Using Bayes Theorem to Quantify and Reduce Uncertainties when Monitoring Varying Marine Environments for Indications of a Leak
Alendal, Guttorm; Blackford, Jeremy; Chen, Baixin; Avlesen, Helge; Omar, Abdirahman

Published in:
Energy Procedia

DOI:
10.1016/j.egypro.2017.03.1492

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Heriot-Watt University Research Portal

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
13th International Conference on Greenhouse Gas Control Technologies, GHGT-13, 14-18 November 2016, Lausanne, Switzerland

Using Bayes Theorem to Quantify and Reduce Uncertainties when Monitoring Varying Marine Environments for Indications of a Leak

Guttorm Alendal\textsuperscript{a,}*, Jeremy Blackford\textsuperscript{b}, Baixin Chen\textsuperscript{c}, Helge Avlesen\textsuperscript{d}, Abdirahman Omar\textsuperscript{d}

\textsuperscript{a}Department of Mathematics, University of Bergen, Bergen, Norway  
\textsuperscript{b}Plymouth Marine Laboratory, Plymouth, UK  
\textsuperscript{c}Heriot-Watt University, Edinburgh, UK  
\textsuperscript{d}Uni Research, Bergen, Norway

Abstract

Monitoring the marine environment for leaks from geological storage projects is a challenge due to the variability of the environment and the extent of the area that migrating CO\textsubscript{2} might seep through the seafloor. Due to the environmental risk associated leaks should not be allowed to continue undetected. There is also a cost issue since marine operations are expensive, so false alarms should be avoided. The main question is then: how large a deviation in the monitoring data should cause mobilization of confirmation and localization procedures? Here Baye’s theorem and Bayesian decision theory is suggested as a tool for quantifying certainties and to implement costs for false positives (false alarms) and false negatives (undetected leaks) in the decision procedure. The procedure is exemplified using modeled natural CO\textsubscript{2} content variability and the predicted CO\textsubscript{2} signal from a simulated leak.

Keywords: Monitoring design, leak detection, Bayesian decision making.

1. Introduction

Geological CO\textsubscript{2} storage project is by regulations such as the London Convention, OSPAR and EU directives, required to have an adequate monitoring program. With proper selection and operational procedures CO\textsubscript{2} geological storage projects will be designed not to leak and a number of different trapping mechanisms will keep the injected CO\textsubscript{2}, being buoyant, inside the intended formation [1]. The injection well is believed to be the most probable leakage pathway but transport of the CO\textsubscript{2} within the formation might cause other pathways to the surface to become possible, or the CO\textsubscript{2} might create new pathways [2], possibly far away from the injection well.

Hence, even if geological monitoring of the reservoir, complex and overburden will be the primary monitoring strategy, there is a need and requirement for a surface monitoring program with the objectives

*Corresponding author.

Email address: Guttorm.Alendal@uib.no()
to 1) maximize assurance of storage integrity, 2) assure that a leak will likely be detected, 3) continue to build an accurate baseline to capture trends and natural variability, and 4) to prevent unjustified accusations of adverse effects from the storage project [3].

For offshore storage projects such a monitoring program can be costly, and the marine environment is hostile for instrumentation. It is therefore suggested that the monitoring program has three levels of modus operandi; 1) anomaly detection modus, 2) confirmation and location modus, and 3) seep quantification modus. Some suggest a fourth step; impact assessment [4].

The focus here is the detection phase, in which the monitoring program looks for anomalies in the environment. A map of probable leak locations, preferably quantifying the internal relative probability between the different sites will govern where it will be most important to search for leaks. This can only be achieved through a thorough site characterization of the overburden.

Equally important will be an understanding of how a leak can be recognized. Probabilistic footprint predictions of a seep have to be achieved through modeling CO$_2$ entering the water column which may materialise in the dissolved phase, or as individual bubbles, bubble trains, or bubble plumes if the leakage flux is high enough [5]. The dynamics of these regimes are different, with the plume dynamics being the most challenging to model [5, 6, 7, 8]. Detection of bubbles can be made from sonars [9, 10], another indication of a leak might be environmental impact, caused by elevated CO$_2$ concentration the vicinity of the source [11] possibly as materializing as new occurrences of bacterial mats [12].

Apart from approaches relying on a thorough understanding of processes, such as the vadose zone gas monitoring approach suggested in Romanak et al. [13], a proper environmental baseline is required in order to detect changes in the environment caused by a leak from the storage complex. Such statistical baseline of important environmental parameters will include currents, natural gas seeps and biogeochemical parameters. Historical data are important in combination with new data collected during site characterization. Long time series will capture natural variability, such as seasonal changes and long-term trends. In particularly it will be important to capture the expected acidification caused by increase of CO$_2$ concentration [14].

Here signals of elevated CO$_2$ concentration away from the seep location is used to illustrate the use of Bayes theorem to decide whether a measurement of CO$_2$ increases our belief that a leak is in progress and quantify our certainty if the decision is that there is no leak occurring. The seep footprints are mainly governed by the varying current conditions, both spatially and temporally, such as the tidal signal or local topography [15, 16, 17].

The leak scenarios used here are discussed in previous publications [18, 19]. The scenarios were simulated in the near zone by the HWU bubble plume model [7, 8] and on a larger scale by an 800m-grid resolution North Sea setup of the three-dimensional terrain-following Bergen Ocean Model (BOM) [19].

Previously these model results have been used to find optimal locations for chemical sensors [20, 21]. It was shown that placing the sensors successively at the location of highest probability is not necessarily the best option; one sensor might detect seeps at several potential leak locations. A threshold for detection, based on a stoichiometric approach [22], was used and an excess concentration above this level immediately concluded that a leak was present.

However, given the cost of mobilizing the resources need to confirm and locate a leak, it will be preferable to have a treatment of any data stream from a monitoring program to quantify with what certainty the alarm of an ongoing leak is based. The main question remains; what level of certainty will be required in order for the monitoring program to sound the alarm? This study argues that Baye’s Theorem and Bayesian decision theory offer that opportunity and it is exemplified using the same model data as in the aforementioned studies.

2. Bayesian Decision Theory

If our belief or probability that a leak is ongoing, the prior $p(L)$, the objective is to obtain an updated belief, the posterior $p(L|x)$, after taking a measurement, $x$. Bayes theorem reads [23]:

$$p(L|x) = \frac{p(x|L)p(L)}{p(x|L)p(L) + p(x|\neg L)(1-p(L))} \tag{1}$$
\[ p(\neg L|x) = \frac{p(x|\neg L)(1 - p(L))}{p(x|L)p(L) + p(x|\neg L)(1 - p(L))} \]  

where the probability of measuring \( x \) if a leak is present is \( p(x|L) \) and similar \( p(x|\neg L) \) is the probability to measure \( x \) in the natural environment, i.e. no leak on going. It is assumed that \( p(L) + p(\neg L) = 1 \).

The environmental variability will have to be accounted for in the \( p(x|\neg L) \) distribution and will have to be achieved through a thorough analysis of baseline statistics. The probability \( p(x|L) \), i.e. likelihood measuring \( x \) in the presence of a leak will have to be based on predictions from models [19] and possibly in-situ experiments [24].

After measuring \( x \) we can either decide that there is a leak or remain assured that there is none, with an estimate on our uncertainty. Subsequent measurements update our belief. At what probability of a leak being present should the alarm be raised? The four possible outcomes when taking a decision, \( (\alpha_1, \alpha_2) = (L, \neg L) \), when the true nature \( (\omega_1, \omega_2) = (L, \neg L) \) are illustrated in Tab. 1. The false positive situations will be false alarms that will mobilize unnecessary resources for location of a nonexistent seep, while the false negatives results in undetected seeps that might cause environmental risks.

<table>
<thead>
<tr>
<th>decision'nature</th>
<th>L</th>
<th>\neg L</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>true</td>
<td>false positive</td>
</tr>
<tr>
<td>( \neg L )</td>
<td>false negative</td>
<td>true</td>
</tr>
</tbody>
</table>

Table 1. The four different outcomes of taken a decision with respect to the real conditions. The two situations in which a leak is correctly detected or ruled out will both be true conclusions. False positives, i.e. deciding that leak is ongoing when it is not, or false negatives, when leaks go on undetected represent wrong decisions and should be avoided.

In general, let \( \lambda(\alpha_i|\omega_j) \) be the cost, or loss, involved in deciding \( \alpha_i \) while the true nature is \( \omega_j \). The risk of deciding \( \alpha_i \) given the measurement \( x \) is now

\[
R(\alpha_i|x) = \sum_{j=1}^{n} \lambda(\alpha_i|\omega_j)p(\omega_j|x) 
\]

and the decision rule is to select the \( \alpha_i \) that gives lowest risk.

In the "leak-no leak" classification scheme addressed here this translates to

\[
R(L|x) = \lambda_{11} p(L|x) + \lambda_{12} p(\neg L|x) 
\]

\[
R(\neg L|x) = \lambda_{21} p(L|x) + \lambda_{22} p(\neg L|x) 
\]

and the least risk is chosen. This can be translated into decide that a leak is present if likelihood ration exceeds a threshold:

\[
\frac{p(x|L)}{p(x|\neg L)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{p(L)}{p(\neg L)}.
\]

The cost parameters, \( \lambda_{ij} \), allows to balance the need to detect a leak with the cost involved with false alarms. The threshold for mobilizing the confirmation and localization procedures can hence be made dependent on the cost involved.

3. An example using \( \text{CO}_2 \) concentration baseline and signal.

To illustrate the use of the theorem \( \text{CO}_2 \) concentration distributions have been created based on limited sets of model results. Time series from the Norths Sea model evaluation set up from Plymouth Marine Laboratory [25] has been used to fit a nonparametric distribution function using standard MatLab routines. These are as continuous curves in Figs. 1 and 2, the distribution based on all data, i.e. a yearly distribution, is in grey colour while the blue curves represent the respective monthly distributions.
Notice the two maximums in concentration present for November (month 11) indicating that the environment has two modes then. Also notice the elevated tails for smaller concentrations for summer and autumn. There is too few data, from one single realization, to draw any conclusions from these distributions. But it illustrates that there will be seasonal dependency in the environmental statistics. These differences will most likely influence the ability to detect leaks.

To simulate signals from a leak, the time series from Ali et al. [19] has been used, indicated by the stapled curves, for the leak location in Figs. 1 and the adjacent grid cell just north of the location in Fig. 2. Since these simulations represent excess CO₂ content the resulting distributions have been convolved with the respective baseline distributions causing the baseline features to recognized in the leak signal.

The shift in distribution between the baseline (continuous line) and the leak situation (stapled line) toward higher concentration is what assists in detecting a leak. As expected this shift is highest close to the source in Fig. 1 compared to some distance away Fig. 2.

To simulate streams of measurement data a series of time series has been produced by randomly pulling a starting point in the respective time series used to find the distribution functions. For each of these time series the time to detect using Eq. 6 is calculated, as presented as box plots in Fig. 3. Notice the different range along the y-axis since the right hand figure represents a measurement taken further away from the source.

Not surprisingly more measurements will be needed to detect a leak further away from the measurement location, and there are more outliers present. For both locations it seems like May (month 5) is a good month to detect a leak, the median time to detection is low and the variability is also small. It seem that the baseline distribution levels to zero for higher concentrations and that the tail for low concentrations is limited, combined with a reasonable shift when adding the leak signal. In both locations the use of annual mean results in the need for more measurements to detect the leak.

4. Discussion

The example illustrates the importance of capturing the tails of the distributions, i.e. incorporate rare events, in the baseline statistics. It is shown that the monitoring program will benefit from resolving seasonal
variations, and it might identify the best time of year to supplement any fixed locations with cruises and campaigns.

However, the example used here should be used with care. The data sets used to illustrate the method are all based on model results and in reality they must be supported by in-situ measurements, especially for the baseline acquisition. In addition the ensemble of model realizations should be much higher in a realistic set up.

Acknowledgements

This work has received funding from the Research Council of Norway, through the CLIMIT program (BayMoDe project no 254711) and the SUCCESS centre for CO₂ storage (193825/S60), and the European Union’s Horizon 2020 research and innovation program under grant agreement No. 654462 (STEMM-CCS).
References