Using the Choquet integral for screening geological CO_2 storage sites

Yingqi Zhang[1*]

1. Lawrence Berkeley National Laboratory, Earth Sciences Division, 1 Cyclotron Road, MS 90-1116, Berkeley, CA, USA

* Contact author
Earth Sciences Division
Lawrence Berkeley National Laboratory
University of California
1 Cyclotron Road, MS 90-1116
Berkeley, California 94720-8126
Tel.: 510-495-2983
Fax.: 510-486-5686
Email: yqzhang@lbl.gov

Abstract

For geological carbon dioxide storage site selection, it is desirable to reduce the number of candidate sites through a screening process before detailed site characterization is performed. Screening generally involves defining a number of criteria which then need to be evaluated for each site. The importance of each criterion to the final evaluation will generally be different. Weights reflecting the relative importance of these criteria can be provided by experts. To evaluate a site, each criterion must be evaluated and scored, and then aggregated, taking into account the importance of the criteria. We propose the use of the Choquet integral for aggregating the scores. The Choquet integral considers the interactions among criteria, i.e., whether they are independent, complementary to each other, or partially repetitive. We also evaluate the Shapley index, which demonstrates how the importance of a given piece of information may change if it is considered by itself or together with other available information. An illustrative example demonstrates how the Choquet integral properly accounts for the presence of redundancy in two site evaluation criteria, making the screening process more defensible than the standard weighted-average approach.

Keywords: Choquet integral, risk assessment, geologic carbon sequestration, geological CO_2 storage, site screening

1. INTRODUCTION

A number of options have been identified for CO_2 storage in geological media: utilization in enhanced oil recovery (EOR) operations, replacement of methane in coal beds, and storage in depleted oil and gas reservoirs, deep saline aquifers, and in salt caverns.[1] Before a detailed storage site characterization is performed, it is desirable to first screen the candidate sites. The certification framework (CF),[2] which examines the effectiveness of carbon trapping in geological formations, provides a simple and transparent way to evaluate the risks of CO_2 leakage from the storage formation. In that study, faults, fractures and wells were considered the only potential leakage pathways. The consequences of CO_2 escaping the storage formation are the impacts CO_2 has on certain compartments (defined as entities that are vulnerable to CO_2 leakage) above the storage formation. Five compartments were considered: ECA – emission credits and atmosphere; HS – health and safety; NSE – near-surface environment; USDW – underground source of drinking water; and HMR – hydrocarbon and mineral resources. When candidate sites are compared based on leakage risks, the risks to all compartments need to be aggregated in a reasonable and defensible manner.

The most common way to aggregate multiple criteria for decision making is to use the weighted arithmetic average, where each weight is given by an expert to represent the importance of a particular criterion. Simplicity and ease of use are the main advantages of this approach. A key drawback of the approach is the assumption that all the criteria are independent of each other, i.e., the measures (here a “measure” refers to a quantitative criterion to evaluate a site) are additive. For measures that are non-additive, i.e., if there are correlations and interactions among the criteria, information that is redundant or mutually

exclusive is not properly accounted for by taking the weighted average, which may lead to a bias in the overall evaluation of a potential storage site.

The five compartments considered in the CF are not independent of each other. For example, if CO_2 leaking to the atmosphere has been identified as a risk, this implies that the other compartments are most likely also at risk. Moreover, risks associated with CO_2 leakage to the near-surface environment are generally strongly correlated to health and safety risks. Due to such correlations (interactions) among the risks that refer to individual compartments, it is inappropriate to use the weighted arithmetic mean to aggregate these risks. Instead, we propose to use the Choquet integral to account for the risk correlation among the compartments. The Choquet integral has been introduced by *Murofushi and Sugeno*[3] and others (e.g., *Grabisch*[4]) as an aggregation tool of non-additive measures (also referred to as capacities). In the next section, we will discuss the general approach and demonstrate its application to evaluate CO_2 leakage risks.

The purpose of this communication is to describe the Choquet integral as an approach for including correlated criteria into initial site screening. The applicability of the approach and the potential impact of including these correlations on the screening decision are demonstrated using a simple, synthetic example. Readers interested in a comprehensive discussion of risk evaluations at CO_2 storage sites are referred to, for example, Stenhouse *et al.*[5] and Maul *et al.*[6].

2. THE CHOQUET INTEGRAL

We explain the Choquet integral in the context of risk evaluation. Because the evaluation criteria are countable, only the discrete Choquet integral is relevant and will be discussed. We consider a finite universal set N, which can be thought of as the index set of

a set of criteria or attributes, where n is the cardinality of the set, i.e., $n = |N|$. A fuzzy measure ν on N is defined as a monotone set function ν: $2^N \rightarrow [0,1]$ to indicate the importance of a criterion or a subset of criteria (also referred to as a coalition). The set function satisfies (1) ν(Ø)=0 (where Ø represents an empty set), ν(N)=1, and (2) for any subsets S and T, $S \subseteq T \Rightarrow v(S) \leq v(T)$ (monotonicity). The monotonicity property states that having more elements in a coalition does not reduce the importance of a coalition.

Consider the case of two elements $i, j \in N$. If $v(i, j) = v(i) + v(j)$, the two elements contain independent information; if $v(i, j) < v(i) + v(j)$, the two elements are substitutive, i.e., simply adding the scores of both elements leads to the inclusion of redundant information; finally, if $v(i, j) > v(i) + v(j)$, the two elements are complementary, i.e., having both elements enhances the overall information content of the fuzzy measure v. If elements in a set of criteria are either substitutive or complementary, the Choquet integral of $x \in \Re^n$, defined in Equation (1), should be used:

$$C_v(x) := \sum x_{(i)}[v(A_{(i)}) - v(A_{(i+1)})] \tag{1}$$

where $x_{(i)}$ is the score for criterion i, and x is permutated on N such that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$, $A_{(i)} = \{(i),\ldots,(n)\}$, and $A_{(n+1)} = \varnothing$.

For demonstration purposes, we consider three compartments (n = 3) for the evaluation of CO_2 leakage risks: HMR, USDW and an HSE (health, safety, and environment) compartment which includes both NSE and HS compartments, a slight deviation from the CF but convenient for this explanation. Therefore, $N = \{R_{HMR}, R_{USDW}, R_{HSE}\}$, where R_{HMR} represents leakage risk into the HMR compartment with the index of the criterion being 1, and so on. The discussion with an expert (Oldenburg CM, 2010, pers.

comm.) provides the following insights: Considered individually, the opinion emerges that HMR is a relatively unimportant criterion, because risk to HMR does not directly harm human beings, other animals, and the environment. In addition, there is a possibility that the resources in the HMR compartment (e.g., oil, natural gas, potash) will never be needed or produced. Both USDW and HSE are much more important criteria because USDW is generally protected by law (e.g., the Safe Drinking Water Act in the U.S.), and CO_2 leakage into the HSE compartment may cause hazards to the safety of humans, other animals, and plants near the ground surface. As an element of a subset, risk in HMR provides relatively independent information. In fact, most hydrocarbon resources vulnerable to CO_2 injection are not in the shallow subsurface but rather are in the deep subsurface. Consequently, knowing that no HMR is at risk does not provide us much, if any, information about risks to other compartments. So even if a risk to HMR has been identified, CO_2 may not be able to reach the USDW and HSE compartments due to additional containment features and trapping mechanisms (e.g., Oldenburg[7]). As a result, the importance of having information on both HMR and USDW or HMR and HSE is equal to the sum of the importance of having individual information. However, risks to USDW and HSE are correlated. If we know there is a risk to HSE, we also know that there is some risk to USDW, although exceptions exist, i.e. the correlation coefficient is less than 1. In other words, these two pieces of information are partially overlapping. Therefore, the importance of having risk knowledge on both compartments is less than the sum of the importance of the individual information, but more than the importance of each individual piece of information, i.e., the correlation between the two is less than 1.

The coefficients ν, provided by stakeholders are defined in Table I. Different stakeholders may have different opinions which need to be accounted for in a multi-criteria analysis. The Choquet integral approach can also be used to aggregate the coefficients given by multiple stakeholders. For simplicity, we consider the case in which the stakeholders agree on the relative importance of the criteria. In this communication, for demonstration purposes, we take as an example the hypothetical geological CO_2 storage site in a Texas Gulf Coast saline formation from *Oldenburg et al.*[2] The scores for leakage risk to each compartment (x_i), obtained based on descriptions by *Oldenburg et al.*,[2] are listed in Table II. They are numbered (1= R_{USDW}, 2 = R_{HSE} , 3 = R_{HMR}) to satisfy $x_{(1)}<x_{(2)}<x_{(3)}$.

The Choquet integral is thus evaluated as follows:

$$\begin{aligned} C_v &= 7\times(v(1,2,3)-v(2,3))+8\times(v(2,3)-v(3))+10\times v(3) \\ &= 7\times(1-0.8)+8\times(0.8-0.1)+10\times 0.1 = 8.0 \end{aligned} \tag{2}$$

If a weighted average had been used, the final score for this site would have been 7×0.5+8×0.7+10×0.1=10.1. The Choquet integral is lower than the weighted mean, properly reflecting the removal of redundant information in all three criteria.

To be able to express ν in a unique way, i.e., $v(S)=\sum_{T\subseteq S} a(T),\ S\subseteq N$, the Möbius transform of a fuzzy measure ν is used:[8]

$$a(S)=\sum_{T\subseteq S}(-1)^{s-t}v(T) \tag{3}$$

The Choquet integral is then written as ($\wedge$ denotes the minimum operation):[8]

$$C_v(x)=\sum_{T\subseteq N} a(T)\wedge x_i \tag{4}$$

Therefore, $a(i)= \nu(i)$, and $a(i,j)= \nu(i,j)- \nu(i)- \nu(j)$. The sign of $a(i,j)$ indicates if criteria i and j contain information that is independent, redundant, or complementary. For our example, both $a(1,2)$ and $a(1,3)$ are zero, indicating that the risk to HMR is independent information. The negative value for $a(2,3)$ indicates that the risks to USDW and HSE are partially overlapping. Finally, a positive value for $a(i,j)$ would indicate that criteria i and j are complementary to each other. Evaluating information on both risks increases the overall information content beyond the sum of the individual information value. Only if all a(S) were zero, will it be the case that the criteria are mutually independent, and a weighted mean can be used to aggregate the scores. In most cases, however, the Choquet integral should be used to account for complementary or redundant information, which is a result of interactions among individual criteria.

3. THE SHAPLEY INDEX

The Shapley index represents the overall importance of a criterion $i \in N$ in a decision problem. It is typically used to interpret the Choquet integral. It is determined by all the $\nu(T)$ that contain i, i.e., $i \in T$:

$$\varphi(v,i) := \sum_{T \subseteq N \setminus i} \frac{(n-|T|-1)!|T|!}{n!}\left[v(T \cup i) - v(T)\right] \tag{4}$$

$N\setminus i$ refers to the set N without i. Recall that $\nu(T)$ are fuzzy measures that represent the importance of subset T. The term $v(T \cup i)$ represents the importance of having both subset T and i. Therefore the term $v(T \cup i) - v(T)$ represents the contribution (or added value) of element i in coalition T. The Shapley index calculates the average contribution of element i in all coalitions. A basic property for the Shapley index is that $\sum_{i=1}^{n} \varphi(v,i) = 1$.

For our example, we have $\varphi(v,1) = 0.1$, $\varphi(v,2) = 0.35$, and $\varphi(v,3) = 0.55$. The average contribution of Criterion 2 or 3 in all coalitions is less than the contribution of the criterion in a coalition with itself. Therefore, the Shapley indices of these two criteria are less than the coefficients ν.

Assume Criteria 1 and 3 are complementary and ν(1,3)=0.85 (all the other ν(s) stay the same as in Table 1). As a result, the Shapley indices for each criterion are $\varphi(v,1) = 0.11$, $\varphi(v,2) = 0.33$, and $\varphi(v,3) = 0.56$. The overall importance of Criteria 1 and 3 are increased by the same amount (i.e., $\varphi(v,1)$ increases from 0.1 to 0.11 and $\varphi(v,3)$ increases from 0.55 to 0.56), because they enhance the information content of each other when both are available. Correspondingly, the importance of Criterion 2 is reduced from 0.35 to 0.33.

In this example, only three criteria are included for illustration purposes. In theory, one could include many more criteria. However, if $n > 3$, the evaluation of the Choquet integral becomes difficult because the experts need to determine a large number (2^n) of coefficients. In this case, *Grabisch*[9] proposed to approximate ν by a *k*-order fuzzy measure. If $k = 1$, the approximation is merely the weighted mean for additive measures. From a practical point of view, a 2-order approximation seems to be appropriate as it considers interactions between pairs of criteria and at the same time is not overly complicated.[9] The Choquet integral with respect to a 2-order fuzzy measure is then written as:

$$C_v(x) = \sum_{i \in N} a(i) x_i + \sum_{\{i,j\} \subseteq N} a(i,j)(x_i \wedge x_j) \tag{5}$$

4. CONCLUSION

In this Short Communication, we have proposed the use of the Choquet integral to evaluate and compare the risks of geological CO_2 storage sites. The approach was demonstrated for a previously documented risk evaluation. Typically, criteria used in the risk assessment are not independent. To appropriately consider correlations among criteria, the Choquet integral should be used for aggregating the scores given by experts or from formal risk assessments. The Shapley index for each criterion can be evaluated to provide the overall importance of a criterion in all coalitions. The proposed approach has practical value at the initial stage for CO_2 storage site selection, when very limited information is available.

Acknowledgment

The author wishes to thank Curtis M. Oldenburg (LBNL) for serving as an expert for the illustrative example, for his useful discussion on CO_2 leakage risks to compartments, and for the review of the manuscript. Suggestions provided by Stefan Finsterle (LBNL) are greatly appreciated. This work was supported in part by the CO_2 Capture Project (CCP) of the Joint Industry Program (JIP), and by Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

REFERENCES

1. Bachu, S, Sequestration of CO_2 in geological media: criteria and approach for site selection in response to climate change. *Energy Conversion and Management*, 41, pp 953-970, (2000).

2. Oldenburg, CM, SL Bryant, and JP Nicot, Certification framework based on effective trapping for geologic carbon sequestration, International Journal of Greenhouse Gas Control, 3, pp 444-457, (2009).
3. Murofushi, T, and M Sugeno, An interpretation of fuzzy measures and the Choquet intergral as an intergral with respect to a fuzzy measure, *Fuzzy Sets and Systems*, 29, pp 201-227, (1989).
4. Grabisch, M, A graphical interpretation of the Choquet integral. *IEEE Transactions on Fuzzy Systems*, 8(5), pp627-631, (2000).
5. Stenhouse, MJ, J Galeb, and W Zhou, Current status of risk assessment and regulatory frameworks for geological CO_2 storage, *Energy Procedia*, DOI:10.1016/j.egypro.2009.02.007. (2009).
6. Maul, P, D Savage, S Benbow, R Walke and RR Bruin, Development of a FEP Database for the Geological Storage of Carbon Dioxide. In: E.S. Rubin, D.W. Keith and C.F. Gilboy, Editors, *Greenhouse Gas Control Technologies Vol. I*, Elsevier Ltd, pp. 701–709, (2005).
7. Oldenburg, CM, Screening and ranking framework for geologic CO_2 storage site selection on the basis of health, safety, and environmental risk, *Environmental Geology*, 54, pp1687–1694, (2008).
8. Marichal, JL, An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria, *IEEE Transactions on Fuzzy Systems*, 8(6), pp800-807, (2000).
9. Grabisch, M, k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92, pp167-189, (1997).

TABLES

$\nu(\emptyset)$	$\nu(1)$	$\nu(2)$	$\nu(3)$	$\nu(1,2)$	$\nu(1,3)$	$\nu(2,3)$	$\nu(1,2,3)$
0	0.5	0.7	0.1	0.9	0.6	0.8	1.0

Table 1. Coefficients ν of each subset evaluation criterion

Score	1 = R_{USDW}	2= R_{HSE}	3=R_{HMR}
x	7	8	10

Table 2. Individual scores for the risks of each compartment. Scores are on the scale of 1 to 10, where a low risk is assigned a high score.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.