Dealing with uncertainty on parameters elicited from a pool of experts for CCS risk assessment

Manceau JC., Rohmer J., Loschetter A., de Lary L., Le Guénan T. 
BRGM
3 Avenue Claude Guillemin - BP 36009
45060 Orléans Cedex 2 FRANCE

Hnottavange-Telleen K. 
GHG Underground
3 Avenue Claude Guillemin - BP 36009
TechPlace, 74 Orion St.
Brunswick ME 04011 USA

1 Introduction

CO₂ capture and storage (CCS) is recognized as a promising solution among others to tackle greenhouse gas emissions and mitigate climate change (IPCC, 2014). It consists of storing CO₂ permanently in appropriate deep geological formations among which saline aquifers are seen to provide the best world-wide geographical distribution and storage capacity. As for any industrial activity, the development of environmentally and healthy safe CCS must rely on robust risk assessment and management in the short as well as in the long term period to comply with regulatory frameworks, such as the CCS Directive in the European Union (Directive 2009/31/EC).

Risk assessment consists of supplying information on the different risks to allow an informed decision-making regarding the level of risk: potentially relevant risks are selected (risk identification), then their consequences on vulnerable elements and their likelihood are further studied (risk analysis) and finally the risks’ acceptability and the necessity for treatment are evaluated (risk evaluation). The integration of the uncertainty issue in risk assessment is crucial in the field of CCS (e.g. de Lary et al., 2015), because of limited experience feedback, of the inherently variable nature of geological objects, of our limited knowledge of these objects (unlike engineered objects for instance) and of the difficulties for understanding and representing the multiple physical phenomena occurring with the injection of CO₂ (fluid flow, geochemical, geomechanical, thermal effects, etc.). Two kinds of uncertainties can be distinguished. Aleatory uncertainty (also referred to as variability or randomness) is associated with the impossibility of predicting deterministically the evolution of a system due to its intrinsic complexity. On the contrary, epistemic uncertainty is not intrinsic to the system under study and can be qualified as “artificial”, because it stems from the incomplete/imprecise nature of available information, i.e. the limited knowledge of the physical environment or engineered system under study (e.g., Dubois, 2010).
In order to deal with uncertainties, current approaches generally use probabilistic methods for representing both aleatory and epistemic uncertainties on model parameters (e.g., Pawar et al., 2015). In situations of scarce/Incomplete data, expressing information in terms of statistical quantities may appear tedious, if not debatable. Besides, as outlined by Dubois and Prade (1994), the probability setting may be often too “rich” to be currently supplied by individuals: the identification of the probability distribution requires more information than what an expert is able to supply, which is often restricted to quantities or a prescribed mode. This raises the question of the relevancy of the fully-probabilistic setting in such a highly uncertain context. Among the few papers handling this issue in the domain of CCS, Bellienfant et al. (2009), de Lary et al. (2015), and Loschetter et al. (2015) demonstrated the interest of using alternative uncertain theories termed “extra-probabilistic” (possibility theory in the aforementioned papers) in early stages of a project to propagate both facets of uncertainties (aleatory and epistemic) separately up to the risk evaluation phase, allowing making decisions accordingly. Loschetter et al. (2015) gave evidence that the probabilistic framework may fail to produce reliable and conservative results in a context of high uncertainties when both aleatory and epistemic uncertainties are all mixed together in a single and unique setting, thus possibly leading to an underestimation of the actual risk level by providing a false impression of confidence in the results.

In the previous afore-mentioned studies, a kind of uncertainty is neglected, namely uncertainty on parameters elicited from a pool of experts, and not from a single expert. In a context of high uncertainty, when very little data is available, experts are commonly requested to provide their opinions on parameters. Not only might each expert express a certain degree of uncertainties on his own statements (this was the issue investigated in previous studies), but the set of information collected from the pool of experts introduces an additional level of uncertainties. It is indeed very unlikely that all experts agree on exactly the same data, especially regarding subsurface parameters which are by definition not precisely known. In some cases their opinions may only slightly differ (e.g. the most plausible value for a parameter is similar for different experts, and they only disagree on the level of uncertainties that taint the said value) while on other cases they may express incompatible opinions for a same parameter. Dealing with these different kinds of uncertainties remains a challenge for a reliable risk assessment.

To support the discussion related to this issue, we present a case study in the field of CCS risk analysis, on a CO₂ storage project in preliminary stage of development. The basic objective of the study is to estimate the extension of the injected CO₂ in a reservoir, based on the expert elicitation of the porosity of that reservoir. When developing a CO₂ storage project, it is indeed necessary to perform a feasibility study. One important purpose is to design the injection operations, which requires assessing notably the expected extent of the injected CO₂ plume. The challenge regarding this assessment is that an underestimation of this area might underestimate the potential risks of the storage (e.g.: risks of leakage through abandoned wells, risk of subsurface use conflicts), while an overestimation might compromise the project implementation or lead to excessive safety margin and characterization campaigns. Despite the challenges, during such a preliminary stage of a project, the available data may be poor, thus making expert elicitation necessary. Therefore the assessment should account for this constraint.

### 2 Presentation of the case study

The data analyzed for this study are a modified subset of data reported by Hnottavange-Telleen and Senel (2013) (the initial dataset has been adapted for this study for methodological purposes). The hypothetical site studied by Hnottavange-Telleen and Senel (2013) was selected for a feasibility study of CO₂ storage operations with the aim of storing 1 Mt of CO₂ during 30 years. The reservoir of interest is known to be a quartz-rich sandstone that overlies granitic basement. Dense information on the properties of this reservoir already exists from a specific area of the reservoir but the area contemplated for the storage is located 50 km away from this “dense information” area. The targeted area is undrilled, thus no local information on reservoir properties exists. Some information on this new area is nevertheless known: the top and thickness of the reservoir are estimated to be respectively 150-300 m shallower and 10-20 % thinner than those of the “dense information” area.

As indicated in the introductory section, the purpose of this preliminary study is to estimate the extension (in m) of the injected CO₂ plume to assess the feasibility of the storage.

Prior information (porosity measurements 50 km away) and general information on the site were provided to experts. 14 experts in the hydrogeological domain were asked to imagine relationships between prior data and the site characteristics to make estimations regarding the reservoir parameters. Experts were asked for the extreme values and the median of the (effective) porosity variable (see the answers in Table 1). The elicitation was done by email. As mentioned beforehand, the truth is unknown and the experts’ elicited data cannot be compared to real data.

It is important to note that experts can be subject to different types of bias during elicitation (see e.g. McBride et al., 2012 for a list of subjective biases commonly encountered in expert elicitation). The bias may be motivational, i.e. the expert has a stake in the issue that distorts consciously and unconsciously his judgment; or cognitive, i.e. the expert fails to process, aggregate, or integrate the available data and information (see e.g. Bonnano et al., 1990). The protocol of elicitation is therefore highly important as it is an essential means for handling those biases.

<table>
<thead>
<tr>
<th>Expert nb.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porosity min</td>
<td>15.2</td>
<td>15</td>
<td>15.6</td>
<td>14.2</td>
<td>13</td>
<td>13</td>
<td>12.7</td>
<td>12.7</td>
<td>12.3</td>
<td>9</td>
<td>12</td>
<td>13</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Porosity median</td>
<td>20.3</td>
<td>20</td>
<td>19.8</td>
<td>18.9</td>
<td>18</td>
<td>17.5</td>
<td>17</td>
<td>16.9</td>
<td>16.9</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Porosity max</td>
<td>25.2</td>
<td>25</td>
<td>23.9</td>
<td>23.6</td>
<td>22</td>
<td>22</td>
<td>21</td>
<td>21</td>
<td>21.5</td>
<td>23</td>
<td>20</td>
<td>20</td>
<td>18</td>
<td>17.5</td>
</tr>
</tbody>
</table>

### 3 Presentation of the method

In order to estimate the CO₂ storage extent from the elicited experts’ data, a classical multi-step approach has been followed. It consists of first representing the data given by the experts, then aggregating the information provided by the 14 experts and finally propagating the resulting uncertainty within a dedicated migration mathematical model. Within this overall method, we apply two different approaches to deal with uncertainties, the classical probabilistic approach and an approach based on other extra-probabilistic uncertainty theories.
We chose for our case study to use the belief functions theory (often called Dempster-Shafer theory), which allows an exact representation of the knowledge when it is expressed with percentiles as in our case study (see e.g. Baudrit and Dubois, 2006). More theoretically, considering the problem on the real line, a Dempster-Shafer structure consists in sets of real values associated with a given positive mass, with the sum of the different masses equaling 1 (Shafer, 1976; Ferson et al., 2003):

\[ m: 2^\mathbb{R} \rightarrow [0,1], \text{such that} \]

\[ \forall A \subseteq \mathbb{R}, 0 \leq m(A); \text{and} \sum_{A \subseteq \mathbb{R}} m(A) = 1 \]

The mass function \( m \) can be seen as representing the degree of belief that the actual value of the variable of interest falls in a given set, but nothing more precise (Destercke and Burger, 2012). Belief (upper confidence degree) and plausibility (lower confidence degree) measures can then be defined as such:

\[ \text{Bel}(A) = \sum_{E: A \subseteq E} m(E) \]
\[ \text{Pl}(A) = \sum_{E: E \cap A \neq \emptyset} m(E) \]

A way of representing Dempster-Shafer structures is through probability box ("p-box"), which can be defined as a pair of cumulative distributions circumscribing an unknown probability measure cumulative distribution (Ferson et al., 2003). From a given Dempster-Shafer structure, one can obtain an associated p-box, with respectively

\[ F(r) = \text{Bel}(]-\infty, r]) \text{ and } F(r) = \text{Pl}(]-\infty, r]) \]

as lower and upper bounds. Even though the transformation between a Dempster-Shafer structure and a p-box is not information preserving (e.g. Baudrit and Dubois, 2006), this way of representing appears as an interesting and useful tool in risk assessment domain when the question to be answered concerns a threshold exceedance (Destercke et al., 2008; Ferson et al., 2003).

4 Representation of the information

The classical protocol for representing the information provided by experts in the probabilistic setting is to find the probability distribution that respects the information and minimizes the information added by the choice of this distribution by e.g. the minimization of the entropy. In our case, the information on the minimum, median and maximum values of the porosity parameter is provided. In such a case, the distribution adding the less information is a uniform distribution both between \([m_{\text{max}}, m_{\text{min}}]\) and \([m_{\text{min}}, m_{\text{max}}]\), i.e. the Cumulative Distribution Function (CDF) is a linear interpolation between these three values (see e.g. Destercke and Chojnacki, 2008). The resulting CDFs for the 14 experts are provided on Figure 1.

Figure 1. Probabilistic representation of the porosity information given by the 14 experts (left: the 14 experts are represented on the figure, right: for a better visibility expert nb.1 information is plotted only)

Regarding the Dempster-Shafer theory, the knowledge (extreme values and median) given by the experts can be exactly represented by the mass function \( m \) defined by \( m([\text{min}, \text{median}]) = 0.5 \) and \( m([\text{median}, \text{max}]) = 0.5 \). The belief and plausibility functions of this Dempster-Shafer structure can be derived as well as the corresponding p-box. These p-boxes are provided on Figure 2.

Figure 2. P-boxes obtained from the Dempster-Shafer structures built from the information of the 14 experts (left: the 14 experts are represented on the figure, right: for a better visibility expert nb.1 information is plotted only)

5 Aggregation of the information

The information supplied by the experts needs to be aggregated in order to further reduce uncertainty and biases (Drescher et al., 2013). The aggregation approaches are commonly divided in two types:
- Behavioral approaches: these approaches allow to find a consensus among the experts, before any processing (e.g. Delphi Method);
- Mathematical approaches: the information provided by the experts is combined according to different mathematical methods.

In the case study presented in this paper, no behavioral approach has been implemented after the collection of the individual expert’s information; the second type of aggregation approach has therefore been followed. One important point regarding the mathematical approaches is the evaluation of the information provided by each expert. Sandri et al. (1995) specify the deficiencies to which the experts are subject when they estimate the values of given parameters:

- Miscalibration: The values given by the expert are inconsistent with real values; for instance, the experts may underestimate or overestimate the parameters;
- Underconfidence: The intervals given by the experts are too large to be informative (although they might contain the real values);
- Overconfidence: The intervals given by the experts are too narrow compared with the range of uncertainty existing on the real values.

To deal with such deficiencies during the fusion process, several ways exist. These deficiencies might be known by the analyst and, in that case, he might estimate coefficients to modify the experts’ information. Otherwise, the performance of each expert can be evaluated through tests for which the analyst knows the true value while the experts do not. In our case study, no such performance evaluation has been performed during the elicitation and, according to the analyst all the experts are a priori equally reliable.

### 5.1 Probabilistic setting

The classical and often considered as the best way of aggregating the experts’ opinion with a probabilistic representation is the arithmetic mean of the experts distributions (see e.g. Cooke, 1991).

\[
P_{\text{arith-mean}} = \sum_{i=1}^{K} w_i p_i \tag{4}
\]

where \( K \) is the total number of expert, and \( w_i \) is the weight attributed to each expert. In the case study, since no performance-based weighting procedure can be set-up before the information fusion, equal weights have been given to the experts. The results of the aggregation, in the probabilistic framework, of the 14 experts’ opinion on the porosity parameter are shown on Figure 3.

![Figure 3. Information aggregation in the probabilistic framework (solid black line)](image)

### 5.2 Extra-probabilistic setting

Dempster-Shafer theory is known, as other theories of uncertainties on which set-operations are possible, to offer some flexibility regarding aggregation of potentially conflicting information (Destercke et al., 2009). In particular, when no information exists on the experts’ reliability, the analysis of the conflicts among those experts can help performing the aggregation, if one assumes that an expert may be judged as more reliable when it agrees with the others (Destercke et al., 2013).

Three main categories of fusion operators exist for theories of uncertainties and can be applied according to the situation encountered:

- Conjunctive operators: these operators make the assumption that all the experts are reliable (meaning in this study that they all provide the true value) and act equivalently as a set-intersection. The uncertainty of the fusion result is therefore lower than the sum of the fusion inputs’ uncertainties. These operators are limited when the experts strongly disagree (they can even provide no results when they totally disagree). A classical conjunctive rule in Dempster-Shafer theory is the so-called Dempster’s combination rule (Dempster, 1967):

\[
m_{DB}(C) = \frac{1}{1-K} \sum_{A \cup B=C \neq \emptyset} m_1(A) \times m_2(B)
\]

with \( m_1 \) and \( m_2 \) two mass functions and

\[
K = \sum_{A \cup B=\emptyset} m_1(A) \times m_2(B). \tag{6}
\]

\( K \) is the mass associated with the conflict occurring from the combination rule application and can be seen as a measure of the conflict between the different experts. The Dempster’s rule, which re-normalizes the obtained structure with the mass associated with the conflict, can lead to counter intuitive results in case of strong conflict (famous counterexample of Zadeh; Zadeh, 1984). Therefore the rule has been modified for instance by Smets (Smets’ rule; Smets, 2007) and Yager (Yager’s rule; Yager, 1992) who assigned the mass associated with the conflict respectively to \( \emptyset \) and to the universal set.

- Disjunctive operators: these operators make the assumption that at least one expert is reliable and act equivalently as a set-union. The uncertainty of the fusion result is therefore higher than the fusion inputs uncertainty and these
operators may lead to uninformative results. A systematic disjunctive consensus rule for Dempster-Shafer structures can be expressed as follows (Dubois and Prade, 1992):

$$m_r(c) = \sum_{A \cap B = c} m_1(A) \times m_2(B)$$

with $$m_1$$ and $$m_2$$ two mass functions

- Trade-off operators: their results lie between the conjunctive and disjunctive operators’ ones. The basic idea is, when the experts are partially in conflict, to find a compromise between reliability and informativeness (i.e. degree of precision of the information) (Destercke, 2008). Two kinds of different trade-off operators have been developed: the adaptive ones that, according to an increasing level of conflict, evolves from a disjunctive-like behavior towards a conjunctive-like behavior, and the non-adaptive ones that do not depend on the context. Such rules have been developed for Dempster-Shafer structures, among which mixing (equivalent to the arithmetic mean for probability distribution) and ’discount+combine’ method (consisting of discounting the experts and then combining them with a rule) can be cited (see e.g. Sentz and Ferson, 2002).

In our case, we propose to use several types of operators in order to analyze the available information and what can be said about the different experts’ opinion (e.g. level of conflict). Two first operators have been implemented, namely the (normalized) Dempster’s rule and the disjunctive rule. The results in terms of probability boxes of the aggregated data sets provided by the experts are provided on Figure 4.

Figure 4. P-boxes obtained from the Dempster-Shafer structures aggregation with (normalized) Dempster’s rule (grey lines) and with the disjunctive rule (black lines)

The differences between both rules, which correspond to two extreme situations, are obvious. Interestingly, the measure of the conflict between the different experts is extremely high ($$K = 0.99$$) (note that the dataset considered in this study is different from that used in Hnottavange-Telleen and Senel (2013), where the conflict was less significant). According to Dubois and Prade (1992), Dempster’s rule can be safely applied when the mass associated with the conflict is nil (i.e. $$K = 0$$), which is far from being the case here. Therefore, the conjunctive rule, which assumes that all the experts are reliable (i.e. they all provide the true value), is not applicable for the case study. On the contrary, as it could have been suspected, the disjunctive rule provides a reliable result but quite uninformative. The resulting p-box is indeed relatively wide.

In this situation, it was of interest to provide a third nuanced representation of the uncertainty information provided by the pool of experts: the alternative approach (termed “trade-off rule”) should find a trade-off between

- providing a vision on the extent of all possible values given by all the experts for the porosity : this is a conservative vision, but hardly efficient for decision making since it is poorly informative.
- providing a vision on the consensus among the experts: this is a very informative vision, but also riskier than the first one, because some conflictual information are discarded.

Since no information was available about the experts’ reliability (no weighting or discounting procedure could be implemented), it has been decided to perform an adaptive trade-off aggregation, in order to increase the informativeness. We decided to rely on the recent Pichon et al. (2015) approach.

This approach consists in testing, during the aggregation process, several hypotheses ”$$r$$ experts out-of 14 are reliable (but without knowing which one, i.e. assuming that all the experts are truthful)”, with $$r \in \{1, \ldots , 14\}$$. The hypotheses ”1 expert out-of 14 is reliable” and ”14 experts out-of 14 (i.e. all the experts) are reliable” correspond respectively to the disjunctive and conjunctive rules, which have been shown, in the case study, to be respectively poorly informative, and not consistent. The purpose of this method is to find the highest $$r$$ providing a consistent answer, i.e. to find the most informative (or specific) but consistent hypothesis.

According to Pichon et al. (2015), the hypothesis ”$$r$$ experts out-of 14 are reliable” can be formulated as follows:

$$m_{r\text{-out-of-14}}(B) = \sum_{\gamma(B) = r} m_1(A_1) \times \prod_{i=2}^{14} m_1(A_i)$$

with

$$\gamma(A) = \bigcup_{\gamma(B) = r} \bigcap_{i=1}^{14} A_i$$

To quantify the consistency of each hypothesis, the mass associated with the conflict occurring from the combination rule application can be used (criterion 1). Pichon et al. (2015) propose to use another consistency indicator which is the maximum value of contour function (defined as the plausibility of singletons) of the aggregation results (criterion 2). Both indicators were used in the following of the work, and Figure 5 (left) shows the consistency of the 14 different hypotheses ”$$r$$ out-of-14 experts are reliable”.

The application of this approach therefore allows an analysis of the information provided by the 14 experts. According to those results, a consistent aggregation result can be obtained by assuming that 1 (equivalent to the disjunctive rule) to 5 experts out of
the 14 are reliable. The results of the aggregation procedure assuming these 5 hypotheses are shown on Figure 5 (right). The best compromise between informativeness (n.b. the informativeness increases when \( r \) increases) and a full consistency can be obtained for the hypothesis "5 experts out of the 14 are reliable", as shown by the associated p-box whose area is significantly reduced compared to that of the disjunctive rule.

![Figure 5](image_url)  
Figure 5. (left) Consistency of the 14 hypotheses "\( r \) out-of-14 experts are reliable". Criterion 1 and criterion 2 are computed. (right) P-boxes obtained from the Dempster-Shafer structures aggregation with the hypotheses "\( r \) out-of-14 experts are reliable", for \( r \) varying from 1 to 5 (\( r \) is indicated on the figure)

### 6 Propagating the resulting uncertainty with a two-phase flow migration model

Uncertainty propagation consists of evaluating the impact of the input parameters uncertainties on the model outputs (Karanki et al., 2009). In the aim of assessing the CO\(_2\) plume extension given the uncertainty existing in the porosity parameter, we rely on a deterministic model developed by Nordbotten et al. (2005). This model has been derived considering a radial flow around an injection well in a deep saline aquifer of constant depth, and constant fluid properties (viscosity and density). In our case, we consider that during the injection period the gravity forces can be neglected compared to the viscous forces due to the high injection rates, and therefore that the CO\(_2\) advancement front is driven by the CO\(_2\)/water mobility ratio. We then used the following formula for computing the maximum extension of the injected CO\(_2\) plume during the injection period (a sketch of the model is provided on Figure 6; see Nordbotten et al., 2005 for more information on the model):

\[
R_{\text{max}} = \frac{\mu_w M_{\text{inj}}}{\mu_C \rho_C \phi a m_i M_{\text{inj}}} \tag{10}
\]

Where \( \mu_w, \mu_C, \rho_C, \phi, M_{\text{inj}} \) and \( H \) are respectively the viscosity of the native brine and CO\(_2\), the CO\(_2\) density, the porosity, the injected mass of CO\(_2\), and the aquifer height. The depth and height of the aquifer are known (respectively ca. 2000 m and 25 m); the viscosity of the fluids at the reservoir conditions (200 bar, 75°C) can be estimated at \( 4.9 \times 10^{-5} \) Pa.s for CO\(_2\) and \( 3.8 \times 10^{-4} \) Pa.s for brine. Similarly the CO\(_2\) density value is estimated at 626 kg/m\(^3\).

![Figure 6](image_url)  
Figure 6. Sketch of the CO\(_2\) migration model within the geological reservoir used for computing the extension R\(_{\text{max}}\) (modified after Nordbotten et al., 2005)

When several input parameters are subjected to uncertainties different propagation methods can be implemented. Monte-Carlo method is classically applied to input parameters represented by probability distributions. Baudrit (2005) presents two different approaches adapted to the Dempster Shafer framework, namely the independent random sets method and the conservative random sets method, which differ regarding the hypotheses made on the dependence of the input parameters’ uncertainty. Note that possibilistic and probabilistic representations of input parameters can also be integrated to the process as they are both encompassed within Dempster Shafer theory. It is however important to note that this step can be significantly challenging, notably regarding the computational burden it may cause in case of numerous variables, complex models and/or complex uncertainty representations.

In the case study, the main uncertainty is on the porosity parameter that has been elicited beforehand. Given the simplicity of the model, the propagation procedure is therefore straightforward. The results of the uncertainty assessment are shown on Figure 7 in terms of the impact of the porosity uncertainty on the extension of the CO\(_2\) plume. The propagation of the probabilistic representation and of the Dempster-Shafer representations of the uncertainty are both represented, considering aggregation with the disjunctive rule and the combination made with the hypothesis "5 out-of the 14 experts are reliable".
The extension of the CO₂ plume is one element to be accounted for the decision regarding the feasibility of the storage project (other elements not assessed in this study are important as well such as the pressure-induced footprint for instance, see e.g. Bouc et al., 2009). An incorrect evaluation of this parameter might indeed either underestimate the level of risk induced by the presence of CO₂ or, on the contrary, artificially overestimate it, and consequently compromise the storage project ("artificial" impossibility of storing the expected quantity in a given perimeter, unessential characterization campaign and associated cost increase). It is therefore important to discuss the results obtained by stating the elements they provide for the decision making.

7 Discussion: how the extra-probabilistic approach can help for decision making

The extension of the CO₂ plume is an essential element to be accounted for the decision regarding the feasibility of the storage project. In a risk analysis perspective, we recommend to use those results together with the results of the disjunctive rule aggregation: the disjunctive rule aggregation provides a conservative view of the possible extension values according to the worst case scenario, whereas the probabilistic aggregation that was notably due to the choice of a specific probability distribution type). In a risk analysis perspective, we recommend to use those results together with the results of the disjunctive rule aggregation: the disjunctive rule aggregation provides a conservative view of the possible extension values according to the worst case scenario, whereas the probabilistic aggregation that was notably due to the choice of a specific probability distribution type.

7.1 Providing multiple views on the uncertainty

Interestingly, the results obtained from the different aggregation approaches are significantly different. Logically, the disjunctive rule, corresponding to a cautious behavior, leads to a large uncertainty in terms of CO₂ plume extension (large p-box, see Figure 7). The probability obtained after averaging the probability distributions derived from the experts’ data set also displays a strong uncertainty on the output parameters. Arithmetic mean can indeed be seen as a kind of equivalent of the disjunctive rule for probability distribution in the sense that it accounts for all the information given by the experts in the same way. However, the information is summed up in a single distribution, constrained by the choice of the distribution type to represent the information given by each expert. This single distribution may lead, if provided alone to the decision makers, to an artificially high level of confidence regarding the extension values to be considered as discussed below.

Let us consider for instance the P50 and P90 extension values (i.e. there is a 50 %-probability, respectively a 90%, that the extension is below the P50-value, respectively the P90-value) as these are often used for decision making in probabilistic risk assessment. After aggregation in the probabilistic framework, these values are fixed and amount to \( P50_{\text{prob}} = 677 \text{ m} \) and \( P90_{\text{prob}} = 773 \text{ m} \). Considering the aggregation made with the disjunctive rule in the Dempster-Shafer framework, we can derive intervals for P50 and P95 values: \( P50_{\text{disj}} = [552 \text{ m} ; 770 \text{ m}] \) and \( P90_{\text{disj}} = [571 \text{ m} ; 925 \text{ m}] \). However, as indicated above, since this rule corresponds to a cautious attitude, it can lead to significantly large uncertainties, which is the case here. It is indeed likely that providing the decision makers only with those P50 and P90 intervals values obtained from the disjunctive rule would not help much the decision.

The trade-off aggregation rule shows good promise by nuancing both the conservative but poorly informative disjunctive aggregation rule and the unique result provided by the probabilistic setting. It provides intervals on percentiles of reduced width, i.e. of reduced epistemic uncertainty: \( P50_{\text{trade-off}} = [605 \text{ m} ; 770 \text{ m}] \) and \( P90_{\text{trade-off}} = [620 \text{ m} ; 778 \text{ m}] \). However, this information alone should not be the only criterion to support decision making: by definition, it makes the basic assumption that the majority opinion is more reliable; therefore it diminishes the importance of some part of the experts’ information, which may lead to extreme values of extension (the reason of non-conservativeness is nevertheless different from that occurring with the probabilistic aggregation that was notably due to the choice of a specific probability distribution type). In a risk analysis perspective, we recommend to use those results together with the results of the disjunctive rule aggregation: the disjunctive rule aggregation provides a conservative view of the possible extension values according to the worst case scenario, whereas the trade-off rule provides a view of the most plausible values if we ignore some information in order to obtain more precise but still consistent (with limited conflict) results during aggregation.

7.2 Informing on the conflict among experts

In addition to the outcomes in terms of p-box obtained, the entire process of the trade-off aggregation within the Dempster-Shafer theory brings other interesting information to be considered by the decision-makers. In particular, it can provide a transparent and readable analysis of the conflict existing among the different experts, and on the contribution of the aggregation steps to the overall uncertainty of the variable of interest. First it allows finding with how many experts assumed to be reliable a totally consistent outcome can be obtained after aggregation. In our case, depending on the criteria for judging the consistency of the aggregated result, this number is between 5 and 7 out-of-14. Secondly, the difference between the uncertainty observed after the application of the disjunctive rule, and the uncertainty observed after the application of the trade-off rule can be seen as an indicator of the part of the conflict among experts in the global uncertainty of the variable of interest. This information may guide the choices to be made after having collected and analyzed the experts’ opinion in order to potentially reduce the imprecision on the porosity parameter. The aggregated representation of this parameter reflects mostly the uncertainty expressed by each of the experts; in such a case, there would be no real way, relying only on experts’ elicitation, to reduce that uncertainty (the uncertainty could be decreased by collecting additional information from other sources). However, when a high conflict situation exists, like in the present case study, the resulting uncertainty reflects significantly the differences in the experts’ opinion. Therefore, the decision makers may, before making a decision, decide to refine the elicitation step (e.g. ask to the experts to find a consensus, or alternatively require other experts’ to join the process or assess the experts’ performance) in order to validate/invalidate the apparent disagreement observed in the existing set of information.
To illustrate the interest of the approach for the presented case study, one can imagine a “synthetic situation” where the conflict between the 14 experts is different from that of the real dataset. This synthetic situation has been built such that all experts almost agree on the extreme and median values of porosity and such that the aggregation results obtained after applying the arithmetic mean to the probability distributions and the disjunctive rule on the Dempster-Shafer structures are similar to the results obtained with the real dataset. The synthetic dataset and results of these two fusion rules are shown on Table 2 and Figure 8. If one accounts only for the results of these two rules, the conclusion to be drawn by the decision makers would not differ much between the two situations. However, the application of the trade-off rule “r-out-of 14 experts are reliable” leads to different results, which is more in accordance with the dissimilarities of the two data-sets: the differences between the two situations in terms of Dempster-Shafer structures obtained after applying the trade-off rule (the hypothesis “5-out-of 14 experts are reliable” also lead to a consistent result for this new dataset) are shown on Figure 9. One can clearly see that even when applying the trade-off rule, the uncertainty remains high with the new data-set (the p-box is very similar to that obtained with the disjunctive rule). In such a case, the conflict level is low and the aggregation process does not add a significant amount of uncertainty, contrary to what is observed for the real dataset.

Table 2. Information on porosity provided by the 14 experts for the “synthetic situation”

<table>
<thead>
<tr>
<th>Expert nb.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porosity min</td>
<td>10.4</td>
<td>10.1</td>
<td>9.7</td>
<td>10</td>
<td>9.3</td>
<td>10.6</td>
<td>9.7</td>
<td>9.9</td>
<td>10.3</td>
<td>10.7</td>
<td>9.9</td>
<td>10</td>
<td>10.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Porosity median</td>
<td>18.5</td>
<td>17.7</td>
<td>18</td>
<td>17.3</td>
<td>16.6</td>
<td>16.7</td>
<td>17.7</td>
<td>17.5</td>
<td>16.7</td>
<td>17.2</td>
<td>18.5</td>
<td>16.6</td>
<td>16.7</td>
<td>17.1</td>
</tr>
<tr>
<td>Porosity max</td>
<td>25</td>
<td>24.1</td>
<td>25.9</td>
<td>25.9</td>
<td>25.9</td>
<td>24.7</td>
<td>24</td>
<td>24.9</td>
<td>25.9</td>
<td>25.1</td>
<td>26</td>
<td>24.6</td>
<td>24.5</td>
<td>25.4</td>
</tr>
</tbody>
</table>

Figure 8. Representation of the experts’ information for the “synthetic situation” (left) Probabilistic representation of the porosity information given by the 14 experts. (right) P-boxes obtained from the Dempster-Shafer structures built from the information of the 14 experts

Figure 9. Results for the initial dataset (solid line) and for the synthetic dataset (dashed line) of the aggregation-propagation process with different aggregation methods: (top) the arithmetic mean in the probabilistic framework (middle) the disjunctive rule in the Dempster-Shafer theory (bottom) the trade-off rule with the hypothesis “5 out-of 14 experts are reliable”
8 Conclusion

A case study has been presented with the objective of illustrating the importance of integrating conflictual information from a pool of experts within an uncertainty analysis performed for risk assessment purposes. Handling potentially conflicting information coming from experts when other data at disposal are scarce or inexistent is an important issue for subsurface operations, especially for projects in the early stages of development. The problem tackled in this study concerns the CO₂ geological storage technology and more specifically the estimation, on a specific storage project, of the extension of the injected CO₂ in a reservoir, based on the expert elicitation of the porosity of that reservoir. For that purpose, information on the minimum, median and maximum values of the porosity parameter have been collected from 14 different experts.

The paper presents a practical framework to deal with uncertainties for the CO₂ extension risk assessment, especially to integrate the information coming from a pool of experts. A classical multi-step method has been followed consisting of first representing the data given by the experts, then aggregating the information provided by the 14 experts and finally propagating the resulting uncertainty within a dedicated CO₂-migration mathematical model. Within this method, two different approaches have been considered, the classical probabilistic approach and an approach based on the belief functions theory (often called Dempster-Shafer theory).

A literature review was made on the different steps in order to select appropriate tools for the assessment. In total the extension assessment with probability representations results in a unique probability distribution; this distribution is very precise but, if provided alone to the decision makers, it may lead to an artificially high level of confidence regarding the extension values to be considered. On the contrary, the assessment with Dempster-Shafer representations using the disjunctive rule for aggregation (classical when the level of conflict is high) provides more conservative results but might not be of great help for decision makers given the significant range of plausible plume extensions in which it results. Moreover, these two assessments do not practically convey the large conflict existing among the experts, which contains important information to be disseminated to the decision-makers. A recent trade-off rule, adapted to that case study, has therefore been set-up to overcome these limitations. This rule consists in testing the hypothesis “r experts out of 14 are reliable” with r decreasing from 14 to 1 and finding which one of these hypotheses allows the best compromise between consistency of the information to be aggregated and informativeness of the resulting fusion. This rule allows deriving a range of extension values for CO₂ plume which might be judged as more plausible (more specific than the disjunctive rule) and assessing the proportion of the uncertainty that is due to the disagreement among the experts. This proportion can then be used for setting-up a strategy to decrease this kind of uncertainty before making the decision (e.g. modification of the elicitation protocol, addition of new experts, assessment of experts’ performance, or the addition of new sources of information such as data collection).

The tools used for treating this case study have been chosen since they were adapted to that specific situation of high conflict level. Therefore, for other situations, other tools or approaches might be more relevant. In addition, for most of the aggregation rules developed for Dempster-Shafer theory, the rules considered in this study assume independent sources, which is a questionable assumption regarding expert’s opinion, as experts are likely to share parts of their expertise or knowledge (Ling and Rudd, 1989). Some works exist on accounting for source dependence/information distinctness in fusion rules (Denoeux, 2008) but it is still considered an open issue (Dubois et al., 2016).

Finally, even though an effort has been put into the explanation of which kinds of results can be transmitted to the decision makers, applying new methods such as those presented in that study remains an important challenge regarding the needs of transmitting operational, understandable and communication-facilitating outcomes. Efforts like the recent studies by Pedroni et al. (2013) or Loschetter et al. (2015) aiming at comparing the results obtained by the standard probabilistic (Bayesian or frequentist) approach and the ones derived from the use of new uncertainty theories, should also be intensified.

Acknowledgements

The authors thank the BRGM-funded project MEDIUM for the financial support of the present work.

References


Dubois, D., Liu, W., Ma J., Prade, H., 2016. The basic principles of uncertain information fusion. An organised review of merging rules in different representation frameworks. Information Fusion 32, 12–39


