigate herding bias (e.g. Polson and Curtis, 2010).

Green boxes indicate new data being brought into the workflow.

Should use the definitions of 'multiple interpretations' and 'considering the geological evolution' as defined in the discussion chapter.

Geoscientists are encouraged to record their thoughts in the knowledge management system at any time during the workflow.

Box 1
The interpretation is completed on paper to avoid the *"use of unnecessary complexity"* (Chellingsworth et al., 2011).

Each geoscientist should approach the exercise with an 'open mind' and consider their assumptions.

Geoscientists should avoid 'anchoring and adjusting' (Tversky and Kahneman, 1974). No regional context is provided to mitigate anchoring on this information (Bond et al., 2008). Acetates can be used to avoid marking the seismic image, to mitigate anchoring on a specific interpretation.

The interpretation should 'honour the data', i.e. not contradict it.

The aim is to devise multiple interpretations of the data (Bentley and Smith, 2008; Bond et al., 2008; Rowbotham et al., 2010; Chellingsworth et al., 2011) and encourage creativity. Utilising multiple geoscientists with a range of prior experience may encourage a greater range of interpretations.

Considering multiple interpretations mitigates confirmation bias (Chamberlin, 1890) and anchoring bias (Bentley and Smith, 2008).

Box 2
Large printed seismic images should be used to improve team collaboration (this research).

Box 3
'Considering the geological evolution' was the most effective interpretational technique (Bond et al., 2012; this research); it is a restoration technique that can be used to test assumptions and identify inconsistencies.

Geoscientists should attempt to draw a series of evolutionary sketches for each interpretation, summarising the geological evolution.

'Considering the geological evolution' also mitigates the overconfidence bias as it is 'challenging' to apply and the assumptions are tested.

Box 4
The regional knowledge should be added in at this stage to determine whether any interpretations can be eliminated.

Box 5
A single 'best' model is not chosen. Multiple interpretations are passed to the decision-makers for risk assessing within the project framework (Bond et al., 2008).

Outcomes
- A key strength is the utilisation of a knowledge management system. The rationale for decisions, as well as the alternative interpretations, are saved. It is quick to re-assess the alternative interpretations if needed.
- Creativity is encouraged since geoscientists are required to complete the exercise on paper and attempt multiple interpretations. Confirmation and anchoring biases are mitigated.
- Anchoring bias is further mitigated because the regional context is withheld at the start of the workflow. Herding bias is mitigated because geoscientists are required to work individually during parts of the workflow. The geoscientists' full prior knowledge is still accessed during the teamwork stages.
- Geoscientists' ability to collaborate in teams is likely to be increased since large printed seismic images allow multiple geoscientists to interact with the data. Avoids having only one geoscientist controlling a computer.
- Interpretations are validated via evolutionary sketches, which force geoscientists to test their assumptions. The technique also mitigates the overconfidence bias.

Chapter 10 – Conclusions

Accurately modelling the subsurface is extremely important to industry and thus, understanding how to make geoscientists more effective should be a priority. Furthermore, in the oil and gas industry, passing on technical skills to the next generation of geoscientists is key because of the so-called 'big crew change', where many experienced technical professionals are due to retire over the next decade, with fewer professionals able to replace them. Exploration is also becoming more complex because more challenging reservoirs are being targeted (e.g. high pressure, high temperature) and because geological datasets continue to increase in size and detail. Training and knowledge transfer in geoscience is essential to meeting the challenges of the twenty-first century (e.g. the extraction of resources and the geological storage of wastes).

Parametric uncertainty is often analysed more rigorously than conceptual uncertainty in the oil and gas industry. However, conceptual uncertainty can be more influential since the underlying assumptions directly affect the resulting geological model. In this research, analysis of the reference experts' interpretations showed that two different tectonic concepts fitted the seismic image. If multiple interpretations had not been encouraged (e.g. by recruiting multiple reference experts) then one of the valid interpretations might have been missed. All future work could then be based on an incorrect assumption; in this case, the chosen tectonic concept. The associated risk in the oil and gas exploration industry might be drilling a dry well or missing a hydrocarbon accumulation.

The analysis of the survey data identified that respondents' experience was more important than their education and work environment. Experience in terms of structural geology ability, how often seismic images were interpreted and the range of geographical experience were all significantly associated to producing a valid interpretation of the seismic image. In each case, the association was positive: more experience was associated to 'more similar' interpretations. Respondents whose main work environment was a 'super-major or major oil company' also tended to produce better interpretations. Certain interpretational techniques were significant even when taking into account respondents' differing backgrounds. Hence, the training of geoscientists matters, regardless of their experience.

The techniques that respondents used most in the interpretation exercise were based on identifying individual features (e.g. faults and horizons), while the more 'advanced' techniques, which were based on geological reasoning, were used least. Managers should therefore ensure that experienced geoscientists are prompted to use the effective techniques, while inexperienced geoscientists are taught to use them, so that the approaches become implicit within the interpretation process. While experience cannot be fast-tracked, the most effective approaches can be explicitly requested to aid geoscientists' development.

The analysis of the survey and workshop data showed that geological reasoning (in the form of 'considering the geological evolution', a restoration technique) was the most powerful approach available to geoscientists. For example, from the survey data, the technique of 'writing about time' was the most influential technique, while the workshop participants who were instructed to 'consider and state the geological evolution' of their interpretation, on average, gained Max RE Scores that were 62% higher than the 'control' participants. Both of these results were statistically significant. Furthermore, via the workshop experiment, a causal link between 'considering the geological evolution' and 'producing a similar interpretation' to the reference experts' interpretations was established, strongly suggesting that evolutionary thinking *causes* geoscientists to produce better interpretations, instead of just being part of a good interpretation.

The process of explaining the geological evolution also identifies interpretations that are invalid, increases the understanding of alternative interpretations and reduces the overconfidence bias. Therefore, training should focus on developing geological reasoning skills by exposing geoscientists to the widest possible range of geology, while discouraging 'added complexity' in the interpretation where possible. Digital resources could be used to supplement traditional field courses and would reduce the need to travel. Training should also make geoscientists aware of cognitive biases and the associated mitigation strategies.

Written descriptions (and/or drawings) are an integral part of the communication of an interpretation. However, few survey respondents explained their interpretation, perhaps implying that they did not deem its communication to be important. One method to systematically improve the communication of interpretations is to require geoscientists to describe them in a knowledge management system. The knowledge management system could also be used to capture geoscientists' evolving assumptions through an interpretation, and would become a valuable resource for future interpreters. In addition, the terminology used in a project should be documented and consistent to avoid confusion.

Hence, by analysing the relevant literature, survey data, reference experts' interpretations and the workshop data, I derived an interpretation workflow to make geoscientists more effective. The effects of cognitive biases are mitigated, geoscientists' ability to collaborate in teams is increased, assumptions are systematically preserved, multiple interpretations are encouraged and their geological evolutions are validated. The resulting multiple interpretations should then be delivered to the decision-makers for risk assessing within the context of the project. The interpretation workflow is most applicable to 'project framing sessions', before the detailed work begins, but might also be applicable to other types of geoscience interpretation with minor changes.

While another large survey investigating the uncertainty in geoscience interpretation has been completed (Bond *et al.*, 2007), this is the first time that so many factors have been statistically quantified, and a workflow has been derived that mitigates cognitive biases, improves team work, validates multiple interpretations and captures interpreters' evolving assumptions. Hence, at a time when geoscience interpretation is becoming more complex, and ever more important, this thesis documents how geoscientists and management can become more effective when working with uncertain data. This thesis is therefore a step forward in understanding how geoscience practice and training can be improved and how the risk arising from uncertainty in geoscience interpretation can be mitigated.

Chapter 11 – Future Work

In this chapter, future work is suggested. The future work relates to the teaching of geoscience and to the application of the interpretation workflow in the oil and gas industry.

- *Would the additional time taken to apply the interpretation workflow be economically viable?*

  As in Chapter 9, the workflow might be most appropriate for 'project framing sessions' for oil and gas exploration. However, this would need to be tested. The time taken to run the workflow could be recorded for multiple framing sessions and then compared against the perceived difference in the final interpretations and understanding of the dataset.

- *How could a knowledge management system be used in the interpretation workflow?*

  It would be challenging to ensure that all geoscientists used the knowledge management system consistently. It is not yet clear how easily multiple interpretations and explanations of the geological evolution could be recorded and then understood by different geoscientists at a later date. The metadata would also have to be quality controlled since it would inform future decisions.

- *How could the 'cost of information' be accounted for in the interpretation workflow?*

  The interpretation workflow assumes that the different types of data have already been collected. However, it might be possible to generalise the workflow to include the cost of collecting new data. Further research would be needed to determine how the 'cost of information' could be accounted for in the workflow.

- *How does the usage of techniques change with experience?*

  To investigate the evolving use of techniques through university and professional development, the techniques used by different experience cohorts could be analysed. Due to the large sample size in this research, experience categories could be defined by 5-year intervals while maintaining reasonable numbers of geoscientists in each category. This would enable the detection of an ‘experience threshold’ where geoscientists become able to apply the effective techniques.

References

Agresti, A., 1996, An introduction to categorical data analysis, New York ; Chichester : Wiley, Wiley series in probability and statistics. Applied probability and statistics, 290 p.:

Akaike, H., 1974, A new look at the statistical model identification: Automatic Control, IEEE Transactions on, v. 19, no. 6, p. 716-723.

Archie, G. E., 1950, Introduction to petrophysics of reservoir rocks: AAPG Bulletin, v. 34, no. 5, p. 943-961.

Arnell, N. W., Tompkins, E. L., and Adger, W. N., 2005, Eliciting information from experts on the likelihood of rapid climate change: Risk Analysis, v. 25, no. 6, p. 1419-1431.

Aspinall, W., 2010, A route to more tractable expert advice: Nature, v. 463, no. 7279, p. 294-295.

Bacon, F., 1620, Novum Organum (Book 1, Aphorism 46).

Bacon, M., Simm, R., and Redshaw, T., 2003, 3-D seismic interpretation, Cambridge, Cambridge University Press, 225 p.:

Baddeley, M. C., Curtis, A., and Wood, R. A., 2004, An introduction to prior information derived from probabilistic judgements: Elicitation of knowledge, cognitive bias and herding, *in* Curtis, A., and Wood, R., eds., Geological Prior Information: Informing Science and Engineering, Volume 239.

Bally, A. W., Gordy, P. L., and Stewart, G. A., 1966, Structure, seismic data and orogenic evolution of Southern Canadian Rocky Mountains: Bulletin of Canadian Petroleum Geology, v. 14, no. 3, p. 337–381.

Bárdossy, G., and Fodor, J., 2001, Traditional and new ways to handle uncertainty in geology: Natural Resources Research, v. 10, no. 3, p. 179-187.

Bentley, M., and Smith, S., 2008, Scenario-based reservoir modelling: the need for more determinism and less anchoring, *in* Robinson, A., Griffiths, P., Price, S., Hegre, J., and Muggeridge, A., eds., The Future of Geological Modelling in Hydrocarbon Development, Volume 309, Geological Society, London, Special Publications, p. 145-159.

Bertram, G., Emery, D., and Myers, K., 1996, Sequence stratigraphy, Wiley Online Library.

Bond, C. E., Gibbs, A. D., Shipton, Z. K., and Jones, S., 2007, What do you think this is? “Conceptual uncertainty” in geoscience interpretation: GSA Today, v. 17, no. 10, p. 4.

Bond, C. E., Lunn, R. J., Shipton, Z. K., and Lunn, A. D., 2012, What makes an expert effective at interpreting seismic images?: Geology, v. 40, no. 1, p. 75-78.

Bond, C. E., Philo, C., and Shipton, Z. K., 2011, When there isn't a right answer: Interpretation and reasoning, key skills for twenty-first century geoscience: International Journal of Science Education, v. 33, no. 5, p. 629-652.

Bond, C. E., Shipton, Z. K., Gibbs, A. D., and Jones, S., 2008, Structural models: optimizing risk analysis by understanding conceptual uncertainty: First Break, v. 26, p. 65-71.

Box, G. E. P., and Draper, N. R., 1987, Empirical Model-Building and Response Surfaces, Wiley.

Bradburn, N. M., Wansink, B., and Sudman, S., 2004, Asking questions : the definitive guide to questionnaire design- for market research, political polls, and social and health questionnaires, San Francisco, Calif. : Jossey-Bass ; Chichester : John Wiley.

Capen, E. C., 1976, The Difficulty of Assessing Uncertainty: Journal of Petroleum Technology, v. 28, no. 8, p. 843-850.

Cartwright, J., and Huuse, M., 2005, 3D seismic technology: the geological 'Hubble': Basin Research, v. 17, no. 1, p. 1-20.

Chadwick, P. K., 1975, A psychological analysis of observation in geology: Nature, v. 256, no. 5518, p. 570-573.

-, 1976, Visual illusions in geology: Nature, v. 260, no. 5550, p. 397-401.

Chamberlin, R. T., 1910, The Appalachian folds of central Pennsylvania: Journal of Geology, v. 18, no. 3, p. 228-251.

Chamberlin, T. C., 1890, The method of multiple working hypotheses: Science (old series), v. 15, p. 92–96.

Chellingsworth, L., Bentley, M., Kane, P., Milne, K., and Rowbotham, P., 2011, Human limitations on hydrocarbon resource estimates – why we make mistakes in data rooms: First Break, v. 29, no. 4, p. 49-57.

Cochran, W. G., 1977, Sampling techniques, Wiley, 3rd ed.

Cooke, R. M., 1991, Experts in uncertainty : opinion and subjective probability in science, New York : Oxford University Press.

Curtis, A., 2012, The science of subjectivity: Geology, v. 40, no. 1, p. 95-96.

Curtis, A., and Wood, R., 2004, Optimal elicitation of probabilistic information from experts, *in* Curtis, A., and Wood, R., eds., Geological Prior Information: Informing Science and Engineering, Volume 239, Geological Society, London, Special Publications, p. 127-145.

Dahlstrom, C. D. A., 1969, Balanced Cross Sections: Canadian Journal of Earth Sciences, v. 6, no. 4, p. 743-757.

Davies, R. J., Cartwright, J. A., Stewart, S. A., Lappin, M., and Underhill, J. R., 2004, 3D seismic technology : application to the exploration of sedimentary basins, London : Geological Society ; Tulsa, OK : AAPG Bookstore, Memoir (Geological Society of London), v. 29, 355 p.:

Davis, J. C., 2002, Statistics and data analysis in geology, New York, N.Y. ; Chichester : John Wiley & Sons, 3rd ed, 638 p.:

Delfiner, P., 2008, Uncertainty in prospect evaluation: Lessons from the movie industry: AAPG Bulletin, v. 92, no. 10, p. 1415-1429.

Dobson, A. J., and Barnett, A. G., 2008, An Introduction to Generalized Linear Models, Taylor & Francis Group, 3rd ed, 307 p.:

Eaton, T. T., 2006, On the importance of geological heterogeneity for flow simulation: Sedimentary Geology, v. 184, no. 3–4, p. 187-201.

Egermann, P., and Lenormand, R., A new methodology to evaluate the impact of the local heterogeneity on petrophysical parameters (Kr, Pc): application on carbonate rocks, *in* Proceedings International Symposium of the Society of Core Analysts, Abu Dhabi, UAE2004.

Elliott, D., 1983, The Construction of Balanced Cross-sections: Journal of Structural Geology, v. 5, no. 2, p. 101-101.

Frodeman, R., 1995, Geological reasoning: Geology as an interpretive and historical science: Geological Society of America Bulletin, v. 107, no. 8, p. 960.

Gao, D. L., 2009, 3D seismic volume visualization and interpretation: An integrated workflow with case studies: Geophysics, v. 74, no. 1, p. 1-12.

Greenbaum, T. L., 1998, The handbook for focus group research, Thousand Oaks, California, Sage Publications, 280 p.:

Groshong, R. H., Bond, C., Gibbs, A., Ratliff, R., and Wiltschko, D. V., 2012, Preface: Structural balancing at the start of the 21st century: 100 years since Chamberlin: Journal of Structural Geology, v. 41, p. 1-5.

Hall, M., 2010, The rational geoscientist: The Leading Edge, v. 29, no. 5, p. 596-601.

Hamilton, W. D., 1971, Geometry for the selfish herd: Journal of theoretical Biology, v. 31, no. 2, p. 295-311.

Harvey, N., 1997, Confidence in judgment: Trends in cognitive sciences, v. 1, no. 2, p. 78-82.

Heckman, J. J., 1979, Sample Selection Bias as a Specification Error: Econometrica, v. 47, no. 1, p. 153-161.

Hosmer, D. W., and Lemeshow, S., 1989, Applied logistic regression, New York : Wiley.

Jain, A. K., and Gupta, S., 1987, Some Evidence on" Herding" Behavior of US Banks: Journal of Money, Credit and Banking, p. 78-89.

Jakosky, J. J., 1950, Exploration geophysics, Trija Publishing Company.

Jenkins, G. W., and Slack, J. L., 1985, Statistics and Probability, Heinemann Educational Books Ltd, 416 p.:

Judge, P. A., and Allmendinger, R. W., 2011, Assessing uncertainties in balanced cross sections: Journal of Structural Geology, v. 33, no. 4, p. 458-467.

Kahneman, D., and Tversky, A., 1973, On the psychology of prediction: Psychological Review, v. 80, no. 4, p. 237-251.

Kali, Y., and Orion, N., 1996, Spatial abilities of high school students in the perception of geologic structures: Journal of Research in Science Teaching, v. 33, no. 4, p. 369-391.

Kleinbaum, D. G., 2002, Logistic regression : a self-learning text, New York : Springer.

La Pointe, P. R., and Fox, A., 2011, Quantification of Conceptual and Parametric Uncertainties in Fractured Reservoir Models, *in* Ma, Y. Z., and La Pointe, P. R., eds., Uncertainty Analysis and Reservoir Modeling, AAPG Memoir 96, p. 57-76.

Levy, P. S., and Lemeshow, S., 1999, Sampling of populations : methods and applications, New York : Wiley, 3rd ed.

Little, R. J. A., and Rubin, D. B., 2002, Statistical Analysis with Missing Data, New York, John Wiley & Sons, Inc, 2nd ed, 408 p.:

Liu, Y., Rigsby, P. G., Sinha, R., Peterson, S., Thomas, J., and Zimmerman, G., 2011, Geologic Modeling and Uncertainty Analysis of a Gulf of Mexico Reservoir, *in* Ma, Y. Z., and La Pointe, P. R., eds., Uncertainty Analysis and Reservoir Modeling, AAPG Memoir 96, p. 77-87.

Loizou, N., 2003, A post-well analysis of recent years exploration drilling along the UK Atlantic Margin: DTI Oil and Gas Directorate (UK), Sharp IOR enewsletter, v. 3.

Ma, Y. Z., 2011, Uncertainty Analysis in Reservoir Characterization and Management: How Much Should We Know About What We Don't Know?, *in* Ma, Y. Z., and La Pointe, P. R., eds., Uncertainty Analysis and Reservoir Modeling, AAPG Memoir 96, p. 1-15.

Mann, J. C., 1993, Uncertainty in geology, Computers in Geology - 25 Years of Progress, Oxford University Press, p. 241-254.

Marshak, S. A., and Mitra, G., 1988, Basic Methods of Structural Geology: Part I, Elementary Techniques, Part Ii, Special Topics, Prentice Hall.

Matthews, J. N. S., 2006, Introduction to randomized controlled clinical trials, Boca Raton, Fla. : CRC Press.

Nagelkerke, N. J. D., 1991, A note on a general definition of the coefficient of determination: Biometrika, v. 78, p. 691-692.

Nickerson, R. S., 1998, Confirmation bias: A ubiquitous phenomenon in many guises: Review of General Psychology; Review of General Psychology, v. 2, no. 2, p. 175.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T., 2006, Uncertain judgements : eliciting experts' probabilities, London ; Hoboken, NJ : John Wiley & Sons, Statistics in practice, 321 p.:

Oppenheim, A. N., 1992, Questionnaire design, interviewing, and attitude measurement, London ; New York : Continuum.

Oskamp, S., 1965, Overconfidence in case-study judgments: Journal of consulting psychology, v. 29, no. 3, p. 261.

Payton, C. E., 1977, Seismic stratigraphy: applications to hydrocarbon exploration, American Association of Petroleum Geologists.

Polson, D., and Curtis, A., 2010, Dynamics of uncertainty in geological interpretation: Journal of the Geological Society, v. 167, no. 1, p. 5-10.

Raafat, R. M., Chater, N., and Frith, C., 2009, Herding in humans: Trends in cognitive sciences, v. 13, no. 10, p. 420-428.

Rankey, E., and Mitchell, J., 2003, That's why it's called interpretation: Impact of horizon uncertainty on seismic attribute analysis: The Leading Edge, no. 22, p. 820-828.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J., and van der Keur, P., 2006, A framework for dealing with uncertainty due to model structure error: Advances in Water Resources, v. 29, no. 11, p. 1586-1597.

Resor, P. G., and Pollard, D. D., 2012, Reverse drag revisited: Why footwall deformation may be the key to inferring listric fault geometry: Journal of Structural Geology, v. 41, p. 98-109.

Rowbotham, P., Kane, P., and Bentley, M., 2010, Bias in geophysical interpretation–the case for multiple deterministic scenarios: The Leading Edge, v. 29, no. 5, p. 590-595.

Shanteau, J., 2000, Why Do Experts Disagree?, *in* Green, B., Cressy, R., Delmar, F., Eisenberg, T., Howcroft, B., Lewis, M., Schoenmaker, D., Shanteau, J., and Vivian, R., eds., Risk behaviour and risk management in business life: Dordrecht, The Netherlands, Kluwer Academic, p. 186-196.

Slovic, P., Fischhoff, B., and Lichtenstein, S., 1982, Facts and fears: understanding perceived risk, *in* Kahneman, D., Slovic, P., and Tversky, A., eds., Judgement under uncertainty: Heuristics and bias, Cambridge University Press, p. 463-489.

Smith, A., 1759, The Theory of Moral Sentiments.

Steele, C. M., and Aronson, J., 1995, Stereotype threat and the intellectual test performance of African Americans: Journal of Personality and Social Psychology, v. 69, no. 5, p. 797-811.

Stewart, S. A., 2007, Salt tectonics in the North Sea Basin: a structural style template for seismic interpreters: Geological Society, London, Special Publications, v. 272, no. 1, p. 361-396.

-, 2011, Vertical exaggeration of reflection seismic data in geoscience publications 2006-2010: Marine and Petroleum Geology, v. 28, no. 5, p. 959-965.

Torvela, T., and Bond, C. E., 2011, Do experts use idealised structural models? Insights from a deepwater fold–thrust belt: Journal of Structural Geology, v. 33, no. 1, p. 51-58.

Tversky, A., and Kahneman, D., 1974, Judgment under uncertainty: Heuristics and biases: Science, v. 185, no. 4157, p. 1124-1131.

U.S. Energy Information Administration, 2011, Annual Energy Review 2010: U.S. Department of Energy.

Wadell, H., 1932, Sedimentation and Sedimentology: Science (New York, N.Y.), v. 75, no. 1931, p. 20.

Williams, G. D., 1993, Tectonics and seismic sequence stratigraphy: an introduction: Geological Society, London, Special Publications, v. 71, no. 1, p. 1-13.

Willis, G. B., 2005, Cognitive interviewing : a tool for improving questionnaire design, London, Sage Publications, 335 p.:

Woodward, N. B., 2012, Evaluation, analysis and prediction of geologic structures: Journal of Structural Geology, v. 41, no. 0, p. 76-85.

Ye, M., Pohlmann, K. F., and Chapman, J. B., 2008, Expert elicitation of recharge model probabilities for the Death Valley regional flow system: Journal of Hydrology, v. 354, no. 1-4, p. 102-115.

Table of Tables

Table of Figures

Appendices

Please see attached envelope at the back of the thesis for a CD-ROM containing the appendices. This section contains a summary of what is included.

Appendix 1 – Statistical glossary

Appendix 2 – Survey methodology

Ethics form for survey (University of Glasgow)

Background questionnaire

Anchoring bias document

George Bertram's comments

Sampling locations for survey

Appendix 3 – Reference experts

Simon Stewart's transcript

Photos of reference experts' interpretations

Geological evolutions for the reference experts' interpretations

Appendix 4 – Sample information

Data tables (responses to questions)

Example words (geological time and geological processes)

Data tables (usage of techniques)

Appendix 5 – Main analysis

Statistical coding and relevant output

Appendix 6 – Industry workshops

Ethics form for workshops (University of Strathclyde)

Workshop background questionnaire (same for both groups)

Interpretation exercise for control group

Interpretation exercise for evolution group

Post-interpretation questionnaire for control group

Post-interpretation questionnaire for evolution group

Responses to countable questions from the 'post-interpretation' questionnaire (frequency tables)

Reports for workshops 1 to 4 (observations and group discussions)

Appendix 7 – Miscellaneous results

- 'Seminar' respondents did not produce better interpretations than 'conference' respondents (p=0.989). See section 3.6.2.
- 'Structural geology experience' and 'seismic interpretation experience' were significantly associated for these respondents (p<0.001). See section 5.2.2.
- 'Confident' respondents did not produce better interpretations than 'doubtful' respondents (p=0.11). See section 5.2.6.
- Due to the fact that so few respondents had used the 'evolutionary thought' technique, the association [years of experience] was not statistically significant (p=0.886). See section 6.4.5.
- No association between 'years of experience' and 'cartoon drawn' (p=0.985). See section 6.4.5.
- No association between 'years of experience' and 'writing about time' (p=0.684). See section 6.4.5.
- The evolution group, though less experienced, achieved a much higher mean 'Max RE Score' than the control group; this result is highly significant (p=0.001). See section 7.4.
- Model-driven participants, on average, attained a Max RE Score of 1 key feature more than data-driven participants. This result was not statistically significant (p=0.251). See section 7.6.1.

The envelope at the back of the thesis contains:

Background questionnaire (A3 print out)

Geological evolutions for the reference experts' interpretations (A4 print outs)

Industry workflow (A3 print out)

CD-ROM containing the appendices