A practical approach for applying Bayesian logic to determine the probabilities of subsurface scenarios: example from an offshore oilfield

P. Craig Smalley* - BP Centre for Business and Technology, Sunbury TW16 7LN, UK

Christopher D. Walker & **Paul G. Belvedere** - BP America, 501 Westlake Park Blvd, Houston, Texas, 77079, USA

* Current address: Department of Earth Science & Engineering, Imperial College, London SW7 2AZ, UK

Abstract

During appraisal of an undeveloped segment of a producing offshore oilfield, three well penetrations revealed unexpected complexity and compartmentalization. Business decisions on whether and how to develop this segment depended on understanding the possible interpretations of the subsurface. This was achieved using the following steps that incorporated a novel practical application of Bayesian logic:

1. Scenarios were identified to span the full range of possible subsurface interpretations. This was achieved through a facilitated cross-disciplinary exercise including external participants. The exercise generated 12 widely differing subsurface scenarios, which could be grouped into 4 types of mechanism: slumping, structural, depositional and diagenetic.
2. Prior probabilities were assigned to each scenario. These probabilities were elicited from the same subsurface team and external experts who performed (1), using their diverse knowledge and experience.
3. The probabilities of each scenario were updated by evaluating them sequentially with 21 individual pieces of evidence, progressively down-weighting belief in scenarios that were inconsistent with the evidence. For each piece of evidence the likelihood (chance that the scenario could produce the evidence) was estimated qualitatively by the same team using a "traffic light" high-medium-low assessment. Offline, these were converted to numerical likelihood values. Posterior probabilities were derived by multiplying the priors by the likelihoods and renormalizing to sum to unity across all the scenarios.
4. The most probable scenarios were selected for quantitative reservoir modelling, to evaluate the potential outcomes of business decisions, given each scenario.

Of the 12 scenarios identified in step 1, most were strongly down-weighted by the sequential revisions against evidence in step 3; after this, only scenarios in the "slumping" group retained significant posterior probabilities. The data showed minimal sensitivity to the initial assumption of prior probability in step 2.

This process had several benefits. Firstly, it encouraged the subsurface team to imagine a full range of scenarios that were likely to bracket the actual subsurface "truth", something that is critical for subsequent decision making. Secondly, it allowed belief in the probability of each scenario to be updated systematically in a way that was strongly conditioned to the evidence, so that the choice of scenarios to take through to reservoir modelling was more objective and evidence-based. Thirdly, it

allowed an assessment of the usefulness of individual pieces of evidence, which could be used to guide value-of-information assessments for subsequent data acquisition. Finally, the process enabled rigorous Bayesian revision methods to be applied in a simple practical way that engaged the subsurface team without exposing them to the underlying mathematics. During field appraisal and development, when the subsurface is revealed gradually as more data are acquired and studied, the process outlined here provides a practical way of generating and modifying belief in a range of subsurface scenarios while minimizing exposure to potential biases and logical fallacies that could affect subsequent decision quality. It also helps to decide which scenarios are sufficiently probable that they need to be represented by detailed reservoir models.

Introduction

Describing and understanding the subsurface of an oil and gas asset is necessarily an incremental process, as gradually more data become available through time. For example, prior to exploration drilling, knowledge of the subsurface of a prospect may be based on seismic data and, if available, information from wells in other parts of the play. After the exploration well has been drilled and a discovery has been made, new information is available from the well, increasing knowledge of the subsurface. During appraisal, this may be augmented by improved seismic coverage and quality, more well penetrations, core analysis and well test data. During development and production there will be further well penetrations and dynamic production data, all of which progressively increase the understanding of the subsurface, constraining the possible subsurface scenarios that could be present and their probabilities. (Note that in this paper, the term “subsurface scenario” is used to describe a discrete internally consistent view of the subsurface)

At each of these stages of field development various business decisions need to be made. Making wise decisions using rigorous decision risk analysis methods requires a good quantification of uncertainty at each stage (Spetzler et al., 2016; SPE, 2016). This paper examines the process by which knowledge of, and uncertainty in, the subsurface is updated as new data or the results of technical studies become available. In particular, the question is addressed of how to apportion appropriate weight between previous existing information and the new information. We focus on field appraisal, a time when there is usually a rapid influx of new information.

There is a standard way of approaching such problems, using Bayesian updating (e.g. Van Wees et al., 2008; Willigers et al., 2013). Nevertheless, examination of the relevant literature indicates that this is not routinely applied in practice in subsurface teams. Possibly this is because the mathematics of probability is perceived as difficult, and the relevance of the Bayesian logic to real subsurface situations is not immediately evident to non-specialists.

This paper describes a simplified intuitive Bayesian approach that, whilst being technically rigorous, is easy to adopt in practice by subsurface teams; this is demonstrated through a case study of the appraisal of an offshore oil field.

Current pitfalls

There are presently various pitfalls that can cause problems when trying to quantify changes to uncertainty in response to new information. These are usually related to various motivational and cognitive biases that are well known to affect perceptions of uncertainty in general and of the subsurface in particular (Milkov, 2015; Montibeller and von Winterfeldt, 2015; see also the chapter "Behavioral Challenges in Decision Making" in Bratvold & Begg, 2010). Examples with potentially serious consequences for subsurface description include:

Not identifying the full range of uncertainty. Many surprises in field development and performance outcomes can be traced back to a subsurface scenario occurring that was never envisaged (Caers, 2011). Scenarios that have never been conceived cannot be planned for or managed. This problem has its roots in biases such as overconfidence (we think we know more than we do) and anchoring (we have an irrational attachment to one particular scenario and thus consciously or sub-consciously ignore scenarios that are significantly different from this).

Over-weighting prior information. This is reflected in a reluctance to change from an existing view, despite new information to the contrary, resulting in an inaccurate assessment of remaining uncertainty. This situation may be in some cases attributable to the degree of personal investment in current scenarios and models; if a large amount of resource has been expended to work up a scenario and to build static or dynamic models to represent it, there may be a "sunk cost" effect (Montibeller and von Winterfeldt, 2015) leading to a conscious or unconscious motivational bias to stick with the modelled scenario rather than creating more work by considering new scenarios.

Over-weighting new information. Having often worked hard to justify expensive data acquisition, and having invested in state of the art data, there may be a tendency to overweight information from this source compared to prior data. This is a manifestation of the availability bias, where excessive weighting is given to information that is more easily called to mind or favoured. This pitfall can lead to lurching from over-belief in one scenario to another with every new piece of information.

These pitfalls can be avoided by adopting two measures: (1) Identifying the full range of subsurface scenarios at the outset, before building detailed reservoir models, through a facilitated cross-disciplinary team exercise with representation from external experts, and (2) using a practical implementation of Bayes' Rule to update the probabilities of each scenario so that the probabilities are appropriately supported by the available data. Step 2 involves viewing relevant data and interpretations as individual pieces of evidence that can systematically be evaluated as to their likelihood of supporting each possible scenario. The process described here is best applied at an early stage, before complex reservoir models have been created; it can be used to help decide which are the most probable subsurface scenarios need to be taken through to more detailed reservoir modelling. The actual reservoir modelling is not addressed in this paper. The theoretical basis for step 2 is described in the following section. This process is then illustrated through a worked case study.

Theoretical basis

The purpose of this paper is to present an approach to Bayesian updating of probabilities that is accessible to subsurface teams. We believe an important element in achieving engagement with such teams is to avoid the explicit use of mathematical formulae, and to keep the process visual, using images rather than numbers. However, some background to the theoretical approach is useful.

When updating probabilities based on new information, Bayes Rule is applicable:

$$P(S|E) = P(E|S) \times P(S) / P(E) \quad (1)$$

Where P(S|E) is the **posterior** (updated) probability of scenario S being true, given that evidence E has been observed; P(E|S) is the **likelihood** that, if scenario S was true, evidence E would be observed; P(S) is the **prior** probability of scenario S being true before evidence E has been considered, and P(E) is the probability of evidence E being observed.

In practice, P(E) is often similar between the different scenarios that are being evaluated. Bearing this in mind, and making the assumption that the identified subsurface scenarios cover the full range of possible subsurface realities, P(E) can be substituted by a normalising factor to ensure the posterior probabilities P(S|E) for all scenarios sum to unity. This leaves:

$$P(S|E) \propto P(E|S) \times P(S) \quad (2)$$

which can be read as the posterior probability being proportional to the likelihood multiplied by the prior probability (Lee, 2002).

In the approach used here, the scenarios are identified first, making a concerted effort to include all possible scenarios. Then a prior probability is assigned to each scenario based on expert opinion. It is known that such probability estimation can be adversely affected by cognitive and motivational biases. Steps are taken at this stage to minimise the effects of bias (e.g., Baddeley et al., 2004; Curtis & Wood, 2004).

Next, a new piece of evidence is considered. For this evidence, the likelihood is estimated by making the assumption that the scenario being considered is true, and then eliciting from experts the probability that the observed evidence could be explained by that scenario. A likelihood of 1 means that the evidence is completely consistent with the what would be expected if the scenario was true, and 0 means that it is impossible that the scenario could have generated the observed evidence; there is, of course, a gradation in between these extremes. The estimated likelihood is generated for each scenario and then multiplied by the prior probabilities. The products for each scenario, when normalized to sum to one, become the posterior probabilities in equation 2. This process is repeated sequentially for each piece of evidence. At each step, the posterior from the previous iteration becomes the prior for the next.

In practice, it is difficult to elicit a completely unbiased view of the initial prior probability. However, if the process described above is repeated for numerous pieces of evidence, the posterior probabilities progressively become dominated by the evidence and less dependent on the initial assumed priors. The scenarios that are the most probable, once they have been through several cycles of updating against the evidence, are the ones that may then be selected for more detailed quantitative reservoir modelling.

Consideration of Equation 2 demonstrates the importance of identifying all relevant scenarios in the first place. If not identified, their prior probability P(S) is effectively zero, and so new data can never change their probability.

This process is now applied in a case study.

Case study from an undrilled segment of a producing oil field

Background

The methods described above were applied in a case study of an offshore oil field. The field in question has been producing for over a decade. This study focuses on the appraisal of an undeveloped fault-bounded extension to the producing field. Figure 1 shows the expected fluid distribution in the potential field extension before appraisal drilling. The area to be appraised was predicted to contain oil down to the regional oil-water contact (OWC) in all three of the known reservoir sands, by analogy with most of the rest of the field. However, the presence of water could not be discounted, as an adjacent area shown in blue on Figure 1 had previously been found unexpectedly to contain water at a level structurally higher than the regional OWC. Three appraisal penetrations were made in the field extension – the original hole (OH) plus sidetracks 2 and 3 (ST2, ST3); these are shown in map form on Figure 2 and cross section in Figure 3.

The OH results came in largely as expected, with all three reservoir sands (upper, middle and lower) present and full to the base of the well with oil. The sands were variably depleted from production in the main part of the field, and this was interpreted to indicate that the mapped fault separating the extension from the producing part of the field (Figs. 1 and 2) had only had a baffling effect to pressure transmission on a production timescale and would probably not form a significant barrier to oil flow over geological time (i.e. during reservoir filling).

The results for ST2 and ST3 were much more complex than expected. ST2 found oil in the upper sand, but water in the middle and lower sands. Frustratingly, oil-water contacts were not observed, and it was not possible to determine whether the contacts differed significantly from the extrapolated regional OWC (Fig. 3). Pressures were only slightly depleted in each sand, not as much as would have been expected if this part of the reservoir was well connected to the producing area of the field or to the OH, so the existence of production timescale flow restrictions between ST2 and the OH was interpreted. The ST3 penetration was placed between ST2 and the OH, in an attempt to tag the OWC. It encountered all three reservoir sands, with oil in the upper and middle sands, and water in the lower sand, but the highest known water in the lower sand was higher than the lowest known oil in ST2 (Fig. 3). The measured or inferred OWC depths (Fig. 3) are comparable with the regional OWC in ST2, but several hundred feet higher in ST3, indicating that the lower sands in ST3 are isolated from the regional aquifer. Using the analytical expression for equilibration of OWCs from Smalley and Muggeridge (2010), the OWC would be expected to equilibrate over the ~400m (1300 ft) inter-well distance in a few tens of thousands of years, so the preservation of OWC differences over geological timescales indicates a significant barrier between the lower sand in the OH and middle sand in ST3. Pressures in ST3 were less depleted than in the OH, and yet more depleted than in ST2. The differences in pressure depletion laterally within each sand indicate flow baffles between the wells on a production timescale.

Water samples from the OH spun from core have higher salinities than those sampled from other parts of the field, including the regional aquifer; this also suggests that the OH is isolated from the main aquifer over a geological timescale, assuming that any mixing is driven by aquifer flow rather than much slower diffusion (Go et al., 2014).

Key development decisions needed to be made for this part of the field: for example, should it be developed at all? If so, should it be developed as a subsea satellite tied back to the main field, or should it be developed by drilling from the existing platform? Would water injection be required and in what pattern? Such decisions are strongly affected by the view of reservoir flow barriers and baffles, the potential mechanisms causing them, and how they might affect the distribution of fluids. This necessitated an examination of possible compartmentalization scenarios. Compartmentalization here refers to the presence of any features that could prevent (barriers) or inhibit (baffles) fluid flow on a production timescale, or on a geological timescale sufficient to affect fluid distribution.

Step 1. Identify scenarios

Based on the data from the wells, a variety of subsurface scenarios were conceived that could potentially explain the mechanisms of reservoir compartmentalization.

When identifying scenarios, motivational and cognitive biases can come into play. An example of motivational bias might be that, because the field team have a stake in the project and may benefit from a successful development of the area being appraised, they may consciously or subconsciously skew the interpretations to be more favourable to development. An example of a cognitive bias could be that the field team are anchored to an existing geological interpretation of the field, and thus fail to identify scenarios that vary significantly from this, thus underestimating the uncertainty in the subsurface.

The effects of biases were minimized by assembling a multidisciplinary team who were familiar with the field, plus others who were familiar with the regional geology, but worked outside the field team. In addition, external (to the region) experts were included. The intent was that the inclusion of different disciplines and external expertise would reduce the chance that the group was anchored towards one particular type of interpretation. The use of external experts who did not have a stake in the success of the project was designed to minimize motivational bias.

The result was a very wide range of possible subsurface scenarios, listed in Table 1, which covered all of the key mechanisms that could cause the observed compartmentalization effects. The scenarios fell into 4 groups, based on mechanisms related to (1) slumping, (2) faulting, (3) deposition or (4) diagenesis. Each scenario was visualized by generating a simple graphic (Figs. 4-7) so that the team all had a common understanding of how that scenario could lead to the observed flow barrier and baffle effects.

The four slumping related scenarios (1a-d) relate to movement soon after deposition due to various mechanisms (Fig. 4):

- Scenario 1a involves small-displacement slumps triggered by movement off a paleobathymetric high causing imbricate faults that act as flow barriers.
- Scenario 1b is where a retrogressive slump failure is initiated by channels down-cutting into earlier layers, creating a gravitational instability that causes movement over the top of an overpressured shale. The resulting faults become flow barriers.
- Scenario 1c is where a large-scale gravity slide down a regional bathymetric gradient is interrupted by encountering another high. The resulting down-dip compression creates thrust faults that compartmentalize the reservoir.

- Scenario 1d involves rapid turbidite deposition onto unstable substrate which then initiates local subsidence, disruption and mud injection; in this scenario, it is the injected mud that creates flow barriers.

The three faulting scenarios (2a-c) relate to structuration after deposition and consolidation of the reservoir sands (Fig. 5):

- Scenario 2a invokes large faults between the appraisal wells that are sufficiently large to compartmentalize the reservoir but are just below the threshold for being visible given the resolution of the seismic data.
- Scenario 2b involves zones of smaller-scale deformation between the wells, for example small faults that are well below the current seismic resolution ("sub-seismic faults"), deformation bands related to bending stresses, or boudinage in sands created by flexural slip in bounding shales.
- Scenario 2c relates to tar-filled fractures, where overpressured source rock from depth has hydraulically fractured into the overburden, and created continuous, through-going, impermeable, bitumen-filled fractures.

The stratigraphic scenarios (3a-c) involve depositional architecture creating compartments by means of off-lapping or down-cutting channels that are lined with shale. These are illustrated in Figure 6:

- Scenario 3a has compartmentalization caused by zones of amalgamation in stacked channel deposits with long axes oriented normal to the line of appraisal wells. In this scenario, there are short-distance facies variations and clay-prone sands and silts deposited towards the edges of the channels.
- Scenario 3b involves shale drapes over compensation lobes. In this scenario, the reservoir sands would be deposited towards the edges of submarine fans where sands are less channelized and shales with larger lateral extent can be developed and preserved. The preserved shale drapes cause the compartmentalization.
- Scenario 3c is where turbidite deposition is influenced by paleotopography of the sea bed. Syn-depositional paleobathymetric relief caused decreased flow velocity as the turbidites approached the topographic high. Finer material was preferentially deposited along particular zones, degrading reservoir quality sufficiently to affect reservoir connectivity.

Two diagenetic scenarios were also identified (4a,b), illustrated in Figure 7:

- Scenario 4a is where the aquifer is heavily cemented and the resulting low permeability creates a bottom seal, inhibiting pressure support from the aquifer being transmitted to the wells.
- Scenario 4b involves differential diagenesis. Small scale variations in primary sediment composition or subsurface conditions (e.g. waters moving up fractures) enhanced diagenesis (e.g. quartz or carbonate cementation) in certain areas, causing permeability reduction between the well locations.

The team was confident that this wide range of diverse scenarios covered all the possible compartmentalization mechanisms that could account for the observations from the appraisal wells.

Step 2. Assign prior probabilities

Based on experience from elsewhere in the field, and in the wider region, it was evident to the team that the 12 scenarios identified were not equiprobable, so the probability of each scenario was estimated. This was done in a workshop environment facilitated by an expert in probability and uncertainty. The team was made aware of cognitive biases that can influence the behaviour of groups, such as anchoring, the authority fallacy, herding and trust heuristics (Milkov, 2015). To minimize these effects, the prior probabilities for each scenario were estimated individually by members of the multi-disciplinary team in isolation. Their estimates were then averaged after eliminating anomalies, rather than being adjusted by group consensus. The probabilities were then normalized to sum to one across all the scenarios. These estimated prior probabilities are shown in Table 1. Clearly, the scenarios related to slumping were seen as the front runners, followed by scenario 2a that invoked larger but un-imaged faults.

The sensitivity to the estimates of prior probabilities is examined later, as are possible improvements to the estimation methodology.

Step 3. Update probabilities by sequential consideration of evidence

The scenarios identified in step 1 and assigned prior probabilities in step 2 were then tested against various pieces of evidence – observations and interpretations derived from a range of data types, shown in Figure 8. For each piece of evidence, the subsurface team as a facilitated group made a qualitative estimate of the likelihood (i.e. $P(E|S)$ in Equation 2) that each scenario could be capable of generating that evidence, by assigning a traffic-light (red, yellow or green) colour to each evidence-to-scenario combination. An assessment of “red” means that it is judged highly unlikely that the scenario could have generated the observed evidence, yellow means that it is somewhat likely, and green means that the evidence is exactly what would be expected if the scenario were true. In addition, some pieces of evidence were marked as “grey”, where it was felt that the scenario had no bearing on the evidence.

The first 19 pieces of evidence were available at or soon after the time that the scenarios were identified. Sometime later, reprocessed seismic data became available which had better resolution and definition, allowing improved imaging of the reservoir layering and faulting, which were two key uncertainties. This dataset provided two important new pieces of evidence. In map view, the two down-dip sidetracks were seen to be situated in an almond-shaped area of poor and disturbed reflectivity. In cross section view, the boundary between the disturbed and background reflectivity was interpreted as a scoop-shaped detachment. Both pieces of evidence were judged as consistent with the slumping-related scenarios, but inconsistent with the faulting-, deposition- and diagenesis-related scenarios, leading to red traffic light assessments (Fig. 8).

To use the evidence to update quantitatively the probabilities for each scenario, it was necessary to convert the indicative traffic light colours (Fig. 8) into numerical likelihood values. This was done offline, after the elicitation workshops described above. In this analysis, the following assumptions were made. “Red” was assigned a likelihood of 0.1, indicating that it was unlikely that the evidence could be consistent with the scenario. Note that no zero likelihoods were assigned; this would have rendered the posterior probability for the affected scenario as zero, and this was seen as unreasonably definite given the uncertainty involved in all of the pieces of evidence. “Yellow” was assigned a value of 0.7, and “Green” was given a 1.0, signifying complete consistency between the scenario and the evidence. “Grey” was also assigned a likelihood of 1, as lack of relevance should not be regarded as positive evidence against a scenario. The likelihoods are shown on Figure 8.

Using these likelihoods, the probabilities of each scenario were updated sequentially for each piece of evidence, i.e. in 21 steps. In each step, the previous posterior probability was used as the new prior, this prior was multiplied by the likelihoods (Fig. 8), and the products normalized to sum to one across all the scenarios, following Equation 2. With the application of each piece of evidence the relative probabilities of each scenario changed (Fig. 9).

After the first 19 pieces of evidence, that is having considered all the data initially available, but prior to the reprocessed seismic data becoming available, the scenario with the highest probability was 2a (large faults) at 0.45, followed by the slumping-related scenarios 1a and 1b. The large faults scenario (2a) involves having faults large enough to compartmentalize the reservoir and cause the observed variations in pressure and fluid distributions, but just below seismic resolution. The reprocessed seismic information was thus a potentially critical piece of evidence, as the improved resolution of the reprocessed seismic data could have revealed such faults. Despite its higher resolution there was still no evidence for such faulting from the reprocessed seismic data. The likelihood that scenario 2a could explain the observations from the reprocessed seismic data (evidence 20, scoop-shaped detachment, and evidence 21, arcuate low reflectivity zone) was thus expressed as a “red” (Fig. 8). The depositional and diagenetic scenarios (3a-c, 4a-b) similarly are unable to explain these observations. On the other hand, the reprocessed seismic observations were consistent with the slump-related scenarios 1a-c, though less consistent with 1d (local subsidence), which would have been unlikely to have resulted in the scoop shaped detachment. The result of applying the two critical pieces of evidence (items 20 and 21 on Fig. 8) was thus quite dramatic, in that the slumping-related scenarios 1a and 1b rose to become the most probable scenarios (~0.4 each), overtaking the large fault scenario 2a, which dropped from a prior probability (after 19 iterations) of 0.45 to a posterior probability (after 21 iterations) of ~0.1 (Fig. 9).

Note that the probability aggregation process, being purely multiplicative, is unaffected by the order in which the pieces of evidence are considered. Whatever the order, the final posterior probabilities are the same.

The posterior probabilities were used to guide the choice of which scenarios to take forward into geocellular and reservoir simulation models. These were, in turn, used to help make development decisions about this field segment.

Discussion

Sensitivities

The approach used here for estimating the probabilities of the identified subsurface scenarios – by estimating an initial prior probability for each scenario, and then using sequential Bayesian updating against different types of evidence to generate improved posterior probabilities – is potentially sensitive to two probability estimations: (1) the initial prior probabilities, and (2) the numerical likelihood values assigned to the qualitative likelihood assessment (i.e. red, yellow, green in Fig. 8).

In the analysis presented so far, and shown in Fig. 9, the initial prior probabilities were estimated by the subsurface team using judgement, based on their own experience and knowledge of the field and region. Whilst steps were taken to try to minimize motivational and cognitive biases (e.g. including the views of independent external experts), assessing prior probability remains a problematic area (e.g., Baddeley et al., 2004; Curtis & Wood, 2004). To test the sensitivity to the priors, the sequential Bayesian updating was repeated but using different values for the priors. Figure 10 shows one such case where the scenarios were initially assumed to be equally probable.

Comparison of Figures 9 and 10 shows that for the first few iterations the probabilities are strongly influenced by the assumed prior probabilities. With subsequent iterations the initial prior becomes less of a control, and the probabilities become dominated by the cumulative effect of the Bayesian updates. This means that, by including many types of evidence, and thus many iterations of sequential Bayesian updating, the scenario probabilities become strongly aligned with the interpretations of the individual pieces of evidence. This is desirable, as this will be less susceptible to the possible systematic biases affecting the prior probability estimation.

Sensitivities were also run on different assignments of numerical values to the red, yellow or green assessments (Fig. 8). The results (not shown) indicate some effect on the absolute scenario probabilities, but minimal effect on the relative order of the scenario probabilities. Indeed, for all reasonable values, the top 5 most probable scenarios were the same. As long as the conversion from qualitative to numerical values is applied consistently, this should not be a dominant factor on the outcomes and the subsequent business decisions made as a result.

Identifying the full range of scenarios

Making good field development decisions, and managing the risks around the outcomes of such decisions, demands a good understanding of the relevant subsurface uncertainties. A common problem with subsurface modelling (described for example in Caers, 2011) is that the full range of uncertainties is not fully comprehended; consequently, they are not incorporated into reservoir uncertainty modelling, resulting in non-optimal decisions and surprise business outcomes. Anecdotal evidence points to two possible causes for this. Firstly, anchoring – where one particular subsurface scenario dominates, perhaps because it was the first one thought up, or because it reflects the interpretational experience of the individuals in the subsurface team (Bond et al., 2002). With anchoring, the team finds it difficult to imagine different scenarios because they are anchored to the existing one. A second potential cause is perhaps more of a motivational bias related to a sunk cost effect – where a complex reservoir model already exists and has had much investment of resources to build and calibrate it. *Within* the existing model it may be possible, or even easy, to investigate the impact of some uncertainties by running numerous cases where the uncertain input values are varied. However, it may be very difficult to model truly different subsurface scenarios, as this could demand building completely new models. Thus, the motivational bias, whether conscious or subconscious, is to restrict the imagination of alternative scenarios that cannot be represented by the existing model.

For good decision making and risk management, it is crucial that the subsurface "truth" lies within the imagined range of possibilities (Spetzler et al., 2016). If the "truth" lies outside the imagined range, there is a large scope for poor decisions and unpleasant surprise outcomes. Consequently, perhaps the most important part of the process outlined in this paper is the scenario identification step. It is critical that this looks wide enough to ensure that all subsurface possibilities are embraced. The benefit of the current process is that any number of scenarios can be imagined without inhibition, safe in the knowledge that: (a) the sequential Bayesian updating process will systematically narrow the wide initial range of scenarios (12 in the present case study) down to a smaller, more manageable number, before starting to build reservoir models; and (b) that this will be done in an objective manner, strongly linked to evidence from subsurface observations and measurements. This process thus helps to mitigate the aforementioned biases.

The impact of different data types

The way in which a particular type of data affects the understanding of the subsurface is readily gleaned from Figure 8. Data that are not relevant to a particular scenario (grey in Fig. 8) have no effect on the posterior probability of that scenario. Where this is repeated across many scenarios, that piece of evidence is less valuable, relative to other data types that have a stronger effect on the posterior probabilities. Perhaps counter-intuitively, though, observations that are consistent with a particular scenario also do not necessarily effect its posterior probability greatly; with a likelihood of 1 (green on Fig. 8) these are equivalent to irrelevant grey data. On the other hand, "red" evidence that cannot be explained by a particular scenario can have a great effect (reduction) on its posterior probability, and data that do this across many scenarios are of particular value in constraining uncertainty. Thus, the product of the likelihoods for each piece of evidence across all scenarios is an indication of how useful that evidence was in distinguishing between the scenarios. For the different types of evidence in Figure 8, the products of the likelihoods vary from $5x10^{-10}$ to 1 (Table 2), with the most useful (lowest likelihood product) being evidence derived from the reprocessed seismic data.

It would be very interesting and useful to gather such likelihood data over many studies in a wide range of situations to create a database of types of evidence and their potential usefulness for constraining different types of subsurface scenario. This would be of benefit for designing data acquisition programs, which could use such analogue information to determine what data would have the greatest chance of providing good evidence to constrain the subsurface uncertainties. This could provide a quantitative basis for estimates of data reliability that are needed for Value of Information studies (e.g. Coopersmith & Cunningham, 2002; Grose & Smalley, 2017).

Potential improvements to the process

Elicitation of priors. In the current implementation of the process, the prior probabilities of each scenario were generated by estimations from team members. No weighting was used, so each team member's estimate was given equal weight. Theoretically, the most correct way of doing this is to rank the individual experts in terms of their estimating reliability (e.g. Bordley, 1982) or the quality of the estimating heuristics they use (Baddeley et al., 2004), and weight their estimates accordingly (see Delfiner, 2008, for a practical example). However, the effectiveness of weighting depends on the ability of the overall process facilitator to assign objective reliable weighting to each contributor (Bolger and Rowe, 2015), which is exceedingly difficult in any situation, but more so in a subsurface team consisting of multiple disciplines, backgrounds and experiences. Without reliable weightings, weighted aggregations of probabilities are not significantly better than unweighted aggregations. As Bolger and Rowe (2015, p10) observed, "good things do not come to those who weight"! In any case, in the current process, the updating of prior probabilities with numerous steps of Bayesian updating renders the final posterior probabilities insensitive to the initial assumption of priors (compare Figures 9 and 10). We thus do not see benefit from expending resources in generating weightings for the prior probability estimates. The current method of using averaged individual estimates, rather than a group consensus, avoids many of the issues with estimating in groups (e.g. Delfiner, 2008) and is a good practical compromise.

Elicitation of likelihoods. For each piece of evidence that was considered, the team made a judgement concerning each scenario in turn: if the scenario was true, what would be the likelihood that the evidence would be observed. In the current case study, this was done in a facilitated group meeting. Such situations are potentially vulnerable to various group biases (e.g. Baddeley et al., 2004), though this was mitigated to some degree by using trained facilitators well acquainted with such potential biases. As discussed previously for elicitation of prior probabilities, a more rigorous approach would be to use likelihood estimates from individuals, and weight these according to their expertise. However, without actual data on the reliability of predictions from each individual, this would be subjective. Furthermore, the process would be highly complex, with the reliability of each team member having to be judged for all 21 individual types of data being considered, potentially hundreds of permutations. We currently feel this complexity would be counter-productive. Our use of the simple traffic-light system scoring system for likelihood estimation (Fig. 8) is not only extremely simple, but avoids many of the issues with quantitative probabilistic elicitations because the three categories (red, yellow, green) are sufficiently distinct that the appropriate category is usually fairly obvious. However, a future possibility would be to track the probability estimating record of individuals through multiple exercises, so that their estimates could be calibrated and weighted appropriately using unobtrusive software tools that did not interfere with the flow of the team workshop.

Quantification of likelihoods. In the current process, likelihoods are elicited as red, yellow or green categories, as described above, and these are translated offline to numerical values (0.1, 0.7 and 1.0 respectively). A future improvement could be to develop a more rigorous process for converting the categories to numbers, which could potentially incorporate different numerical scales for different types of data. The principles of fuzzy logic (Zadeh, 1978) may be helpful in this regard, though the mathematics would need to be performed in the background to avoid compromising the simple flow of the process for subsurface teams.

Aggregation of probabilities. In the approach used here, the likelihood estimated for each piece of evidence is incorporated simply by applying Equation 2 iteratively. The final posterior probabilities are then the products of the arrays of prior probabilities and likelihoods. While some authors have used broadly similar methods (e.g. Curtis and Wood, 2004), some have proposed more sophisticated methods to aggregate probabilities (e.g. Bordley, 1982; Allard et al., 2012). More sophisticated aggregation techniques are advantageous for two main reasons: (1) they allow weighting of inputs from different estimators based on their reliability; and (2) they can account for data redundancy related to the lack of independence between some of the pieces of evidence, and indeed between members of the team providing the estimates of prior probabilities and likelihoods.

We do not believe more sophisticated probability aggregation would have added significant value in the present case, but this is something that could be considered in the future. For the first advantage to materialize, the participants in probability estimation would need to be calibrated as to their prediction reliability as described above. For the second advantage, data redundancy between the participants would also need to have been assessed as to their individual areas and depths of expertise, something that is possible, though could be complicated by the confidentiality of personal data. Redundancy between different types of data should be a more tractable issue, requiring research to generate a table of dependencies between data types that could be built in to the selected probability aggregation algorithm.

With any future improvements to the process, it is crucial that any complex calculations are kept "behind the scenes" and do not impact the enthusiasm and focus of the subsurface team or the flow of the workshops used to implement the process. Unless the process is seen as simple, it will probably not be used at all.

Conclusions

Statistical techniques such as sequential Bayesian updating may seem intimidating for non-mathematicians, a possible reason why it has not been widely adopted by subsurface teams. This paper demonstrates a technique that applies Bayesian logic in a practical manner that is visual, rather than being overtly mathematical, and is thus easily engaged with by subsurface teams. The process frees teams to imagine a wide range of subsurface scenarios covering the full range of possible Earth realities, in the knowledge that these can subsequently be systematically reduced to a more manageable set of scenarios that are constrained objectively by evidence. After assigning prior probabilities to each scenario based on current knowledge and professional judgement, the sequential Bayesian updating method is easily applied in a team setting using a qualitative traffic-light approach to indicate likelihoods for the data (probability that the scenario could have generated the observed evidence). These can subsequently be converted to a quantitative scale by assigning numerical values to each traffic light colour. The likelihoods are multiplied by the prior probabilities and renormalized to sum to one across all of the scenarios to generate posterior probabilities. This process is iterated for many types of evidence (21 in the case study presented), at each iteration the posterior from the previous iteration becoming the prior for the next. The final posterior probabilities for each scenario are strongly constrained by the available evidence, making maximal use of all the available data, and being insensitive to the initial prior probability assessments. The product of the likelihoods for each piece of evidence is a measure of their usefulness, a useful input to value-of-information studies.

This process is proposed as a practical, consistent and objective process to prioritize subsurface scenarios to take forward for subsequent more detailed static and dynamic reservoir modelling.

Acknowledgments

BP and the anonymous case study field partners are thanked for permission to publish.

References

Allard, D., A. Comunian, and P. Renard, 2012, Probability aggregation methods in geoscience. Mathematical Geosciences, v. 44(5), p. 545-581.

Baddeley, M.C., A. Curtis, and R. Wood, 2004, An introduction to prior information derived from probabilistic judgements: elicitation of knowledge, cognitive bias and herding. In: A. Curtis and R. Wood (eds.), Geological Prior Information: Informing Science and Engineering: Geological Society Special Publication 239, p. 15-27.

Bolger, F., and G. Rowe, 2015, The aggregation of expert judgement: Do good things come to those who weight? Risk Analysis, v. 35, p. 5-11.

Bond, C.E., R.J. Lunn, Z.K. Shipton, and A.D. Lunn, 2002, What makes an expert effective at interpreting seismic images? Geology, v. 40, p. 75-78.

Bordley, R.F., 1982, A multiplicative formula for aggregating probability assessments. Management Science, v. 28, p. 1137-1148.

Bratvold, R.B., and S.H. Begg, 2010, Making Good Decisions. Society of Petroleum Engineers, 207pp.

Caers, J., 2011, Modelling Uncertainty in the Earth Sciences. Wiley, 229pp.

Coopersmith, E.M., and P.C. Cunningham, 2002, A practical approach to evaluating the Value of Information in the upstream petroleum industry. SPE paper 77582, presented at the SPE ATCE, San Antonio, Texas, 29 September-2 October 2002.

Curtis, A., and R. Wood, 2004, Optimal elicitation of probabilistic information from experts. In: A. Curtis and R. Wood (eds.), Geological Prior Information: Informing Science and Engineering: Geological Society Special Publication 239, p. 127-145.

Delfiner, P., 2008, Uncertainty in prospect evaluation: Lessons from the movie industry. AAPG Bulletin, v. 92, p. 1415-1429.

Go, J., I. Bortone, A.H. Muggeridge, P.C. Smalley, 2014, Predicting Vertical Flow Barriers Using Tracer Diffusion in Partially Saturated, Layered Porous Media. Transport in Porous Media, v. 105, p. 255-276

Grose, T., and P.C. Smalley, 2017, Risk-based surveillance planning: A practical Value-of-Information approach for data acquisition in producing fields: SPE Economics and Management, January 2017, p. 1-11.

Krueger, J.I., and D.C. Funder, 2004, Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking behaviour and cognition. The Behavioral and Brain Sciences, v. 27, p. 313–327.

Lee, P.M., 2012, Bayesian Statistics: An Introduction, 4th edition. Wiley, 486p.

Milkov, A.V., 2015, Risk tables for less biased and more consistent estimation of probability of geological success (PoS) for segments with conventional oil and gas prospective resources. Earth-Science Reviews, v. 150, p. 453-476.

Montibeller, G., and D. von Winterfeldt, 2015, Cognitive and Motivational Biases in Decision and Risk Analysis. Risk Analysis, v. 35, p. 1230-1251.

Smalley, P.C., and A.H. Muggeridge, 2010, Reservoir compartmentalization: Get it before it gets you. In: S.J. Jolley, Q.J. Fisher, R.B. Ainsworth, P.J. Vorlijk and S. Delisle (eds.), Reservoir Compartmentalization, Geological Society, London, Special Publication 347, p. 25-41.

SPE, 2016, Guidance for Decision Quality for Multicompany Upstream Projects. Society of Petroleum Engineers, Technical Report 181246, 36p.

Spetzler, C., H. Winter, and J. Meyer, 2016, Decision Quality: Value Creation from Better Business Decisions. John Wiley and Sons, Hoboken, NJ, 225p.

Van Wees, J.-D., H. Mijnlieff, J. Lutgert, J. Breunese, C. Bos, P. Rosenkranz, and F. Neele, 2008, A Bayesian belief network approach for assessing the impact of exploration prospect interdependency: An application to predict gas discoveries in the Netherlands. AAPG Bulletin, v. 92, p. 1315-1336.

Willigers, B.J.A., S.H. Begg, and R.B. Bratvold, 2013, Combining Geostatistics with Bayesian Updating to Continually Optimize Drilling Strategy in Shale Gas Plays. Society of Petroleum Engineers paper 164816, presented at the EAGE Annual Conference & Exhibition incorporating SPE Europe, London, UK, 10–13 June 2013.

Zadeh, L.A., 1978, Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, v. 1, p.3-28.

Authors

Craig Smalley has worked in petroleum geology for over 30 years in various technical and senior advisory roles. He developed and led BP's global subsurface uncertainty and risk management community of practice. He now consults in field appraisal and development, reservoir compartmentalization, risk management and decision making. He has visiting academic positions at Imperial College London and Oxford University.

Chris Walker received a PhD in Earth and Environmental Science from Columbia University and a MESc from Oxford University. He joined BP in 2006 and has worked a variety of Production, Development and Technology roles primarily assessing fault seal and interpretation uncertainty. He currently works as a Structural Geologist in the Western Hemisphere Exploration Technology Center.

Paul Belvedere has 28 years of industry experience as a Petroleum Geologist.   He spent his early career at ARCO and Vastar, working projects from Exploration to Production in the mid-continent region. Since joining BP, Paul has been primarily involved in the Gulf of Mexico, currently as a Development and Production geologist on deepwater assets, and serves part time as BP Production Geoscience Advisor for the Western Hemisphere. He holds a BS in Geology from Midwestern State University and MS in Geology from the University of Tulsa.

Tables

Group	Scenario	Prior probability
1. Slumps	1a. Slumps off bathymetric high	0.231
	1b. Undercut regressive slope failure	0.202
	1c. Thrusts caused by gravity slide	0.116
	1d. Local subsidence of unstable substrate	0.169
2. Faults	2a. Large but un-imaged faults	0.116
	2b. Smaller scale deformation	0.012
	2c. Tar-filled fractures	0.012
3. Sedimentologic and stratigraphic	3a. Zone of amalgamation	0.029
	3b. Compensation lobes with continuous shale drapes	0.029
	3c. Paleotopography	0.058
4. Diagenesis	4a. Cemented aquifer	0.012
	4b. Differential diagenesis	0.012

Table 1. Scenarios imagined to explain the observed reservoir compartmentalization, and their assigned prior probabilities.

Evidence	Product of likelihoods
New seismic data: scoop shaped detachment	5E-10
New seismic data: arcuate low reflectivity zone	2E-08
Pressure: inter-well baffling	2E-02
OWC	3E-02
Uniform stratigraphy	8E-02
Basin boundary	1E-01
Faults above reservoir	1E-01
Amplitude dim	1E-01
Shale correlation	1E-01
High curvature	2E-01
Previous faults missed	2E-01
Sub-regional isochore	2E-01
Local isopach map	2E-01
Uncharged clayey interval	2E-01
Tar presence	3E-01
Field analogues	3E-01
Analogue fault style	5E-01
Missing section	5E-01
Tar conduits	5E-01
Deformation in fast-track seismic data	1E+00
Pressure ramp; overpressured shales	1E+00

Table 2. Products of likelihoods across all 12 scenarios, sorted from lowest to highest. The evidence with the lowest values have the greatest constraints on the posterior probability of the scenarios.

Figure 1

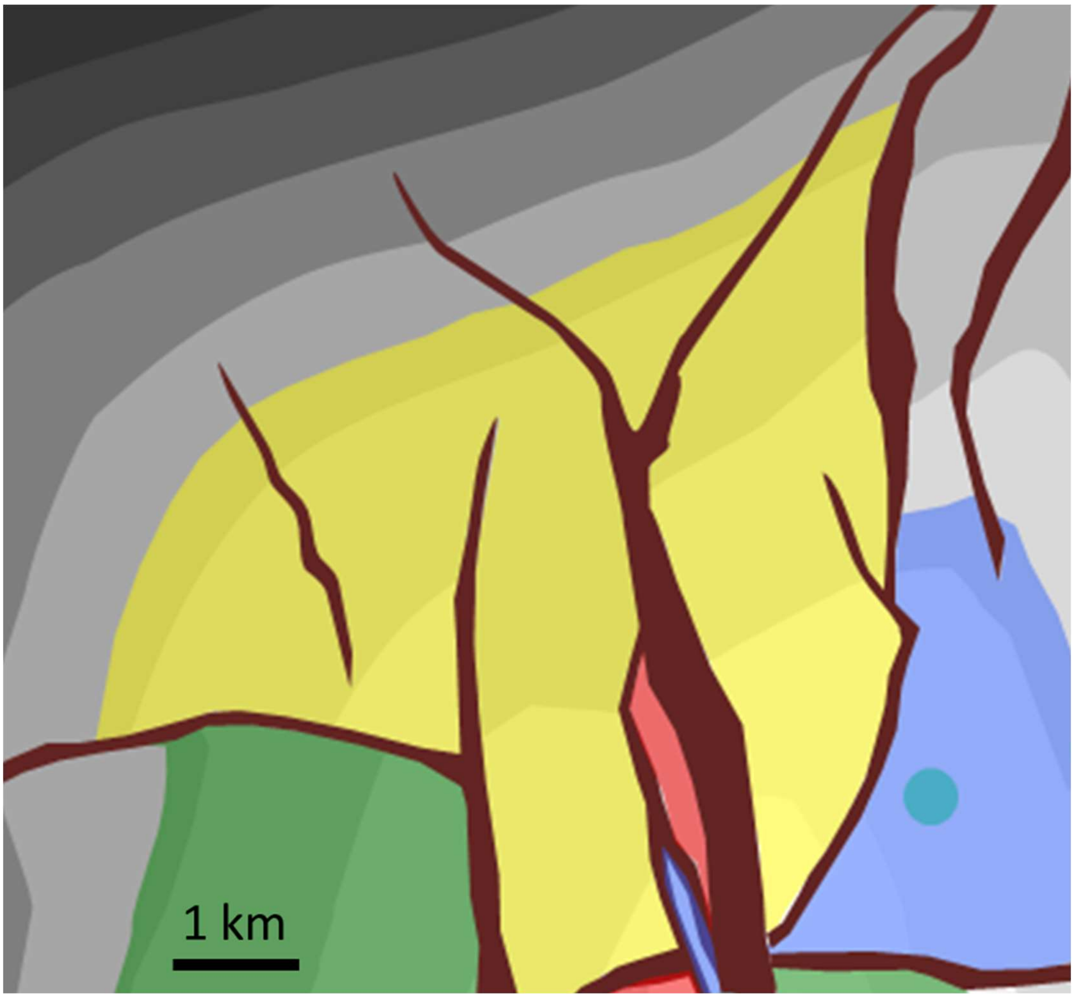

Figure 1. Reservoir structure map showing interpretation prior to appraisal drilling results. The map shows major seismically-visible faults (heave gaps shown as brown), with compartments coloured by pore fluid: green is known oil, blue is known water. In the area being appraised the colors follow a common risk segment map convention: yellow is moderate risk of encountering water, red is a high risk of encountering water. Grey denotes probable water below the sub-regional oil-water contact. The blue circle is a water penetration. The shade of colour relates to reservoir depth; darker shades are deeper. The contour interval is 500 ft (150 m). The scale bar is 1km (0.62 mile).

Figure 2

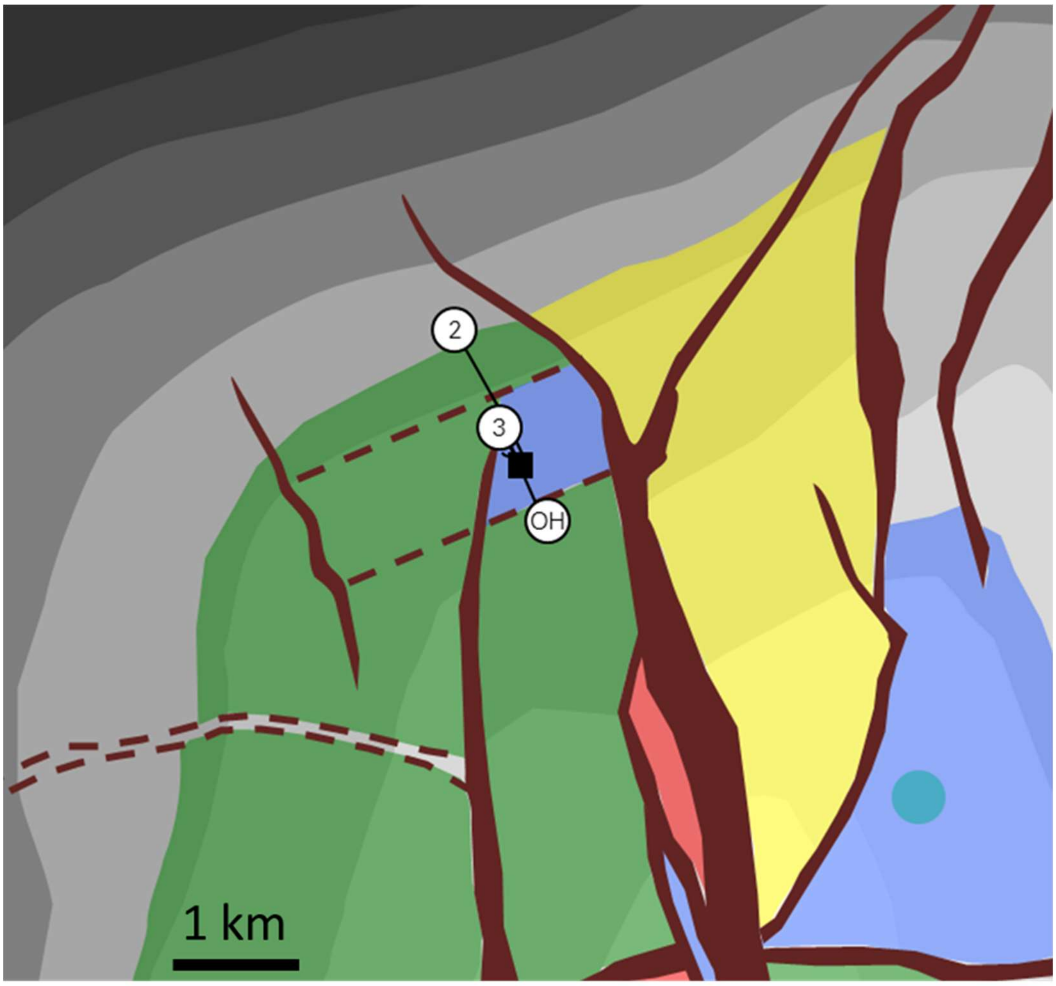

Figure 2. Reservoir structure map showing an interpretation after the appraisal drilling results, but before the current study was performed. Legend as for Figure 1. The appraisal penetrations are shown: black square is wellhead, black circles are bottomhole locations of the original hole (OH) and sidetracks 2 and 3, black lines are well paths. The contour interval is 500 ft (150 m). The scale bar is 1km (0.62 mile).

Figure 3

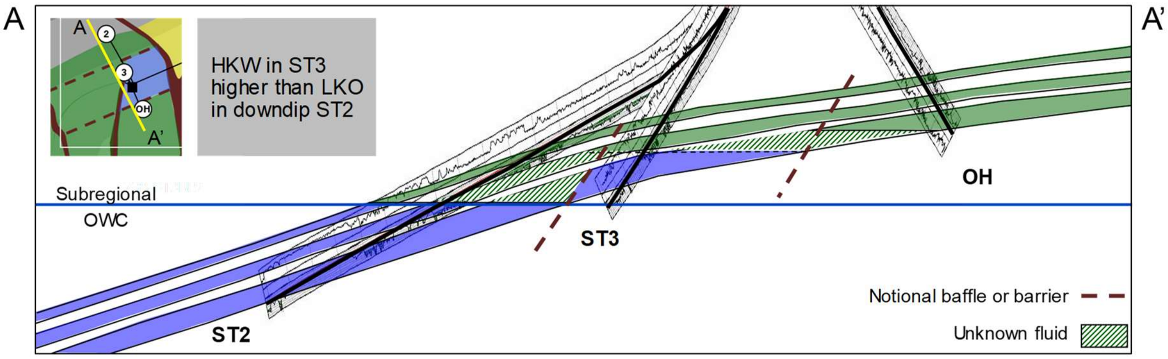

Figure 3. Cross section, roughly NNW-SSE, through the bottomhole locations of the three appraisal penetrations (OH = original hole, ST2 = sidetrack 2, ST3 = sidetrack 3), showing encountered fluid type (green is oil, blue is water, hatched green is unknown). HKW = highest known water; LKO = lowest known oil.

Figure 4

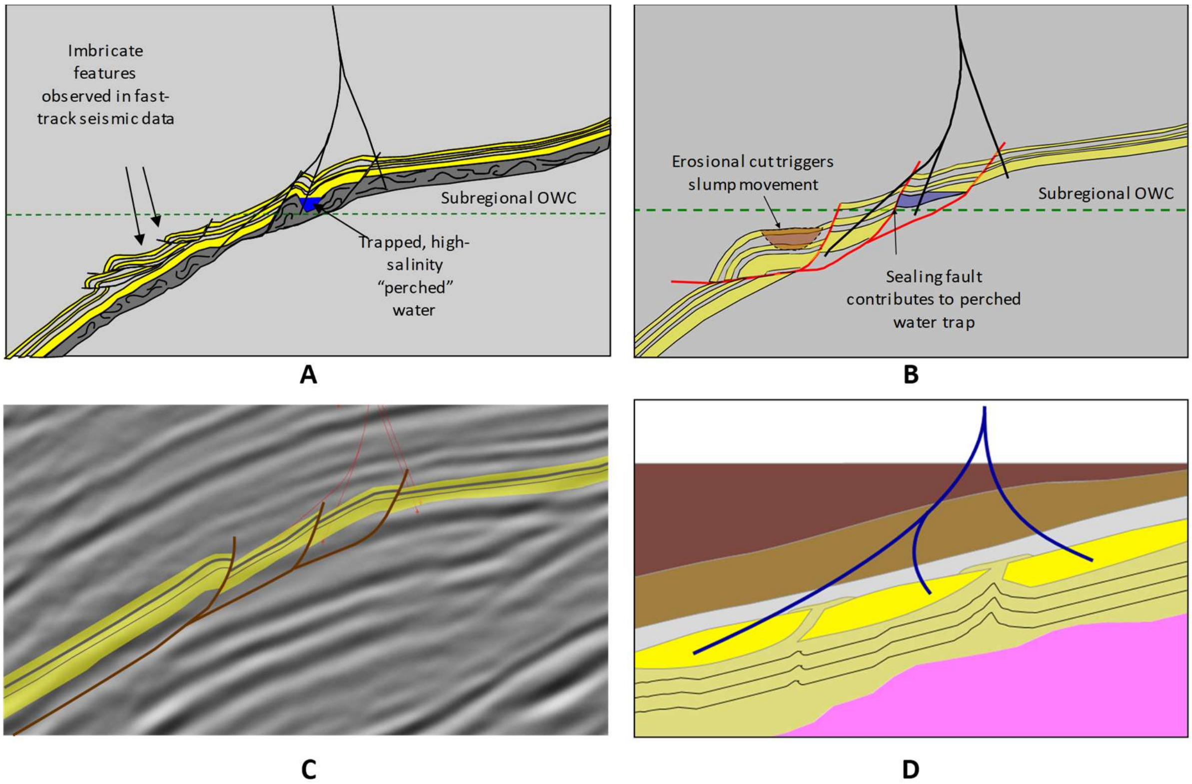

Figure 4. Cartoons illustrating the possible conceptual subsurface scenarios related to slumping. (A) Slumps off a bathymetric high; (B) Undercut regressive slope failure; (C) Thrusts caused by gravity slide; (D) Local subsidence of unstable substrate.

Figure 5

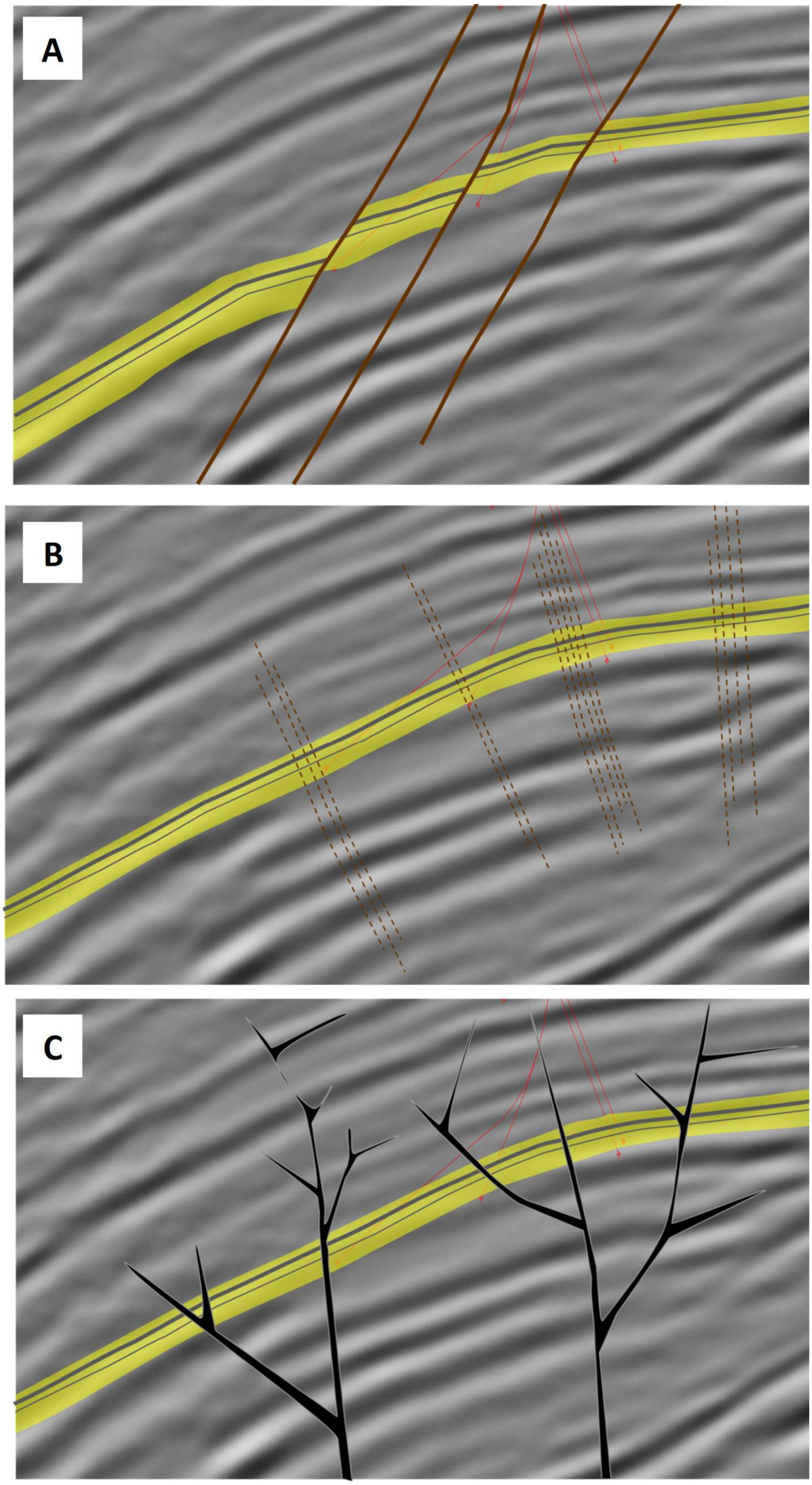

Figure 5. Cartoons illustrating the subsurface scenarios related to faulting. (A) Large but un-imaged faults; (B) Smaller scale deformation; (C) Tar-filled fractures. Well trajectories shown in red.

Figure 6

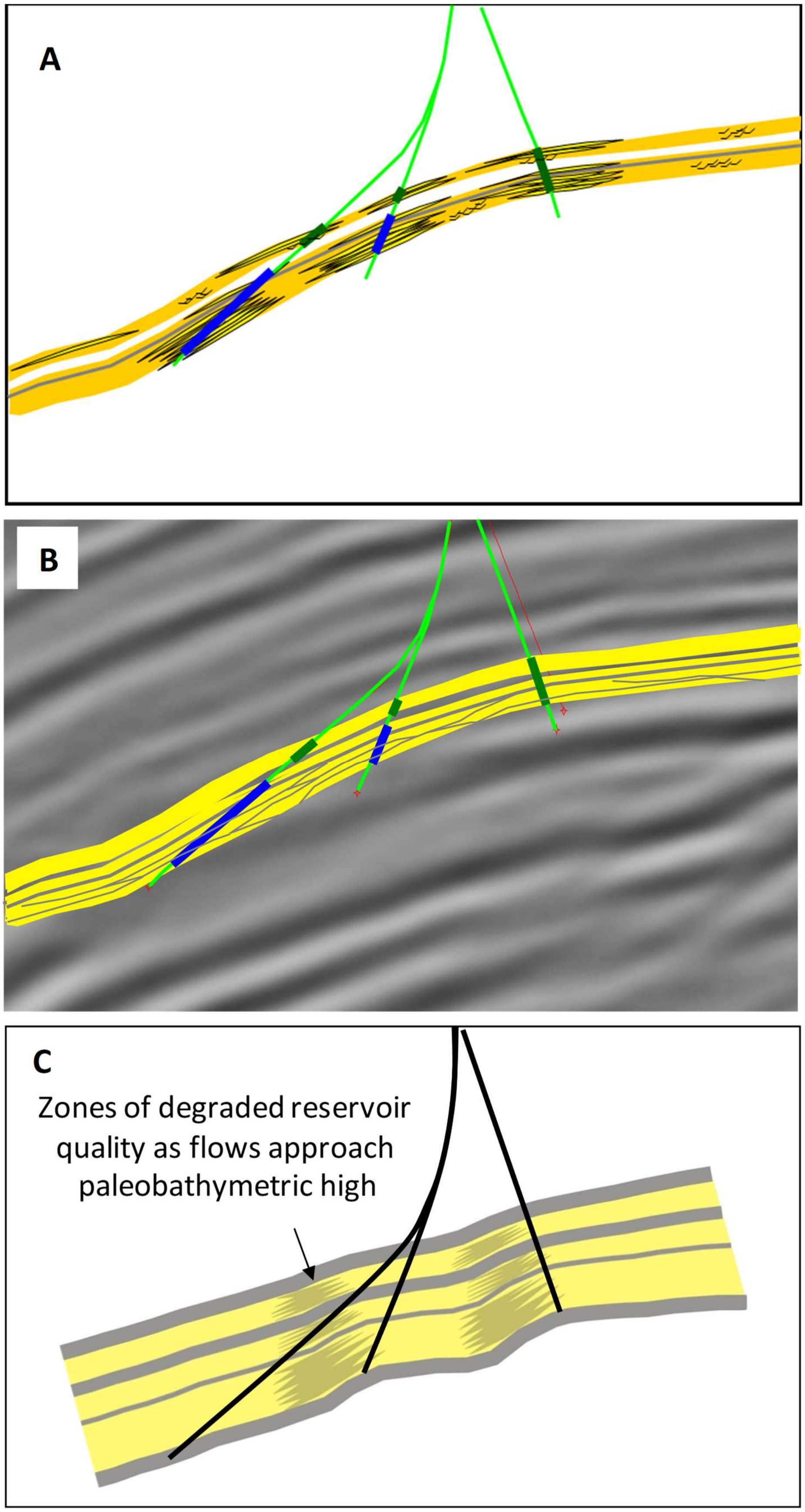

Figure 6. Cartoons of subsurface scenarios related to deposition. (A) Zones of amalgamation; (B) Compensation lobes with continuous shale drapes; (C) Paleotopography.

Figure 7

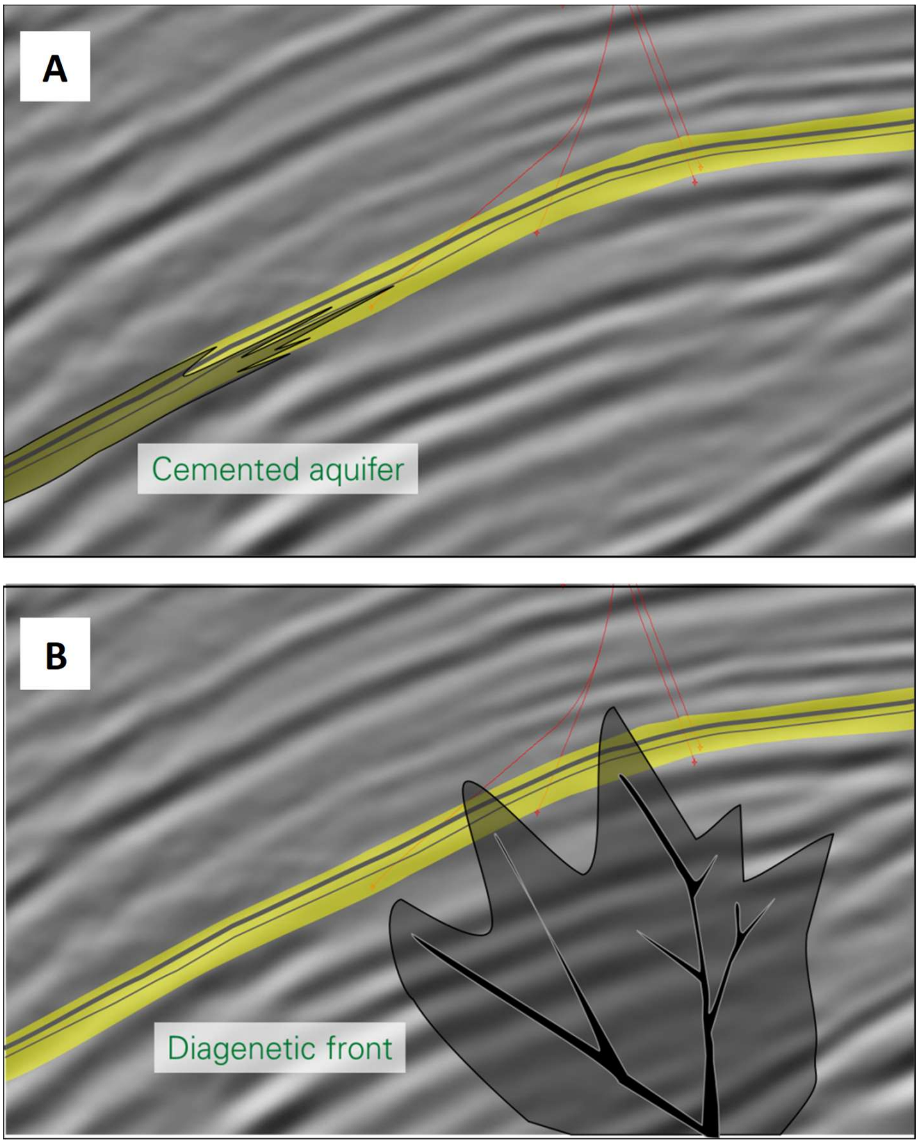

Figure 7. Cartoons of subsurface scenarios related to diagenesis. (A) Cemented aquifer; (B) Differential diagenesis.

Figure 8

	Scenario	1a	1b	1c	1d	2a	2b	2c	3a	3b	3c	4a	4b
	Evidence	Slump off bathymetric high	Regressive slope failure	Thrusts caused by gravity slide	Local subsidence	Large faults	Smaller scale deformation	Tar-filled fractures	Zone of amalgamation	Shale drapes	Paleotopography	Cemented aquifer	Differential diagenesis
1	Subregional isochores	1	1	0.7	0.7	1	1	1	0.7	0.7	1	1	1
2	Local isopach map	1	1	1	1	1	1	0.7	0.7	0.7	1	1	0.7
3	Deformation in fast-track seismic	1	1	1	1	1	1	1	1	1	1	1	1
4	Pressure: inter-well baffling	1	1	1	1	1	1	1	0.7	0.7	0.7	0.1	0.7
5	Pressure ramp; overpressured shales	1	1	1	1	1	1	1	1	1	1	1	1
6	Basin boundary	0.7	0.7	1	0.7	1	1	1	0.7	1	1	0.7	0.7
7	Uniform strat	0.7	0.7	0.7	0.7	1	1	1	0.7	1	0.7	1	0.7
8	Oil-water contact	1	1	1	1	1	1	1	0.7	0.7	1	0.1	0.7
9	High curvature	0.7	0.7	0.7	0.7	1	1	1	1	1	1	1	0.7
10	Analog fault style	1	1	1	1	1	1	0.7	1	1	1	1	0.7
11	Faults above reservoir	0.7	0.7	0.7	0.7	1	1	0.7	1	1	1	1	0.7
12	Missing section	1	1	1	1	1	1	0.7	1	1	1	1	0.7
13	Previous faults misssed	0.7	0.7	0.7	0.7	1	1	1	1	1	1	1	0.7
14	Tar conduits	1	1	1	1	1	0.7	1	1	1	1	1	0.7
15	Tar presence	1	1	1	1	0.7	0.7	1	1	1	1	1	0.7
16	Uncharged clayey interval	1	1	1	1	1	1	1	1	0.7	0.7	0.7	0.7
17	Amplitude dim	1	1	1	1	1	0.7	0.7	0.7	1	0.7	0.7	0.7
18	Shale correlation	0.7	0.7	0.7	0.7	1	0.7	1	0.7	1	1	1	1
19	Field analogues	1	1	1	0.7	1	1	1	1	0.7	1	0.7	1
20	New seismic data: scoop shaped detachment	0.7	1	0.7	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
21	New seismic data: arcuate low-reflectivity zone	1	0.7	0.7	0.7	0.1	0.1	0.1	0.1	0.1	0.7	0.1	0.1

Figure 8. Matrix showing the subsurface scenarios (columns) and pieces of evidence (rows). The colours relate to the traffic light scores initially assigned (green – evidence highly likely from scenario; yellow – evidence fairly likely; red – evidence unlikely; grey – evidence contains no information related to the scenario). The numbers are the scores that were later assigned to the traffic light colours. Items 20 and 21 were derived from reprocessed seismic data sometime later than the other evidence.

Figure 9

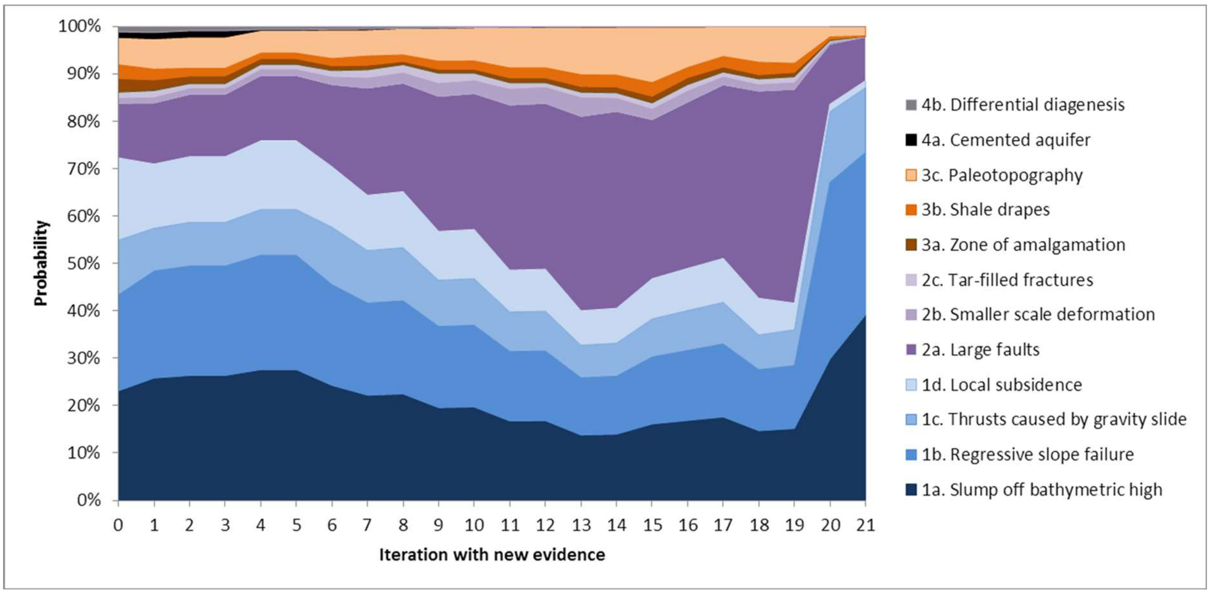

Figure 9. The probability of each subsurface scenario, showing how the view of their probability evolves when considering each individual piece of evidence, and applying the likelihoods shown in Figure 8, in the same order. The initial probabilities (at step 0) are the prior probabilities for each scenario assigned by the subsurface team.

Figure 10

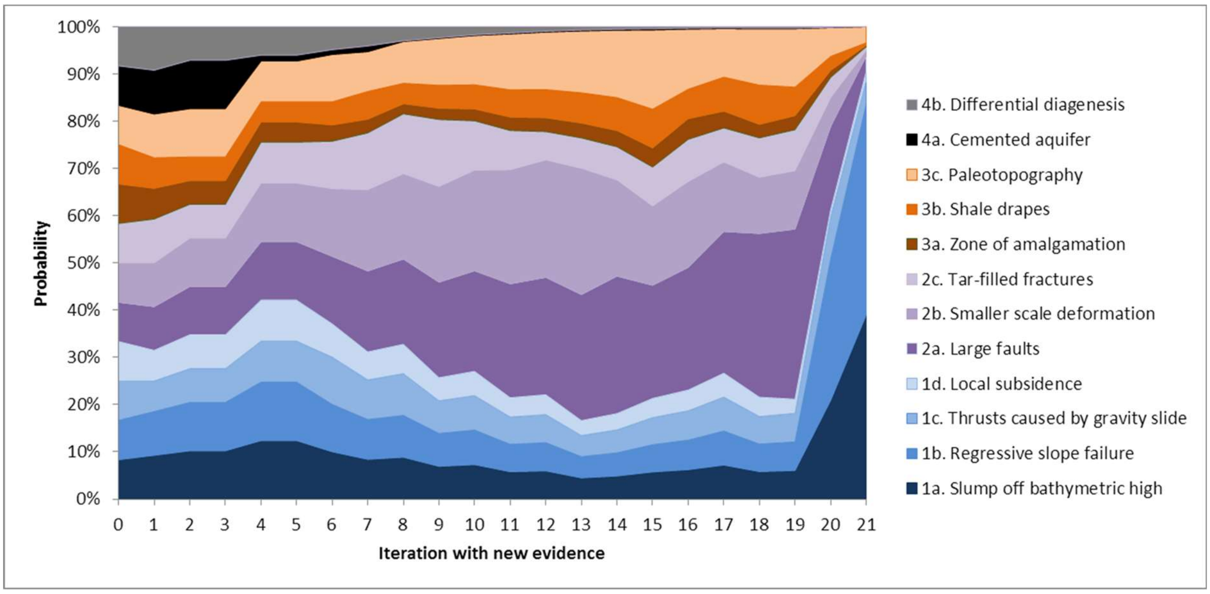

Figure 10. The probability of each subsurface scenario, showing how the view of their probability evolves when considering each individual piece of evidence, and applying the likelihoods shown in Figure 8. This differs from Figure 9 in that the initial probabilities (at step 0) are here assumed to be equal for each scenario (i.e. assuming no prior information). The results after 21 iterations compare closely with those of Figure 9.