SPE -124703

Process for tracking the evolving perception of risk during CO2 storage projects

Debbie Polson[1], Andrew Curtis[1], Claudia Vivalda[2] and Stephanie Saunier[2]

[1]University of Edinburgh, School of GeoSciences, Edinburgh EH9 2JT and ECOSSE (Edinburgh Collaborative of Subsurface Science and Engineering)

[2] Schlumberger Ltd

This paper was prepared for presentation at the 2009 SPE Offshore Europe Oil & Gas Conference & Exhibition held in Aberdeen, UK, 8–11 September 2009.

Abstract

The paper describes a process for constructing a risk register to be used to track the evolving perception of risk during a CO_2 storage project. A project-specific risk register is developed through a structured elicitation process to determine initial perception of risk and through discussion with experts. Regular updating of the register by experts is used to track changes in the assessment of risk as the project progresses and inform decision and actions during the project with the aim of reducing risk to an acceptable level.

Introduction

The current attention to climate change has increased the pressure on all nations to reduce emissions of greenhouse gases. Carbon capture and storage (CCS) is potentially an important technology to reduce emissions of CO_2, one of the key greenhouse gases, particularly as fossil fuels are likely to continue to supply a major part of the world's energy needs for decades to come (Jepma et al, 1998; Bajura, 2001).

In CCS the CO_2 generated by large point source emitters, primarily power plants, is captured and transported to a storage site where it is injective into the Earth's subsurface and stored in the pore spaces of rock. Potential storage sites include saline reservoirs, depleted oil and gas reservoirs, and coal seams. While the technology for CCS already exists (e.g. Moritis, 2003; Torp et al, 2004) and is currently used in enhanced oil recovery, the scale of CO_2 sequestration required to impact significantly on anthropogenic emissions, dwarves existing projects (Nordbotten et al, 2005). In addition, long term storage of CO_2 is required if negative climate impacts are to be avoided.

Carbon dioxide injected into deep geological formations will initially be more buoyant than the existing formation fluids (Bachu, 2003). This will result in the migration of CO_2 upwards through the formation and offers a potential leakage pathway. In principle a sealing caprock over a storage reservoir should trap the CO_2. However over-pressurisation of the seal, fractures, faults and injection wells all offer possible sites of failure of the containment system and hence routes for CO_2 migration away from the target reservoir (Celia et al, 2002).

Leakage of CO_2 and displacement of the formation fluids may result in acidification and/or contamination of groundwater and soils, with uncontrolled leakage posing a risk to near-surface ecosystems (Celia et al, 2002; Farrar et al, 1995). There is also a significant financial risk in rolling out CCS on a major scale, with significant investment required both in improving technology and in identifying and characterising potential storage sites.

To minimise these potential negative impacts a complete risk assessment of the full capture, injection and storage process is required for all CCS projects. We present a case study developed within a project, henceforth referred to as Case Study 1, in which a complete, simulated assessment process was carried out for two theoretical storage sites for CO_2. We describe a risk assessment process designed to highlight high risk areas to the project, which feeds directly into the decision making process regarding future development and risk mitigation activities.

Risk Analysis Method

Risk refers to the probability and scale of eventualities which may negatively impact on a project. A six step process has been developed for the compilation of a project-specific risk register based on that of

Jammes et al, 2008. The register contains a list of Features, Events and Processes (FEP) which define relevant scenarios and behaviour of CO_2 in the storage system which are assessed by experts for their likelihood (probability) of impacting the project, and the severity (scale) of this impact. Most engineering aspects of the project will have well established risk assessment methods already in place and hence are not included here in a detailed way.

This process took place during the initial months of Case Study 1, and the resulting register was used to assess risk at regular intervals over the lifetime of the project, allowing the evolution of the perception of risks to be tracked. The process is summarised below:

Step 1. *List Construction*. Construct or obtain a comprehensive list of all known possible features, process and events (FEPs) that might conceivably pose a project risk. Construct or obtain preliminary likelihood and severity scales.

Step 2. *Initial Group Elicitation.* General elicitation of lower and upper bound of likelihood and severity of each FEP affecting one or more of the possible storage sites using all available experts. For each FEP also elicit the level of expertise of each participating expert.

Step 3. *Individual Consultations*. Discussion with project partners to ensure all areas of potential risk are included in risk register and define project specific criteria for likelihood and severity scales.

Step 4. *Propose Register*. Based on the results of steps 2 and 3, a risk analyst produces a proposed risk register of project-specific FEPs and proposed likelihood and severity scales. This takes the form of each of the original FEPs categorised as *to keep*, *to remove*, and *for discussion* by group.

Step 5. *Group Discussion.* Meeting of partners to discuss proposed list of FEPs to be kept and produce a finalised register.

Step 6. *Reflection and Validation*. After a pre-agreed period, all project partners to agree on a final risk register.

In step 1 a comprehensive list of all known possible risks that might affect a CO_2 storage project is compiled. For our case study this was compiled partly from the Quintessa CO_2 FEP register (Quintessa Ltd), which was extended to include additional FEPs. The likelihood and severity scales were taken from Hnottavange-Telleen et al (2008).

Using the FEP list from step 1 and the preliminary scales for likelihood (L) and severity (S), the participants were asked to assess the lower and upper bounds of the likelihood and severity for each FEP for the injection project as a whole, without distinguishing between the various possible sites that might in the end be chosen. The lower bound L x S gives the minimum perception of risk and the upper bound L x S give the maximum perception of risk.

The purpose of this initial elicitation is to identify areas perceived as definitely low risk which should not affect the project. Therefore the upper bound of risk is used to rank the FEPs as this ensures that any FEP which has the potential to be moderate or high risk is identified as such. The range (upper bound – lower bound) is used to give an indication of uncertainty in the risk. While it would be possible to also elicit a best-guess value, it is not necessary at this stage. Experts were also asked to rank their level of expertise on a scale of 1 to 5 (1=novice, 5=expert) for each FEP.

The combined likelihood and severity (L x S) matrix is shown in Table 1. FEPs which scored an L x S value of less than 5 are considered to be low risk. FEPs in the range $5 < L \times S < 10$ are moderate risk, in the range $10 < L \times S < 20$ are high risk and > 20 are very high risk. For FEPs in the high and very high risk bands, the risk is considered to be unacceptable and mitigation activities are required to reduce the risk. FEPs in the moderate banding might be tolerable but ideally actions should take place to reduce these. The lowest band of FEPs are considered to pose a minimal and acceptable risk to the project. The purpose of risk analysis is to allow decisions and actions to be taken that reduce all FEPs to moderate or low risk.

After the initial elicitation session, individual discussions took place with all partners in the project to clarify aspects of the initial session, and to ensure that all areas of risk in the project were covered, as well as allowing likelihood and severity scales to be better defined. This process is also used to add any additional FEPs not already included. For Case Study 1 this resulted in a new FEP category entitled 'Technical and Financial Viability', containing 5 FEPs related to industrial aspects of the project. These discussions are also used to develop likelihood and severity scales which are particularly attuned to project partner concerns (e.g. reflecting the potential scale of financial losses they are willing to accept), and to the specific sites considered. While the likelihood scale may be generic for many projects, it is important that the severity scale reflects the unique characteristics of, and encompasses the full range of consequences for, the project. For example in Case Study 1 the financial severity scale was change to reflecting the potential scale of financial losses the project partners were willing to accept,

In order to make tracking of the evolution of risk manageable, the FEP list was reduced, focusing it on those areas perceived to be in the moderate and high risk bands for the project. This was done to minimise the number of FEPs partners were required to assess therefore minimising the work and time involved. The risk assessment is repeated at regular intervals. Therefore it is important that project partners are prepared to complete the assessment regularly and minimising the work involved helps to ensure participation. Using the results from the initial elicitation session, the upper risk bound was used to rank the FEPs from low to high risk. To ensure a minimum level of expertise, only experts who ranked themselves as at least expertise level 3 (competent) were allowed to contribute to the pruning process for each FEP.

The first step was to remove all FEPs which were ranked to have a low upper bound of risk by all individuals. There were a number of FEPs where there was disagreement between participants of similar expertise with some ranking the FEP as low risk and others ranking the FEP as moderate risk. Such FEPs were considered borderline for exclusion from the register and were categorised as "for discussion by the group". All other FEPs were considered sufficiently risky for inclusion in the final register.

Based on the results of the initial elicitation and individual consultations, the risk analyst develops a proposed register with FEPs categorised as either *keep*, *remove* or *discuss*. The list of FEPs, including a complete description of each, is then distributed to the partners along with the proposed likelihood and severity scales. The partners are given time to review the FEPs, ensuring they are clear on the FEP meaning, and are able to propose changes to the register including the addition of new FEPs. For Case Study 1 this process resulted in 65 FEPs in the *keep* category, 32 in the *remove* category and 46 in the *discuss* category.

The likelihood and severity scales developed by the risk analysis from the consultation process are also distributed to the partners. For Case Study 1 this resulted in the severity scale being divided into two separate scales: 'health and safety' and 'finance, environment, research and industry viability'. The two scales were assessed separately to clearly distinguish between impacts on human health and other consequences.

The next step is to bring together all partners to discuss the proposed register as a group and reach semi-final agreement on scales and on FEPs to monitor. Having introduced the partners to the proposed likelihood and severity scales, participants are requested to suggest alterations to the proposed scales. After discussions they agree to the scales to be used for the duration of the project.

The group must then reach agreement on which FEPs to include in the final register. Given sufficient time, ideally the FEPs in all 3 *keep*, *remove* and *discuss* categories will be considered. However this part of the process is time-consuming and time with all experts together may be limited. Efforts should therefore focus first on reaching agreement over the FEPs "*to remove"* category. Participants are asked to read through the list and for each FEP confirm that it can be excluded from the final register. Attention then moves to the "*for discussion"* list with the default option being to include each FEP in the final register. Participants are asked to read through the list and confirm if each FEP in turn should be included in the final register. Finally the FEPs in the *keep* category should be discussed to confirm that no FEPs are included that could be removed, and to allow participants the opportunity to verify that all areas of potential risk are covered by the register.

For Case Study 1, time restrictions prevented discussion of FEPs in the *keep* category; however, agreement was reached on which FEPs in the *remove* and *discuss* categories to exclude, and a consensus was achieved over the final register.

After the final register has been agreed, the partners are given the opportunity to reflect and suggest any final changes within an agreed and fixed time period, whereafter the register is finally fixed. The finalised likelihood and severity scales for Case Study 1 are shown in Tables 2, 3 and 4.

Having reached final agreement, the register is distributed to all partners and experts for their risk evaluation. This first assessment is used as a baseline for the project against which future perception of risk can be compared as a means of judging the success of risk mitigation activities.

Each FEP is assigned to a small number of experts for subsequent assessment, with each expert given responsibility for assessing only a small number of FEPs. At regular intervals throughout the project (quarterly in Case Study 1), each expert is asked to complete their FEPs in the register according to a structured elicitation process. Experts are asked to assess the best-guess, lower bound and upper bound for the likelihood and severity. The best-guess value of risk is used to rank the FEPs as low to high risk and the range between the lower and upper bounds indicates the uncertainty. In addition to assessing the likelihood and severity, experts are asked to comment on their reasoning. These assessments are used to track the evolution of risk perception, and reasoning behind changes in perception, throughout the project lifetime.

Elicitation Process

Any judgement is subject to cognitive bias, introduced by the use of heuristics (Kahneman et al, 1982). Experts are not immune to these biases and indeed are more prone to some than non-experts. Steps must therefore be taken to reduce their effects when asking experts to make judgements in situations of uncertainty. Common biases are over-confidence, anchoring and adjustment, availability, and motivational bias. *Over-confidence* is intrinsic to experts as they tend to be expert in a selectively narrow field. It is often manifest as bounds on risk estimates being too narrow. *Motivational bias* occurs where an expert is not completely independent and their judgement is influenced by some conflict of interest. A common example is where an employee believes that a confident answer; or one consistent with a company view is more desirable. *Availability bias* results from the use of heuristics to judge the likelihood of an event. Instead of using the recollection of the number of similar events and their frequency to judge the probability of a future event, people tend to use ease of recollection as a short-hand for frequency. This can lead to bias with more recent or memorable events judged as more likely than more distant or hard to recall events. *Anchoring and Adjustment* is the tendency for people to 'anchor' a numerical judgement to some initial estimate of its value. In the light of new information they tend to adjust their initial estimate, rather than re-evaluate the situation from scratch, resulting in a bias towards the initial 'anchor' value. Another manifestation can occur in each risk assessment because experts are asked to assess lower, best guess and upper bounds for each FEP. The natural tendency in such situations is for an expert to assess the 'best guess' first, followed by the lower and upper bounds. This will tend to lead to a narrowing of the distribution with both the lower and upper bound anchored to the best guess. The lower bound estimate is judged to be higher due to anchoring to the best guess and an upper bound assessed as lower.

There is evidence to suggest that group elicitation may reduce individual bias, however group interaction has the potential to introduce other complex forms of bias. For example, interactions can lead experts to become overconfident in their assessments through a process known as 'group polarization' where the final consensus judgement is more extreme than the average of the group member's individual judgements (Sniezek, 1992). A group may also show a herding tendency (Keynes, 1921, 1937) where the experts incorporate the judgement of others in the group. While this process can lead to stable consensus distributions, there is a risk that the group can be guided in the wrong direction. Even in a well structured and managed elicitation session, herding behaviour can occur (Polson et al, 2009) and such sessions are time consuming making repeated assessments using this method difficult.

Explaining these possible biases to experts can reduce their impact, though not remove them entirely. Motivational bias in particular is relatively easy to overcome by stressing the importance of unbiased results. Availability bias is harder to reduce; however, directing experts to think of all relevant similar situations throughout their career rather than focusing on the most memorable or recent example can reduce its impact.

It is therefore important that experts are warned of these biases early in the process before the first assessment in step 2. In completing each assessment thereafter, experts should be reminded of the biases and directed to complete their assessments in a particular order: first the lower bound should be assessed, then the upper bound and finally the best guess. This should reduce the affect of anchoring to the best guess and widen the range of risks to more realistic levels.

Experts are directed to independently complete an assessment for each FEP for the likelihood and severity. The results from all experts are collected and analysed before revealing the results to any individual project partner.

Results

For each site, two assessments were completed by each expert, one using the "health and safety" severity scale and one using the "financial, environmental, research and industry viability" severity scale. The results for the first assessment are described below using the average results of the experts for the best guess, lower and upper bounds.

A second assessment was carried out 3 months later with the results showing little change from the first: little new information had emerged during the intervening time period to change the perspective of the experts. Most experts anchored to their first assessment with these experts showing little to no change in their responses. A handful of experts did however complete the assessment independently of the first with the changes in their results reflecting their uncertainty in answering.

Site One

Health and Safety

Shown in Figure 1 are the average best guess, lower and upper bound L x S risk rankings for each FEP for Site One using the health and safety severity scale. The FEPs are plotted according to increasing best guess L x S with highest ranked risk at the top. As the best guess of risk increases, so the range (upper-lower) also tends to increase, demonstrating how experts become more uncertain as the perception of risk increases. It should also be noted that when analysing individual expert's results, a pattern emerged which seemed to show that while experts are confident in assigning very low levels of likelihood and/or severity (i.e. range 1-1, 1-2), they are reluctant to assign very high levels of likelihood and/or severity (i.e. range 4-5, 5-5).

Only 11 FEPs rank as high risk or above, using the best-guess estimate. These are 'Construction of Pipeline', 'Geological Heterogeneities', 'Fractures and Faults', 'Undetected Features', 'Formation Pressure', ' Construction and Site Logistics', Financial Viability', 'Effects of pressurisation of reservoir on cap rock', 'Hydrogeology', 'Hydrological regime and water balance' and 'Near-surface aquifers and surface water bodies'. Five of these FEPs are related to the geology of the site, three to the industrial issues, two to the near-surface terrestrial environment and one to CO_2 interactions. If the upper bound of risk is considered then 48 FEPs would rank as high risk.

Financial, Environmental, Research and Industry Viability

Figure 2 shows the average best guess, lower and upper bound L x S risk rankings for each FEP for site one using the 'Financial, Environmental, Research and Industry Viability' severity scale. The FEPs are plotted according to increasing best guess L x S with highest ranked risk at the top. As with the health and safety scale, as the best guess of risk increases, so the range (upper-lower) also tends to increase.

Only 9 FEPs rank as high risk or above, using the best-guess estimate. These are in 'Financial Viability', 'Effects of pressurisation of reservoir on cap rock', 'Lithology', 'Geological Heterogeneities', 'Fractures and Faults', 'Undetected Features (Geology)', 'Formation Pressure', 'Hydrological regime and water balance' and 'Near-surface aquifers and surface water bodies'. Five of these FEPs are related to the geology of the site, two to the near-surface terrestrial environment, one to CO_2 interactions and one to industrial issues. If the upper bound of risk is considered then 56 FEPs would rank as high risk. The ranking of FEPs tends to be similar using both severity scales, however there are some differences, most notably 'Construction of Pipeline' and 'Construction and Site Logistics'. Both are industrial issues that were assessed as high risk on the 'Health and Safety' scale but moderate risk on the 'Financial, Environmental, Research and Industry Viability' scale

Site Two

Health and Safety

Shown in Figure 3 are the average best guess, lower and upper bound L x S risk rankings for each FEP for Site Two using the 'Health and Safety' severity scale. The FEPs are plotted according to increasing best guess risk with highest ranked risk at the top. Similarly to Site One, as the best guess of risk increases, so the range (upper-lower) also tends to increase,

Only 6 FEPs rank as high risk or above, using the best-guess estimate. These are in order 'Construction of Pipeline', 'Construction and Site Logistics', 'Financial Viability', 'Hydrological regime and water balance', 'Near-surface aquifers and surface water bodies' and 'Hydrogeology'. Three of these FEPs are related to industrial issues, two to the near-surface terrestrial environment and one to geology. If the upper bound of risk is considered then 54 FEPs would rank as high risk. For Site Two it is clear that the perception of risk associated with the geology of the site is lower than for Site One.

Shown in Figure 4 are the average best guess, lower and upper bound L x S risk rankings for each FEP for Site Two using the 'Financial, Environmental, Research and Industry Viability' severity scale. The FEPs are plotted according to increasing best guess risk with highest ranked risk at the top. Similarly to the other assessments, as the best guess of risk increases, so the range (upper-lower) also tends to increase.

Only 7 FEPs rank as high risk or above, using the best-guess estimate. These are in order 'Financial Viability', ' Effects of Pressurisation on Cap Rock', 'Hydrological regime and water balance', 'Near-surface aquifers and surface water bodies', 'Storage Concept', 'Hydrogeology' and 'Public Perception and Security. Two of these FEPs are related to industrial issues, two to the near-surface terrestrial environment, one to geology, one to CO_2 interactions and one to pre-closure CO_2 storage. If the upper bound of risk is considered then 62 FEPs would rank as high risk. As with the 'Health and Safety' assessment of Site Two, the perception of risk associated with the geology of the site is lower than for Site One.

Comparison of Results

Figure 5 shows the health and safety risk plotted against the financial, environmental, research and industry viability risk. There is a strong correlation between the two risk categories, with a correlation coefficient of 0.88 for Site One and 0.78 for Site Two. However as the risk increases the correlation between the two categories decreases. This shows that as perception of risk increases, the impacts are being assessed differently in each area.

Figure 6 shows the highest best guess L x S value of the two severity scales, for each FEP for both Site One and Site Two. This allows a direct comparison to be made between the perception of risk for each site. The perception of risk for both sites is generally similar in many areas, however there is a clear distinction in geology where site one is perceived as high risk while Site Two is perceived as moderate risk.

Uncertainty

It is clear that as an experts perception of risk increases, so does their uncertainty. Figure 8 shows the range (upper-lower) L x S against best-guess L x S. For all four assessments there is a strong linear relationship between the perception of risk (best-guess L x S) and the uncertainty (range).

In the second assessment experts confirmed that their perspective had not changed due to any new information or knowledge, so any changes in the results reflected the expert's uncertainty in answering. Most experts anchored to their first assessment with the result that the second risk assessment produced very similar results to the first.

However 4 experts completed the assessment independently to the first. Given the time between assessments, the changes between the two can give a good indication of the expert's uncertainty in answering. Each expert showed an average change per judgement of less than 0.8 suggesting that overall their uncertainty was low. From Figure 7 which shows the distribution of the changes in score for each judgement, it is clear that changes are generally small. Larger changes (±3 or more) are consigned to the negative side of the distribution indicating a decrease in the perceived likelihood or severity of an event at the second assessment. This suggests that where there is significant change, the expert's perception of risk has decreased as the project has progressed even though there is no new information, perhaps reflecting an increase in confidence as they become more familiar with the work.

Discussion

The highest risk overall for both sites is financial viability followed by construction of pipeline and effects of pressurisation of cap rock. However while hydrogeology is the only geological feature ranked as high risk for Site Two, lithology, heterogeneities, fractures and faults, undetected features and formation pressure are all also ranked as high risk for Site One. The areas perceived as higher risk for Site Two are CO_2 storage concept, and public perception and security.

The results from the first risk assessment show that Site One is perceived as slightly higher risk than Site Two and that much of the additional risk is associated with geological features. This would suggest that Site Two is either simpler or the more geologically secure of the two. There is more geological data available for Site Two and it is generally better quality than Site One suggesting that the geology of the region is better understood resulting in lower uncertainty. It is likely that it is the lower uncertainty associated with Site Two that leads to a lower perception of risk. Where the geology is know and understood the perception of risk is lower even where there are specific characteristics that have been identified that could pose a potential risk to the long term storage of CO_2. In trying to mitigate the geological risk associated with Site one, further investigation of the geology is key to reducing the uncertainty, which in turn may lead to a reduction in the perception of risk.

The purpose of risk analysis is to inform decision makers of areas where the risk is unacceptably high with results feeding directly into project decisions. In Case Study 1 decisions as to which data acquisition activities were carried out were partly influenced by the results of the first risk analysis. In deciding which particular activity to undertake, four factors were considered including the importance of the additional data. This was broken down into three parts including 'Gap in Existing Knowledge', 'Generic Value of Information' and 'Criticality or Risk' which was defined as the degree to which the new information may address or impact on areas identified as high risk by the assessment process. Here the risk assessment which highlighted high perception of risk associated with the geology of Site One directly informed which data acquisition activities took place with the result that the specific areas of the project identified as high risk due to uncertainty in the data or lack of data were addressed directly. While this process will not necessarily ensure reduction of risk, it will reduce uncertainty, helping to distinguish areas where the perception of risk is due to uncertainty and where perception of risk is due to specific concerns regarding well understood characteristics of the site.

In addition to informing data acquisition and risk mitigation activities, risk analysis can also advise on the absolute and relative security of different storage sites. In Case Study 1, Site Two is perceived to be the more geologically secure of the two sites. Should a decision into a preferred storage location or the suitability of a site be required in the future, the risk analysis results would form a significant part of the decision process as to whether a particular site was safe or viable.

Furthermore, the risk assessment process identified financial viability as the highest risk FEP for both sites. In order to understand the reasons for this a decision was made to break this FEP down further in five key points with the aim of addressing which specific aspects of project finance were critical to the risk. This in turn allows mitigation efforts to focus in these key features of the financial risk.

Acknowledgements

This work was supported by Technology Strategy Board (TSB) and the Engineering and Physical Sciences Research Council (EPSRC).

References

Bachu, S. Screening and ranking of sedimentary basins for sequestration of CO_2 in geological media. 2003. Environment. Geol. 44, 277-289.

Bachu, S., Geological sequestration of anthropogenic carbon dioxide: applicability and current issues. In Geological Perspectives of Global Climate Change; Gerhard, L., Harrison, W. E., Hanson, B. M., Eds.; AAPG Studies in Geology 47; American Association of Petroleum Geologist: Tulsa, OK, 2001; pp 285-304.

Celia, M. A, Bachu, S. 2003. Geological sequestration of CO2: is leakage unavoidable and acceptable? In Proceedings of the Sixth International Greenhouse Gas Technologies Conference, Kyoto, Japan, October 1-5, 2002; Gale, J., Kaya, Y., Eds.; Pergamon; Vol. I, pp 477-482

Farrar, C. D., Sorey, M. L., Evan, W. C., Howie, J. F., Kerr, B. D., Kennedy, B. M., King, C. Y. and Southon, J. R. 1995. Forest-killing diffuse CO_2 emission at Mammoth Mountain as a sign of magmatic unrest. Nature, 376, 675-678.

Hnottavange-Telleen, K. and Krapac, I. G. 2008. Risk-analysis methods and early results from MGSC phase III. 7th Annual Carbon Capture & Sequestration Conference

Jammes, L., Vu-Hoang, D., Faivre, O., Ramakrishnan, T. S. and Sengul, Mahmut. 2006. Measuring and modeling for site characterization a global approach. Proceedings CO2SC Symposium

Kahneman, D., Slovic, P., and Tversky, A. (eds), 1982. *Judgement under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge.

Keynes, J. 1921. A Treatise on Probability. Macmillan, London.

Keynes, J. 1937. The general theory of employment. Quarterly Journal of Economics, v. 51, p. 209-223.

Nordbottem, J. M., Celia, M. A., Bachu, S. and Dahle, K. 2005. Semianalytical solution for CO_2 leakage through an abandoned well. Environ. Sci., 39, 602-611.

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 Expert's Probabilities.* Wiley.

Quintessa Ltd. The generic CO_2 geological storage FEP database. http://www.quintessa.org/co2fepdb/PHP/frames.php

Sniezek, J. A. 1992. Groups under uncertainty: An examination of confidence in group decision making. Organizational Behaviour and Human Decision Processes, v. 52, p. 124-155.

Torp, T. A. and Gale, J. 2004. Demonstrating storage of CO_2 in geological reservoirs: The Sleipner and SACS projects. Energy, 29(9-10), 1361-1369.

Vivalda, C. and Jammes, L 2008. Probabilistic performance assessment methodology for long term subsurface CO2 storage. PSAM9 Conference.

Vivalda, C. and Jammes, L 2008a. The key role of expert judgement in CO2 underground storage projects. ESREL 2008 and SRA Europe Annual Conference

Table 1. Combine Likelihood and Severity (L x S). 1=negligible, 2-4=low, 5-9=moderate, 10-20=high, 21-25=very high.

		Likelihood				
		1	**2**	**3**	**4**	**5**
Severity	**1**	1	2	3	4	5
	2	2	4	6	8	10
	3	3	6	9	12	15
	4	4	8	12	16	20
	5	5	10	15	20	25

Table 2. Finalised likelihood scale for the Case Study 1 project.

Likelihood		*If there were 100 similar projects, impact related to this risk element (FEP) would occur…*
Improbable	1	Probably not at all, never
Unlikely	2	Fewer than three times among the 100 projects
Possible	3	5 to 10 times among the 100 projects
Likely	4	In around half of the 100 projects
Probable	5	In most or nearly all of the projects

Table 3. Finalised health and safety severity scale for the Case Study 1 project.

Severity of Impacts		Project Values *Health and Safety*
Light	1	Minor injury or illness, First Aid
Serious	2	Reversible health effect, lost time injury less than 3 days.
Major	3	Irreversible health effect, lost time > 3days
Catastrophic	4	Life threatening health effect, Fatality
Multi-Catastrophic	5	Multi-fatality

Table 4. Finalised financial, environmental, research and industry viability severity scale for the Case Study 1 project. Categories are assessed together and the highest ranking category used to rate severity.

Severity of Impacts		Project Values *Financial*	*Environment*	*Research*	*Industry Viability*
Light	1	< £500k	No modification to initial state	Little to no progress to 1 of 4 goals	Project lost time > 1day. Minor citations e.g. moving vehicle citations.
Serious	2	£500k - £5m	Modification to initial state within acceptable limits	Little to no progress to 2 of 4 goals	Project lost time > 1week. Regulatory notice with out fine. Local allegations of unethical practice or mismanagement
Major	3	£5m-£25m	Modification to initial state above acceptable limits but without damage	Little to no progress to 3 of 4 goals	Project lost time > 1month. Permit suspension. Major local opposition or substantial negative local media coverage
Catastrophic	4	£25m-£50m	Modification to initial state above acceptable limit with repairable damage	Little to no progress to 4 of 4 goals	Project lost time > 1 year. International media coverage of law violations, questionable ethical practices or mismanagement.
Multi-Catastrophic	5	>£50m	Considerable modification to initial state which is not repairable with existing technologies	No gain in understanding applicable to future projects	Negative public experience results in legal ban on similar projects

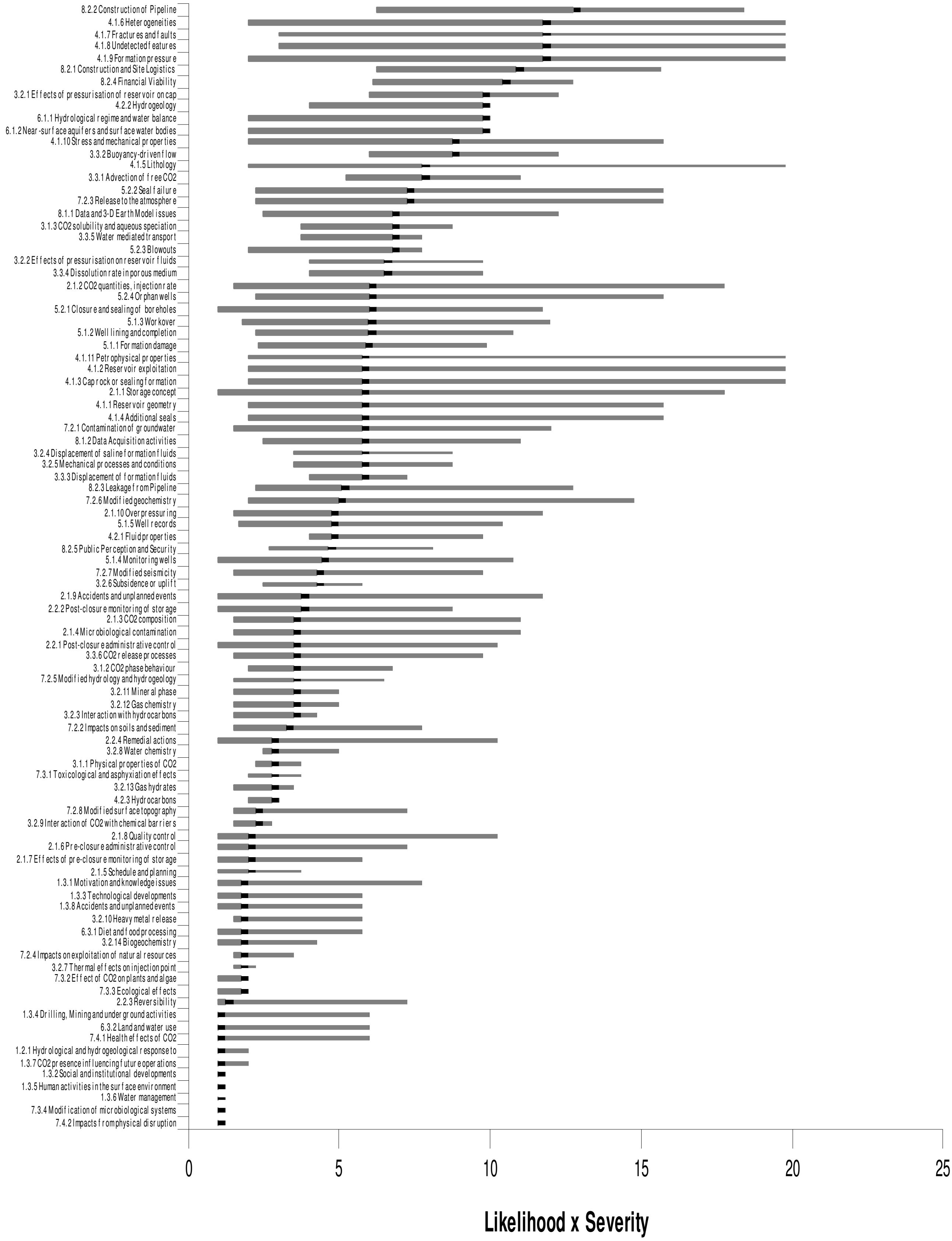

Figure 1. Likelihood (L) x Severity (S) best guess (■) and range (■) from average individual L and S values for lower, best-guess and upper bounds for Site One and health and safety severity scale.

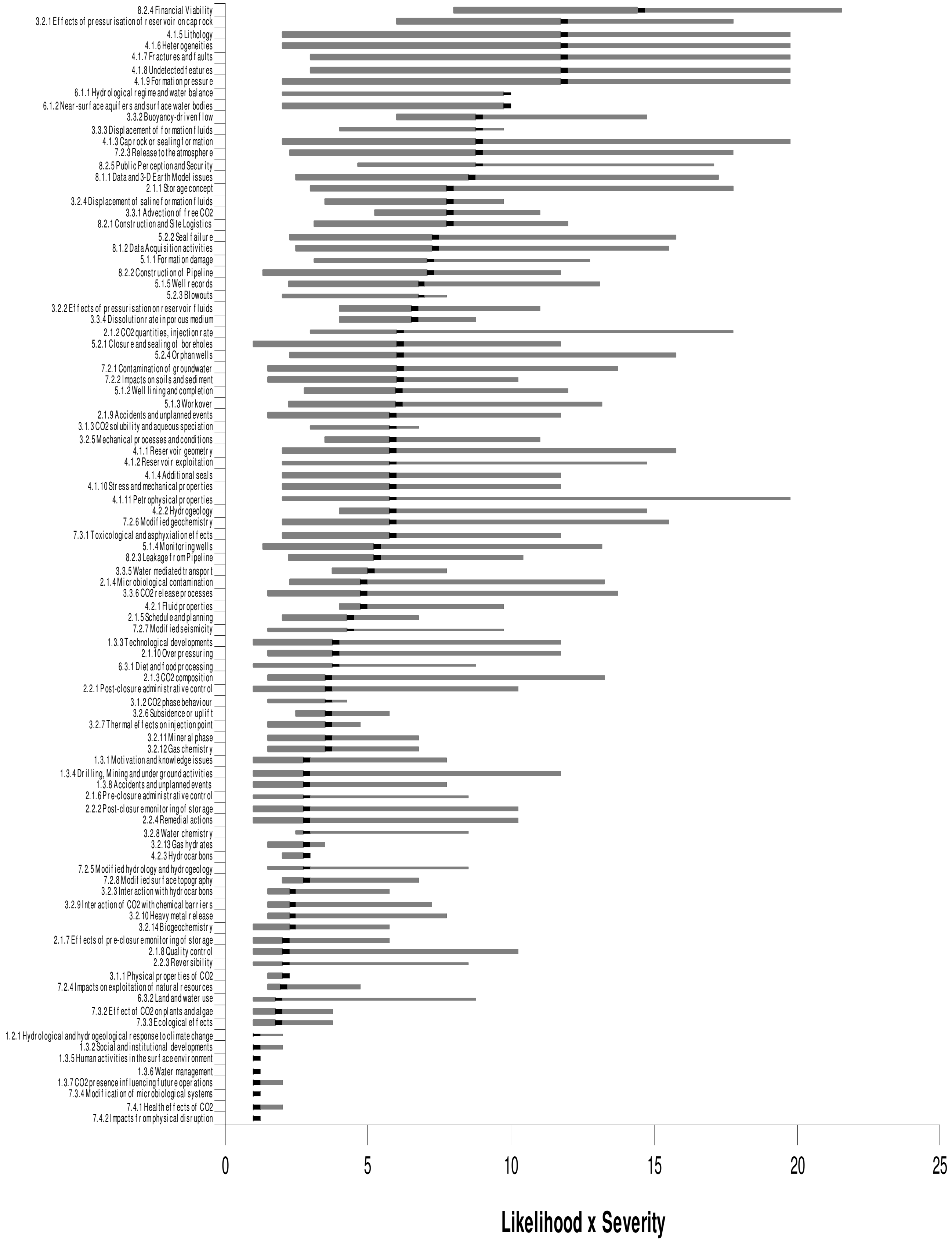

Figure 2. Likelihood (L) x Severity (S) best guess (■) and range (■) from average individual L and S values for lower, best-guess and upper bounds for Site One and financial, environmental, research and industry viability severity scale.

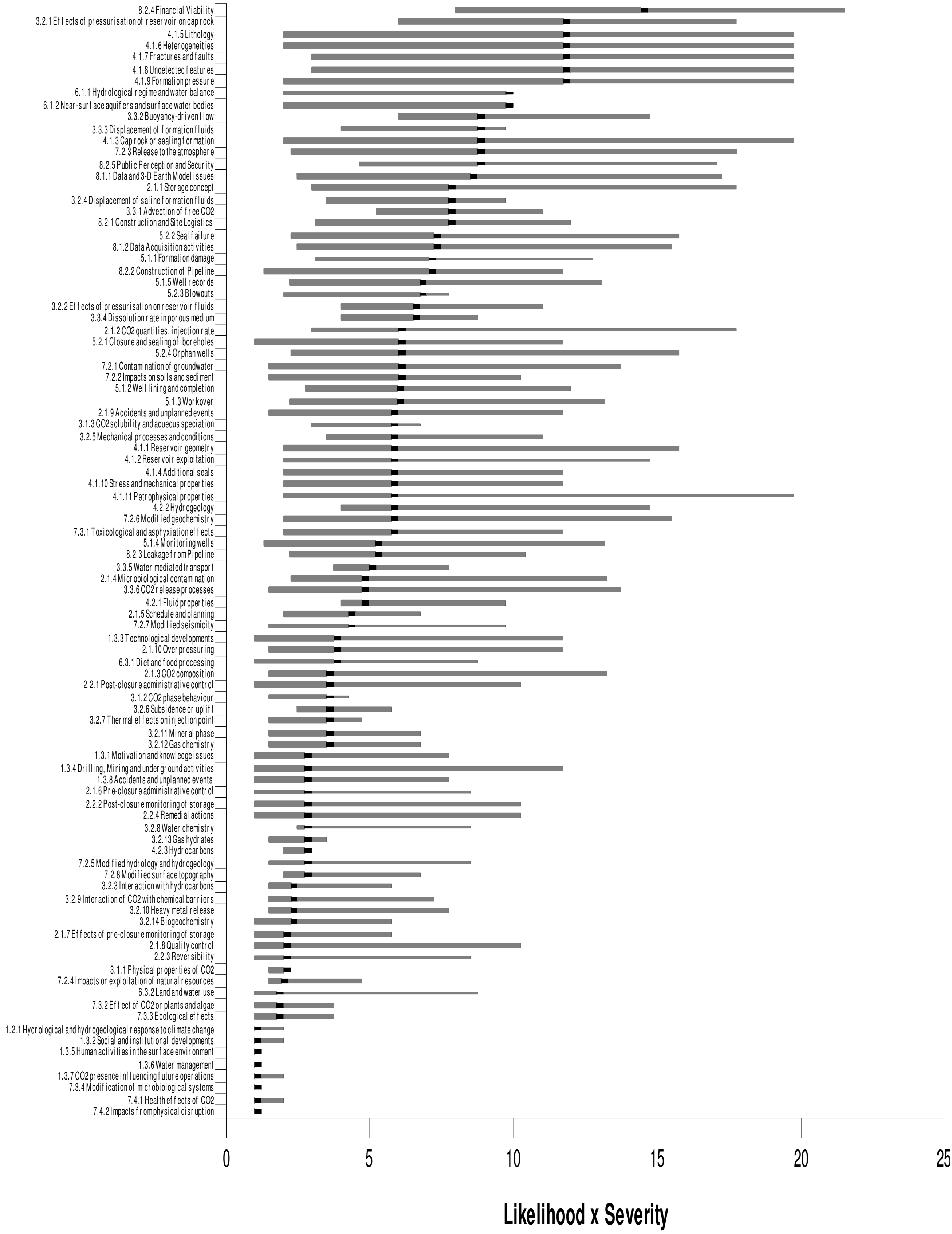

Figure 3. Likelihood (L) x Severity (S) best guess (■) and range (■) from average individual L and S values for lower, best-guess and upper bounds for Site Two and health and safety severity scale.

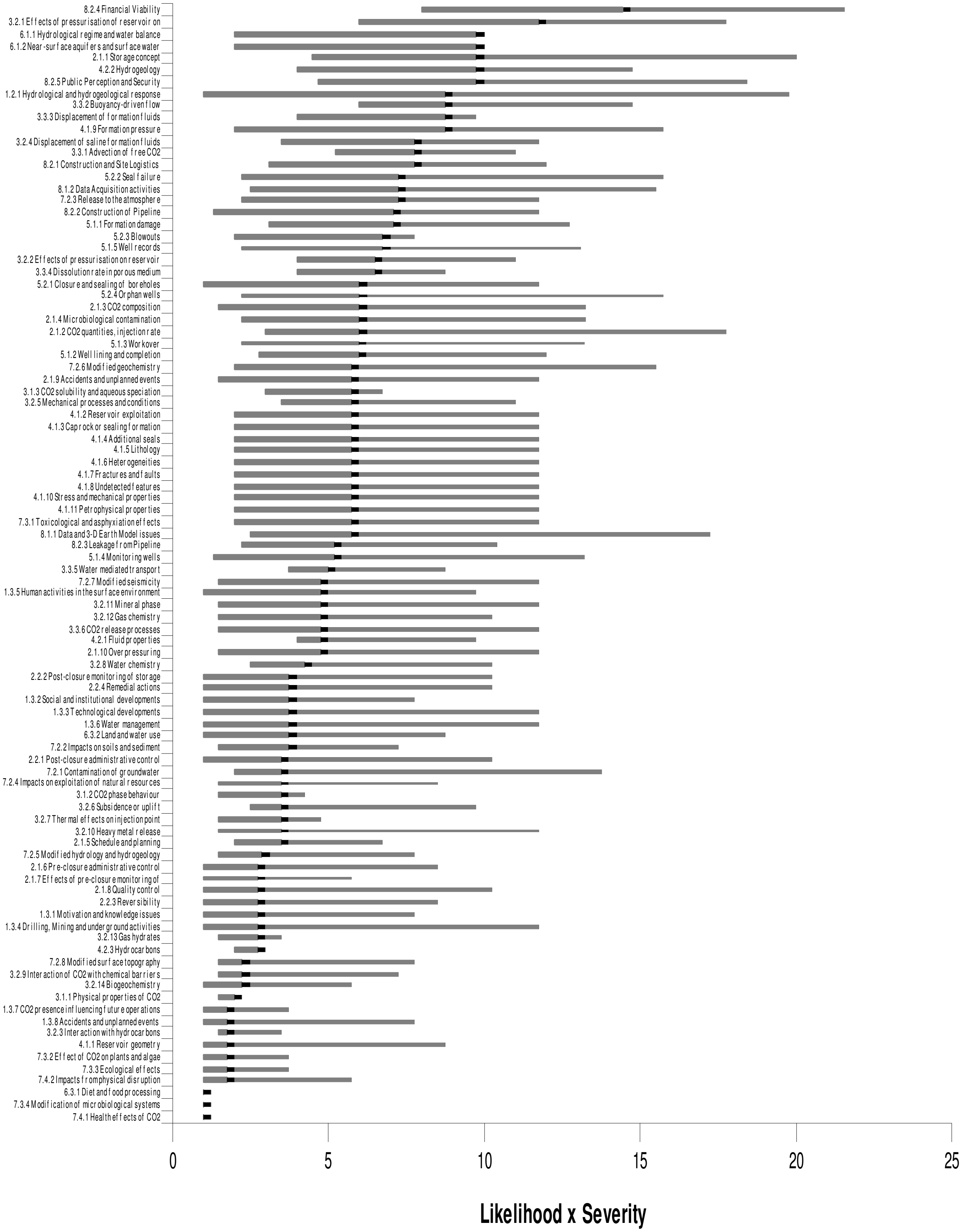

Figure 4. Likelihood (L) x Severity (S) best guess (■) and range (■) from average individual L and S values for lower, best-guess and upper bounds for Site Two and financial, environmental, research and industry viability severity scale.

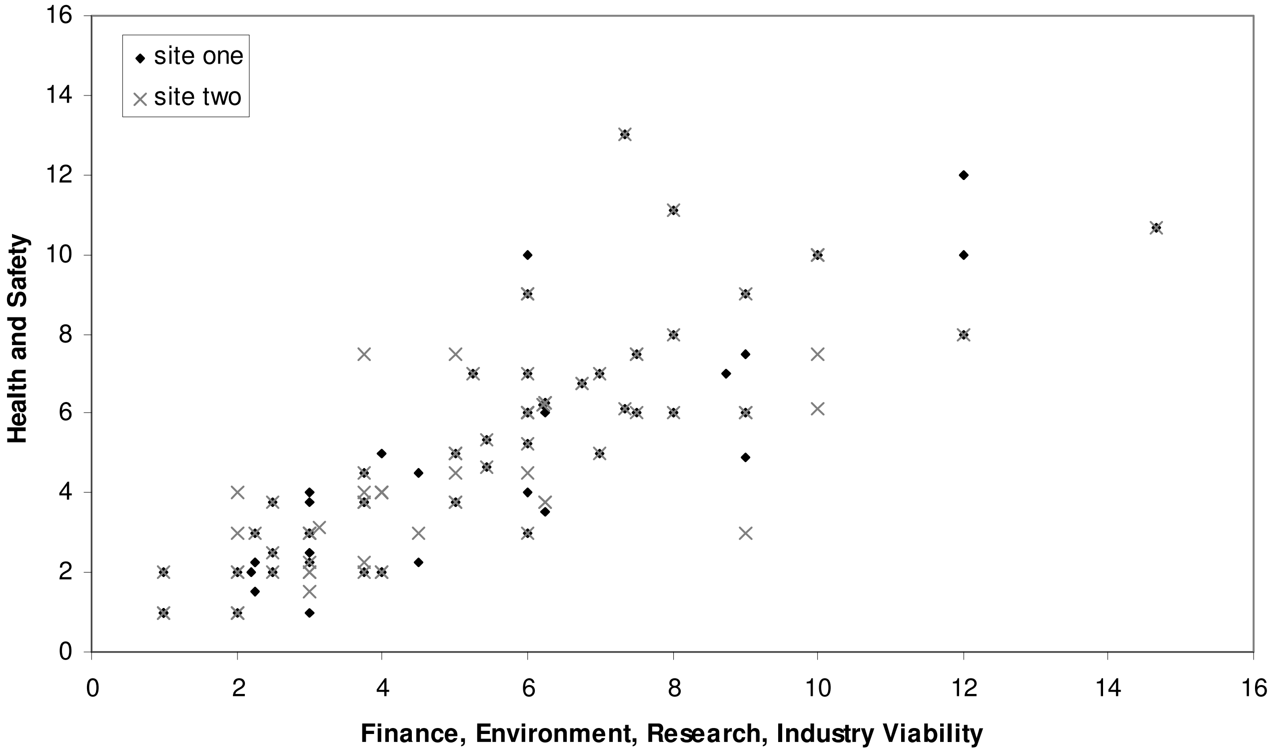

Figure 5. Health and safety risk versus financial, environmental, research and industry viability risk.

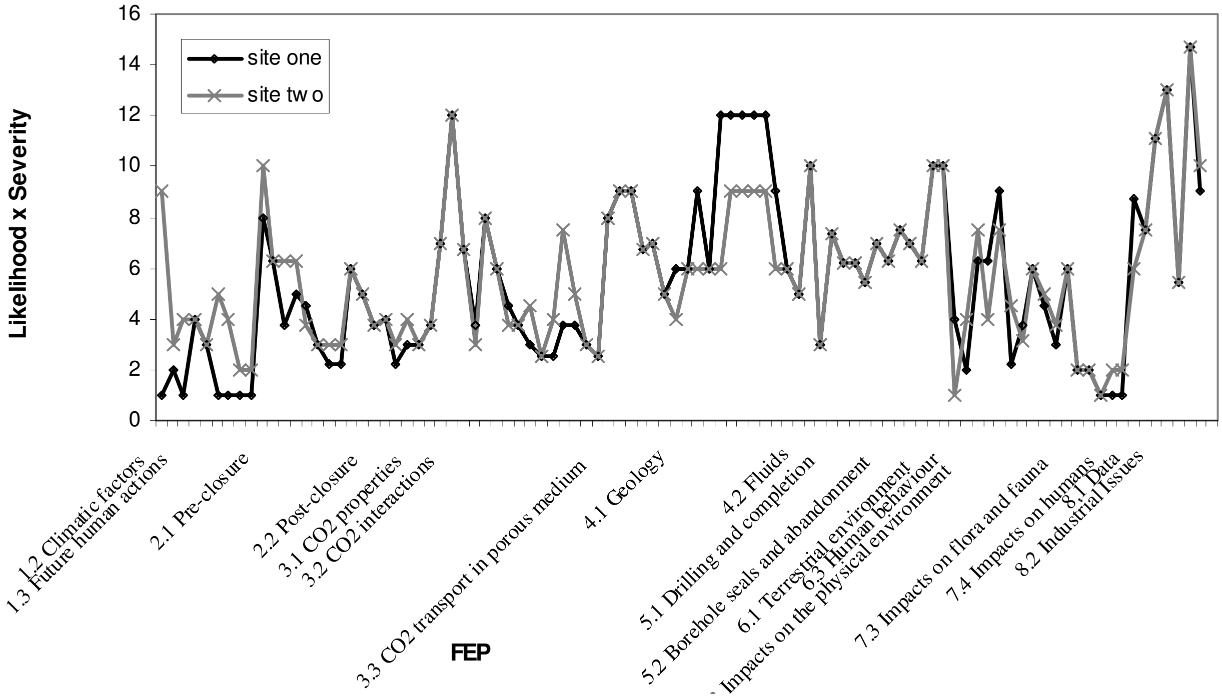

Figure 6. Highest best guess L x S from two severity scales for both Site One and Site Two.

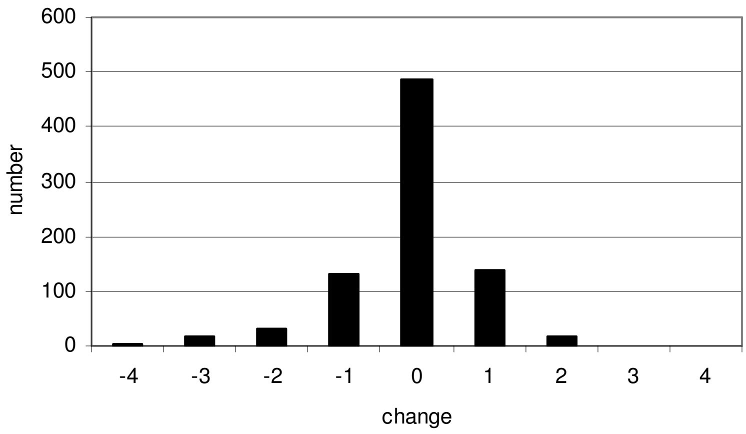

Figure 7. Change in score for each judgement.

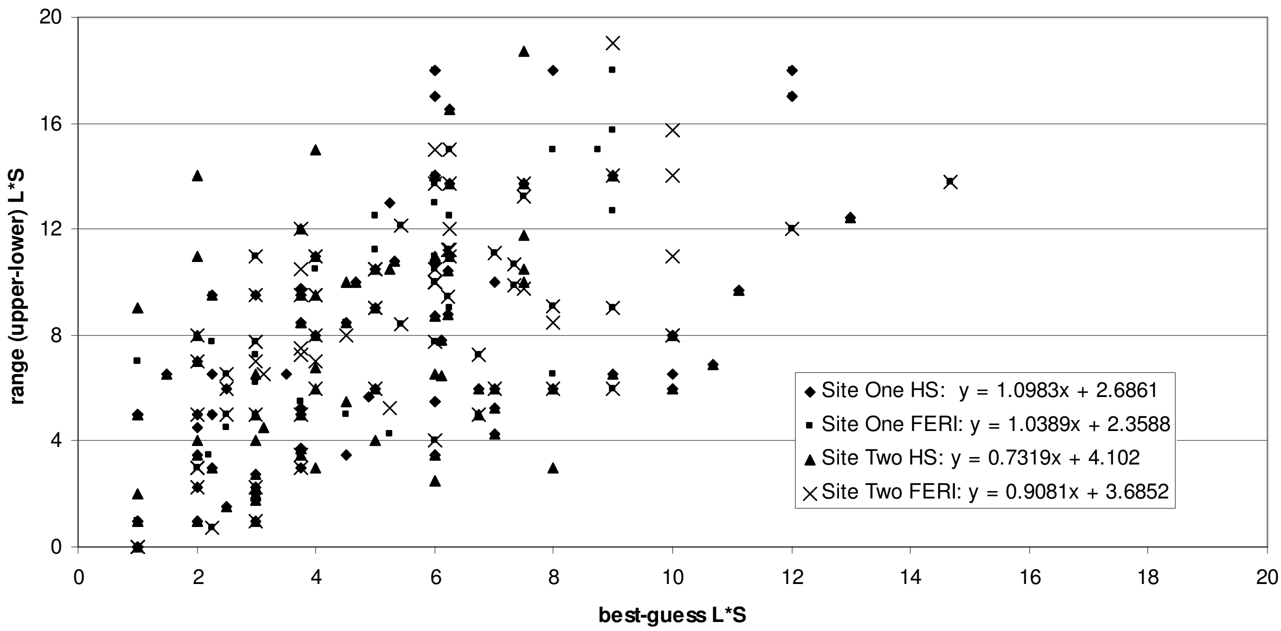

Figure 8. Range (upper-lower) likelihood x severity (L x S) against best-guess L x S