

## Optimal sensors placement for detecting CO<sub>2</sub> discharges from unknown locations on the seafloor



Anna Oleynik<sup>a,\*</sup>, Maribel I. García-Ibáñez<sup>b,c</sup>, Nello Blaser<sup>a</sup>, Abdirahman Omar<sup>b</sup>, Guttorm Alendal<sup>a</sup>

<sup>a</sup> University of Bergen, Department of Mathematics, Allegaten 41, 5008 Bergen, Norway

<sup>b</sup> Norwegian Research Center, NORCE, Climate, Bergen, Norway

<sup>c</sup> School of Marine Science and Policy, University of Delaware, Newark, DE 19716 USA

### ARTICLE INFO

#### Keywords:

Subsea CO<sub>2</sub> seepage  
Monitoring design  
Offshore  
Chemical sensors  
Optimal sensor placement

### ABSTRACT

Assurance monitoring of the marine environment is a required and intrinsic part of CO<sub>2</sub> storage project. To reduce the costs related to the monitoring effort, the monitoring program must be designed with optimal use of instrumentation. Here we use solution of a classical set cover problem to design placement of an array of fixed chemical sensors with the purpose of detecting a seep of CO<sub>2</sub> through the seafloor from an unknown location. The solution of the problem is not unique and different aspects, such as cost or existing infrastructure, can be added to define an optimal solution. We formulate an optimization problem and propose a method to generate footprints of potential seeps using an advection–diffusion model and a stoichiometric method for detection of small seepage CO<sub>2</sub> signals. We provide some numerical experiments to illustrate the concepts.

### 1. Introduction

Insulating the captured CO<sub>2</sub> from the atmosphere, by injecting it into geological formations, is the final step, and the whole purpose, of the Carbon Capture, and Storage (CCS) technology. Many promising storage sites are offshore, and especially the North Sea is considered a promising region for large scale storage (Halland et al., 2013).

Even though the offshore geological storage complexes are chosen and storage operations are designed to assure long term confinement, there is a risk that some of the injected CO<sub>2</sub>, being buoyant for the initial decade after injection, can migrate toward the surface and seep into the water column (Metz et al., 2005). If a seep occurs, it would reduce the climate change mitigation efficacy (Haugan and Joos, 2004; Torvanger et al., 2012), have impact on the carbon trading framework (García and Torvanger, 2019), and might, at least in the vicinity of the discharge, damage the ecosystem (Jones et al., 2015). In addition, even if all the stored CO<sub>2</sub> is successfully contained within the intended formation, a storage project could suffer from accusation of environmental impact (Boyd et al., 2013; Romanak et al., 2013).

This gives motivation for designing monitoring programs that would not only comply with regulations (Dixon et al., 2015), but also to rule out aforementioned unjustified accusations. A monitoring program can also be viewed as part of marine mapping and survey programs,

e.g., the Mareano programme (Buhl-Mortensen et al., 2014). It can be part of Ocean Observatories Initiative (OOI) (Smith et al., 2018), helping in assessing the overall health of the marine environment (Halpern et al., 2012), and be a tool for Marine Spatial Planning (MSP) (Domínguez-Tejo et al., 2016). As such, the monitoring program will be useful for communicating risks and benefits from large scale storage underneath the ocean, subsequently assisting in gaining public acceptance (Mabon et al., 2014).

Since the storage site must be monitored for a long period after the injection, and the area in which migrating CO<sub>2</sub> might reach the seafloor is large, the marine monitoring program will impose additional costs and challenges to the storage project (Oldenburg and Lewicki, 2006; Blackford et al., 2015, 2017). In particular, access to offshore sites for monitoring purposes will be harder and more costly than for onshore sites, partly due to seawater being hostile to instrumentation. Thus, the monitoring program must be not only effective but also cost efficient.

While geophysical monitoring technologies will be the backbone of the monitoring program for offshore CO<sub>2</sub> storage projects (Jenkins et al., 2015; Vermeul et al., 2016), the detection threshold of such techniques is of the order of 10<sup>3</sup> t CO<sub>2</sub> (Jenkins et al., 2015) and, thus, secondary monitoring strategies must be in place. Monitoring for CO<sub>2</sub> seeps through the seafloor, e.g., monitoring changes in bottom fauna or in the pelagic ecosystem (Wegener et al., 2008; Blackford et al., 2010), detecting bubbles from ship

\* Corresponding author.

E-mail address: [anna.oleynik@uib.no](mailto:anna.oleynik@uib.no) (A. Oleynik).

sonars (Brewer et al., 2006; Noble et al., 2012), or elevated concentration of dissolved gases (Alendal and Drange, 2001; Drange et al., 2001; Botnen et al., 2015; Vielstädte et al., 2015; Uchimoto et al., 2018) could be used to detect possible leakage at low levels. Seafloor monitoring can reduce the probability and magnitude of adverse events to occur if small seeps are not detected. However, the high variability of the marine environment, both in current conditions (Alendal et al., 2005; Ali et al., 2016) and in biochemical activities (Artioli et al., 2012; Romanak et al., 2012; Botnen et al., 2015), poses new challenges compared to the classical environmental monitoring procedures developed during decades of offshore petroleum activities.

Here we focus on seafloor monitoring strategy based on measuring CO<sub>2</sub> concentrations using fixed installations. Seawater CO<sub>2</sub> is highly dynamic in both space and time. The high variability arises from natural processes, such as photosynthesis/respiration, biosynthesis/dissolution of calcium carbonate (CaCO<sub>3</sub>) and changes in salinity, which affect the complex seawater CO<sub>2</sub> system. Therefore, in order to detect seeps from subsea reservoirs one needs to define an anomaly threshold able to distinguish the seepage signals from the natural variability of seawater CO<sub>2</sub>.

Botnen et al. (2015) demonstrated a stoichiometric approach for detection of small CO<sub>2</sub> signals that might arise when extra CO<sub>2</sub> stemming from subseafloor seeps dissolve into the water column. This method, henceforth referred to as the C-seep method, will be briefly described in Section 3.1. For a more thorough description the reader is referred to Botnen et al. (2015) and Omar et al. (2018). The C-seep method is based on the so-called back-calculation method (e.g., Gruber et al., 1996) and lies in between the statistical power analysis for obtaining an environmental baseline (Yang et al., 2011) and the pure process based monitoring approach suggested by Romanak et al. (2012). One of the utilities of the C-seep method is that it lowers the concentration threshold for a signal to become statistically significant and here we use the method to define detection limits.

In addition to the environmental statistics, the design of monitoring programs relies on probable seep scenarios that can only be predicted from characterization of the local geology and through flow migration process models. As the marine waters are in constant motion and are characterized with high variability, the footprints of leaks are thus highly anisotropic and strongly depend on the local oceanic and atmospheric conditions (Alendal et al., 2005; Ali et al., 2016). The role of numerical modeling in this context is summarized in Blackford et al. (2018).

Studies on how to design monitoring programs for detecting a seep from an unknown location, incorporating the natural variability, have been performed for fixed installations (Hvidevold et al., 2015, 2016; Greenwood et al., 2015; Ali et al., 2016) and for Autonomous Underwater Vehicles (AUV) (Maeda et al., 2015; Alendal, 2017). These studies did not take into account spatial heterogeneity of advective velocities and used very limited number of simulated leak scenarios.

In particular, Hvidevold et al. (2015, 2016) optimized the layout of a fixed array of chemical sensors on the seafloor, using the probability of detecting a seep as metric. It was assumed that all possible leaks would present the same footprint and the layout of the sensors was solved in order to reach the highest probability of detecting the seep. Here we introduce two main improvements to the previous studies; Firstly, we formulate an optimization problem in terms of the classical set cover problem, solution to which could be approximated using well established approximation algorithms. This formulation also allows for different cost functions to be minimized, which could be, e.g., the number of sensors, the cost associated with maintaining sensors, the probability of a leak at specific locations. We give more details on the set cover problem associated to the seep localization in Section 2. Secondly, we use different and more realistic footprints based on a physical model, as described in Section 3. We give illustrations of the concepts throughout the manuscript using model simulations, and some constructed examples. In particular, we use an area in the southern North Sea as area of study to illustrate our results, see Section 4. Finally, we discuss possible extensions of the approach in Section 5.

## 2. Sensor placement algorithm

In this section we formulate the optimization problem and define possible cost functions. Let  $\Omega$  be a bounded subset of  $\mathbb{R}^d$ , and  $U_i \subseteq \Omega$ ,  $i = 1, \dots, N$ , be subsets of  $\Omega$  of a positive Lebesgue measure which we denote as  $|U_i|$ . We can think of sets  $U_i$  as the footprints associated to a potential leak  $i$ , which may occur, that is being active, with some probability  $p_i$ . We say that a set  $U_i$  is detected at a point  $x \in \Omega$  if  $x \in U_i$ . Obviously one point  $x$  may detect several sets. Here we would like to discuss how can we place a number of sensor points so that they detect all the sets  $U_i$  and their placement is optimal with respect to a given cost function. For a given cost function  $\Gamma: \mathcal{P}(\Omega) \rightarrow \mathbb{R}_{\geq 0}$  we define optimality as the set of points  $\{x_1, \dots, x_n\}$  which lead to a minimal cost. Here we use cost functions that can be written as the sum of pointwise costs  $\gamma(x)$  incurred by each sensor point  $x$ , that is,

$$\Gamma(\{x_1, \dots, x_n\}) = \sum_{i=1}^n \gamma(x_i) \quad (1)$$

for a pointwise cost function  $\gamma: \Omega \rightarrow \mathbb{R}_{\geq 0}$ .

The most straightforward criteria for optimality is the number of points  $n$  to be minimal. This corresponds to a pointwise cost function  $\gamma(x) = 1$ . We would like to emphasize that we do not distinguish between two different points  $x$  and  $y$  if they detect the same sets. That is, if  $x, y \in U_i$  for  $i \in I$  and  $x, y \notin U_j$  for  $j \notin I$ , then  $x$  and  $y$  are in the same equivalence class.

Let  $I(x) = \{i \in \{1, \dots, N\}: x \in U_i\}$  be the index set corresponding to a point  $x \in \Omega$ . Assume  $x \in U_i$ ,  $i \in I(x)$  and  $x \notin U_j$ ,  $j \notin I(x)$  in the examples below. The pointwise cost function can depend on the area of the equivalence class, that is,

$$\gamma(x) = |\Omega| - |\bigcap_{i \in I(x)} U_i| \quad \text{or} \quad \gamma(x) = |\bigcap_{i \in I(x)} U_i|^{-q}, \quad q > 0. \quad (2)$$

Let  $N_k$  be a number of sets detected by a point  $x_k \in \{x_1, \dots, x_n\}$ , that is  $N_k = |I(x_k)|$ . We define

$$N_{\text{over}} = \sum_{k=1}^n N_k - N, \quad (3)$$

which we call the overdetection number, or simply overdetection, associated with  $\{x_1, \dots, x_n\}$ . In order to maximize the overdetection number while keeping the number of sensors low, we introduce the pointwise cost function

$$\gamma(x) = (N + 1) - |I(x)|. \quad (4)$$

Observe that if  $\{x_1, \dots, x_n\}$  detect all the sets  $U_i$ ,  $i = 1, \dots, N$ , then  $N_{\text{over}} \geq 0$ . In particular,

$$N_{\text{over}} = n(N + 1) - \Gamma_N(x_1, \dots, x_n) - N,$$

where  $\Gamma_N(x_1, \dots, x_n)$  is given by Eq. (1) with the pointwise cost function as in Eq. (4).

Now let  $U_i$ ,  $i = 1, \dots, N$ , be associated with certain probabilities  $p_i$ . If  $p_i$  is a probability that leak  $i$  is active, and assuming that exactly one leak is active, the natural way to define a pointwise cost function would be

$$\gamma(x) = 1 - \sum_{i \in I(x)} p_i. \quad (5)$$

In this case minimizing  $\Gamma(\{x_1, \dots, x_n\})$  would correspond to placing sensors so that they detect leaks, associated with higher probability of being active, more times. This cost function could be view as a particular case of  $\Gamma_N$  when  $p_i = 1/N$  in Eq. (4). In a similar fashion, one can define the probability of overdetection as

$$P_{\text{over}} = (n - 1) - \Gamma_P(x_1, \dots, x_n), \quad (6)$$

where  $\Gamma_P(x_1, \dots, x_n)$  is associated with the pointwise cost in Eq. (5).

Observe that even within one approach there are multiple ways to choose the cost function, see Eq. (2). This choice should be motivated

by the application. The only requirement is that the pointwise cost function  $\gamma$  is non-negative.

We summarize the problem as follows. Find  $x_j$ ,  $j = 1, \dots, n \leq N$  in  $\Omega$  such that (i) all the sets are detected and (ii) the total cost of placing  $\{x_j\}$ , that is  $\Gamma(\{x_1, \dots, x_n\}) = \sum_{j=1}^n \gamma(x_j)$  is minimal.

This is a typical example of a set cover problem with the universe  $\mathcal{U} = \{U_i\}$  and the collection of sets  $S = \{U_{k \in I(x)} \mid \forall x \in \Omega\}$ . Dealing with a small number of sets or sets with rather simple intersections, is trivial. However, the complexity of the problem increases considerably when the number of intersecting sets gets large. In fact, the problem is computationally difficult to solve and different approximation algorithms have been developed to deal with problems of this kind. For the review of the problem and numerical methods see, e.g., Hochbaum (1997) and Schrijver (1986). Here we use a linear integer program formulation, which we present below.

Let  $\{x_k\}$ ,  $k = 1, \dots, N_{\text{grid}}$ ,  $N_{\text{grid}} \gg N$ , be a grid defined on  $\Omega$ . For every grid point  $x_k$  we create a vector  $v^{(k)} \in \{0, 1\}^N$ ,

$$v_i^{(k)} = \begin{cases} 1 & \text{if } x_k \in U_i \\ 0 & \text{if } x_k \notin U_i, \end{cases}$$

and calculate the pointwise cost  $\gamma_k = \gamma(x_k)$ . Then we define the matrix  $A \in \{0, 1\}^{N \times N_{\text{grid}}}$  with columns  $v^{(k)}$  and the cost vector  $w \in \mathbb{R}^{N_{\text{grid}}}$  as

$$V = (v^{(1)}, v^{(2)}, \dots, v^{(N_{\text{grid}})}), \quad w = (\gamma_1, \dots, \gamma_{N_{\text{grid}}}).$$

In order to simplify the problem, we can remove repeated columns and zero columns in  $V$  and the corresponding elements in  $w$ , but keep the ones that corresponds to the minimum of  $w$  for the duplicated columns. By doing so we obtain a smaller matrix  $\tilde{V} \in \{0, 1\}^{N \times M}$ ,  $M > N$  and  $\tilde{w}$ , respectively. Each column of the matrix  $\tilde{V}$  corresponds to the sets  $\Omega_i$ ,  $i \in I(x_k)$  that could be detected with the point  $x_k$ .

Finally, we formalize the problem as

$$\begin{aligned} & \min_{z \in \{0, 1\}^M} \tilde{w}^T z, \\ & \text{s. t. } \tilde{V} z \geq 1. \end{aligned} \quad (7)$$

Each nonzero element  $z_j$  in  $z \in \{0, 1\}^M$  corresponds to a point  $x_j$  that can be placed anywhere in the set  $X_j$  defined as

$$X_j = \bigcap_{i \in I(x_j)} U_i.$$

Obviously, the number of nonzero elements in  $z$  is at most  $N$  and the total cost  $\Gamma$  is given as  $\tilde{w}^T z$ . The matrix  $\tilde{V}$  and  $\tilde{w}$  in Eq. (7) could be further pruned by removing the columns corresponding to  $x_j$  that do not have desired properties. For example, dealing with fine grids, one may remove the columns corresponding to  $x_j$  that has too small area  $|X_j|$ .

In the next section we propose a method to generate the sets  $U_i$  associated with leak footprint.

### 3. Generating footprints

#### 3.1. Determining detection limits using the C-seep method

The C-seep method isolates the effect of leakage  $\text{CO}_2$  on the Dissolved Inorganic Carbon (DIC) by comparing two measurements acquired at the reference station (ref) and the station being monitored ( $m$ ). The method first minimizes the DIC differences between the two stations that arise from differences in natural processes (Botnen et al., 2015). This is achieved by correcting the DIC of the monitored station back to that of the reference station. It then assumes that the remaining natural variability of the two stations are identical and, thus, seepage of  $\text{CO}_2$  can be computed as the difference between the corrected DIC and the DIC at the reference station (Omar et al., 2018).

Below we present shortly the calculations. A description of the variables can be found in Table 1.

From Botnen et al. (2015) the corrected DIC is calculated as

**Table 1**

A list of variables used in calculation of  $C_{\text{seep}}$  in Eq. (11) and the measurements for stations 5, 6, and 7 as in Fig. 2(a).

Variable	Description (units)	Examples of measured and calculated values		
		ref = 6	m = 5	m = 7
S	Salinity	35.1147	35.1033	35.1549
P	PO <sub>4</sub> ( $\mu\text{mol kg}^{-1}$ )	0.5933	0.5611	0.6869
A	Alkalinity ( $\mu\text{mol kg}^{-1}$ )	2302	2304	2306
C	DIC ( $\mu\text{mol kg}^{-1}$ )	2122.7576	2111.8200	2134.8600
$\tilde{C}$	corrected DIC ( $\mu\text{mol kg}^{-1}$ )		2159.3208	2133.9360

$$\tilde{C}(m) = (C(m) - r_{C:P} \Delta P - 0.5 \Delta \tilde{A} - A_0) \frac{S(\text{ref})}{S(m)} + A_0, \quad (8)$$

where  $\Delta P$  is the difference of phosphate (PO<sub>4</sub>)

$$\Delta P = P(m) - P(\text{ref}),$$

$\Delta \tilde{A}$  is the difference of corrected alkalinity

$$\Delta \tilde{A} = (A(m) - A_0) \frac{S(\text{ref})}{S(m)} + A_0 - A(\text{ref}), \quad (9)$$

and  $r_{C:P}$  is the carbon to PO<sub>4</sub> Redfield ratio, which relates the DIC and PO<sub>4</sub> produced/consumed during organic matter cycling (Redfield, 1934). The term  $A_0$  stands for the estimated alkalinity when salinity  $S = 0$ , assuming that salinity and alkalinity obey the linear relationship

$$A = aS + A_0, \quad (10)$$

see, e.g., Friis et al. (2003).

Assuming that  $C(\text{ref})$  is not influenced by seeps we estimate the excess of  $\text{CO}_2$  at the monitoring station as

$$C_{\text{seep}}(m) = \tilde{C}(m) - C(\text{ref}). \quad (11)$$

In Table 1 we give examples of the measurements for three stations 5, 6 and 7, see Fig. 2(a). Here we used the publicly available data seawater  $\text{CO}_2$  measurements from a cruise in February 2002 (Olsen et al., 2016).

Observe that when  $m = \text{ref}$  we simply have  $\tilde{C}(\text{ref}) = C(\text{ref})$  and, thus,  $C_{\text{seep}} = 0$ . In reality, this is however not the case due to errors in measurements and estimated parameters. Including these errors in the model results in

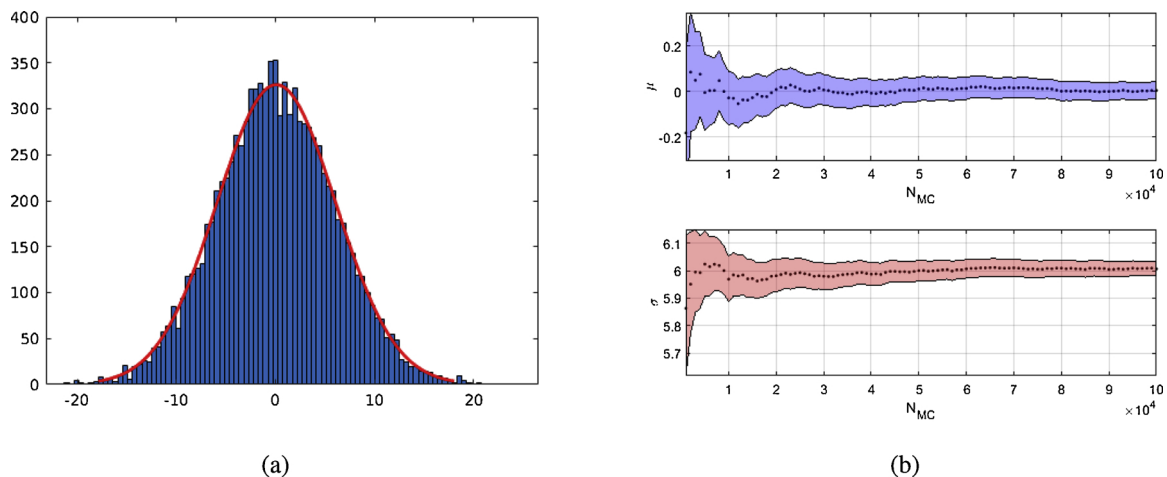
$$C_{\text{seep}}(m) = \tilde{C}(m) - C(\text{ref}) + \varepsilon, \quad (12)$$

with error distribution  $\varepsilon$ .

Given a threshold  $\theta$ , we can then attribute the excess of  $\text{CO}_2$  to the seep if  $C_{\text{seep}}(m) > \theta$ . Choosing too high thresholds  $\theta$  will not allow for seep detection, while too low thresholds may lead to false alarms.

In order to estimate  $\varepsilon$  we performed Monte Carlo (MC) simulations of the underlying model in Eqs. (8)–(11), assuming normal distribution for the measurements and model parameters errors. In particular, we assumed that salinity measurements have standard deviation  $\sigma_S = 0.003$  and the standard deviations of alkalinity and phosphate are chosen to be 0.1% and 0.05% of measured values, respectively. For the Redfield ratio we use  $r_{C:P} = 117 \mu\text{mol kg}^{-1}$  with the standard deviation equal to  $14 \mu\text{mol kg}^{-1}$ , based on Anderson and Sarmiento (1994). Finally, we assumed that  $A_0$  in Eq. (10) has normal distribution with 1817.46  $\mu\text{mol kg}^{-1}$  mean and the standard deviation 48.02  $\mu\text{mol kg}^{-1}$  based on Omar et al. (2010).

We plot the histogram of the error  $\varepsilon$  from the MC simulations with  $\text{ref} = m = 6$  in Fig. 1(a) fitted with a normal density function. We chose the station labeled 6 for the reference station as in Omar et al. (2018). There is, however, no particular reason for this choice. In Fig. 1(b) we plot the estimated mean  $\mu$  and standard deviation  $\sigma$  as a function of the sampling number  $N_{\text{MC}}$  together with the 95% confidence intervals for this distribution. The confidence intervals were calculated as in Harding



**Fig. 1.** (a) Histogram of the error  $\epsilon$  obtained by the MC sampling for  $ref = m = 6$  and  $N_{MC} = 10^4$ , with superimposed fitted normal density. (b) Estimation of the mean  $\mu$  and standard deviation  $\sigma$  with 95% confidence intervals, for the model with  $m = ref = 6$  using MC simulations with the number of samples  $N_{MC}$  between  $10^4$  and  $10^5$ .

et al. (2014). In particular, we obtained the 95% confidence interval for  $\mu$  as  $[-0.0313, 0.0432]$  and for  $\sigma$  as  $[5.98086, 6.03351]$ . This gives us motivation for using  $\theta = 2\sigma$ .

In Fig. 2 we illustrate the outcome of 10 000 random draws of MC sampling as the box plot for  $ref = 6$  and  $m = 1, \dots, 10$ . The shaded areas corresponds to detection threshold  $|C_{seep}(m)| < \theta$  where we set  $\theta = k\sigma$ ,  $k = 1, 2, 3$ .

The measurements used here did not contain any seepage signal. Therefore, the box plot indicates the degree at which the assumption of identical background DIC in the reference and different monitoring stations is met. For stations that have similar background DIC as the reference station we require  $|C_{seep}| < \theta$ . Positive values above  $\theta$  mean that monitored station has higher background DIC than the reference station, whereas negative values indicate that the monitored station has lower DIC levels compared to the reference station. Station 6 can be considered as a good reference for all stations except, perhaps stations 3 and 9, when choosing  $\theta = 2\sigma$ . By choosing  $\theta = \sigma$  we risk to have too many false alarms, in particular from station 3 and 9, while by choosing  $\theta = 3\sigma$  we may miss a seep if it occurs. The above consideration of the C-seep is based on a limited data from one winter cruise. Including more data would provide us better knowledge of parameters and errors and result in improved estimates of C-seep.

Moreover, in order for a seep to be detected in a large area reference and monitoring stations must be placed in a such way that they capture the signal originating from most of the potential leak locations and, at the same time, can be used as reference stations for each other. In this paper we focus on the placement of measurement stations. In order to decide if they are representing the natural variability of the region, and thus can be used as the reference stations, the measurements must be collected from the identified locations. This is however outside of the scope of the present paper.

The choice of  $\theta = 12.0 \mu\text{mol kg}^{-1}$  corresponds to  $\theta = 0.5445 \times 10^{-3} \text{ kg m}^{-3}$ , see Table 2, which is more than twice larger than  $0.2260 \times 10^{-3} \text{ kg m}^{-3}$  used in Hvidevold et al. (2016).

### 3.2. Simulating $\text{CO}_2$ footprints

As we mentioned before, designing a monitoring program of subsea CCS reservoirs is challenging due to both the variability of the environment and ocean dynamics. Assuming that the measurements can be corrected for the natural variability, as in Section 3.1, we now focus on the ocean dynamics.

Transport of contaminants, such as  $\text{CO}_2$ , in the ocean is typically modeled using General Circulation Models (GCMs) with additional

transport equations for tracers. These models are computationally demanding and, hence, only allow to simulate a few leak scenarios. Under the assumption that the contaminant is a passive tracer, i.e., does not influence on the water density, the tracer transport equations can be integrated off-line. The GCMs can be used to produce characteristic spatial and temporal velocity fields, accounting for tides, storm events, and topographic steering of the currents. Such current statistics, preferably supported by in-situ current time series, together with an advection–diffusion model, being orders of magnitude less computationally demanding than the GCMs, can be used to simulate many more leak scenarios.

Let the transport of a contaminant be given by the advection–diffusion model

$$\frac{\partial c}{\partial t} = D\Delta c - W \cdot \nabla c + f, \quad x \in \Omega, \quad t \in [t_0, t_0 + T], \quad (13)$$

with

$$c(x, t)|_{x \in \partial\Omega} = 0, \quad c(x, t_0) = 0. \quad (14)$$

Here  $c(x, t)$  is the concentration of a contaminant,  $\Omega$  is a bounded connected domain in  $\mathbb{R}^d$ ,  $d = 2, 3$ ,  $W(x, t) \in \mathbb{R}^d$  is a velocity field, and  $D(x, t) \geq 0$  is the diffusion coefficient. The source term  $f(x, t)$  is assumed to be in the form

$$f(x) = q\delta(x - z), \quad (15)$$

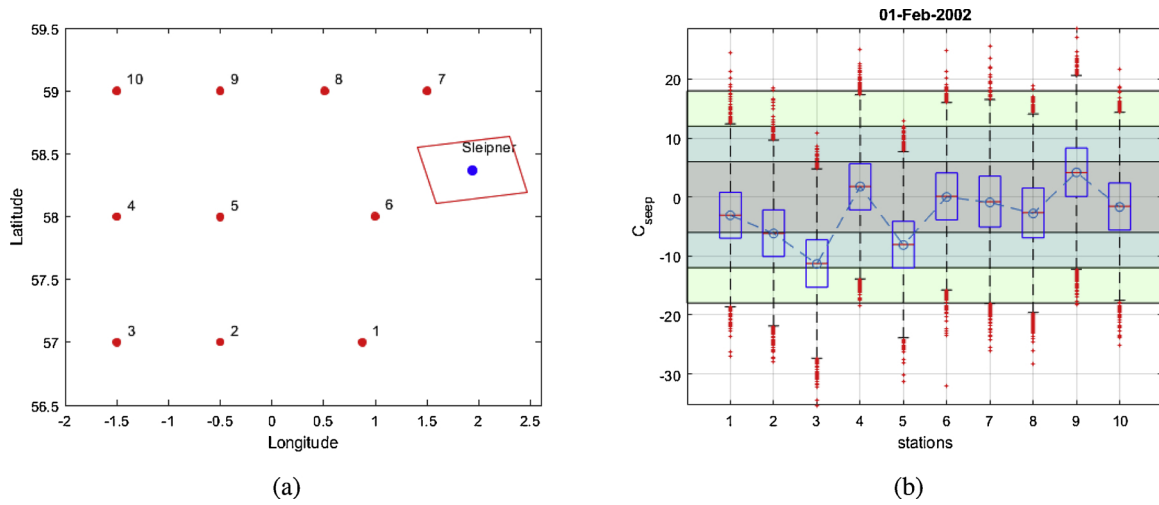
where  $\delta$  is the  $d$ -dimensional delta function,  $q > 0$  is the intensity, or the seepage rate, and  $z$  is the location of the source. For simplicity we assume that  $q$  is constant.

In applications, the point source  $z$  is substituted by a small region around  $z$  which amounts to replacing  $\delta(x - z)$  with functions of small support. Thus,  $\delta(x - z)$  can be viewed as a limiting case when the support is getting smaller and smaller. Here, we assume that the domain  $\Omega$  is large enough and the sources positioned far from the boundary so that the contaminant does not reach  $\partial\Omega$ . The latter assure that spurious effects from the open boundary conditions enforced on the lateral boundaries do not affect the results.

In our examples we use  $W(x, t)$  obtained from a 800 m resolution regional Bergen Ocean Model (BOM)<sup>1</sup> set up for North Sea (Ali et al., 2016). In the vertical the model uses 41 sigma-layers, distributed with higher resolution (1 m) near the free surface and the sea floor. The resulting current is dominated by semi diurnal tidal signal with an average speed close to 10 cm/s, and an amplitude less than 10 cm/s, for

<sup>1</sup> <https://org.uib.no/bom/>.





**Fig. 2.** (a) Stations locations (red dots) and Sleipner A (blue dot), the red polygon marks the area used in the numerical simulations in Section 3. (b) Example of the C-seep method with the box plot. The gray area corresponds to  $[\mu - \sigma, \mu + \sigma]$ , blue to  $[\mu - 2\sigma, \mu + 2\sigma]$  and green to  $[\mu - 3\sigma, \mu + 3\sigma]$ . Here  $\mu = 0$  and  $\sigma = 6.0$  obtained as in Fig. 1,  $N_{MC} = 10^4$ ,  $ref = 6$ ,  $m = 1, \dots, 10$ . The central red mark indicates the median, the bottom and top edges of the box indicate the 25th and 75th percentiles. The whiskers extend to the extreme points, not considered outliers, and the outliers are marked as red crosses.

**Table 2**

Columns as follows: (1–2) Detection threshold  $\theta$  of  $CO_2$  in different units under the assumption of the water density  $\rho = 1029 \text{ kg m}^{-3}$ , (3–4) flow rate of  $CO_2$  per  $m^3$ , (5) total flow rate of  $CO_2$  (via the  $800 \times 800 \times 1 \text{ m}^3$  model cell).

Detection threshold		Leak input		
$\mu\text{mol kg}^{-1}$	$\text{kg m}^{-3}$	Per cubic meter per second $\mu\text{mol m}^{-3} \text{s}^{-1}$	$\text{kg m}^{-3} \text{s}^{-1}$	Total input per second $\text{kg s}^{-1}$
$1.20 \times 10^1$	$5.445 \times 10^{-4}$	$4.42 \times 10^0$ (high)	$1.94 \times 10^{-7}$ (high)	$9.53 \times 10^0$ (high)
		$4.42 \times 10^{-2}$ (low)	$1.94 \times 10^{-9}$ (low)	$9.53 \times 10^{-2}$ (low)

details see Ali et al. (2016). For simplicity, we considered only the 1 m thick bottom layer. In our simulations  $\Omega$  is a  $72.8 \times 74.4 \text{ km}^2$  rectangular area that, in the geographic coordinate system, corresponds to the region marked in Fig. 2(a). We use  $93 \times 91$  grid cells of  $800 \times 800 \text{ m}^2$ . In Fig. 3(a) we plot the mean of the current speed at the bottom layer. To illustrate variability in speed and direction, we plot the wind rose diagram for the currents. In particular, in Fig. 3(b) we plot the diagram for the currents over the whole area  $\Omega$  and in Fig. 3(c) and (d) at particular locations, which are marked in Fig. 3(a) with the red circle and black square. As the horizontal diffusion is insignificant for this grid size, we set  $D = 0$ .

When we would like to emphasize the parameter dependence of the solution  $c(x, t)$  we add the parameters of interest after semicolon, e.g.,  $c(x, t; t_0)$  or  $c(x, t; t_0, q, z)$ .

Since the model above Eqs. (13)–(15) is linear, a multiple leak scenario solution, that is, when  $f(x, t) = \sum q_j \delta(x - z_j)$  instead of Eq. (15), can be calculated as

$$c(x, t) = \sum_j q_j c_j(x, t),$$

where  $c_j(x, t)$  is the solution of Eqs. (13)–(15) with  $q = 1$  and  $z = z_j$ .

Let  $q$  be the intensity,  $t_0$  and  $T$  the seep starting time and its duration, respectively, and  $\theta$  the detection threshold obtained as in Section 3.1. Then a leak footprint can be defined as, e.g., the maximal footprint

$$U_{t_0, T}^{\max}(q, \theta) = \{x: c(x, t; q) > \theta \text{ for some } t \in [t_0, t_0 + T]\}. \quad (16)$$

Since  $c(x, t; q) = q/\bar{q} c(x, t; \bar{q})$  we have  $U_{t_0, T}^{\max}(q, \theta) = U_{t_0, T}^{\max}(\bar{q}, \bar{q}q/\theta)$ , for  $\theta, \bar{q}, q > 0$ , which is a useful property when generating footprints with different parameters  $q$  and  $\theta$ . In addition, as the model is mass conserving, we can be guaranteed that all leaks with the flux rate larger than  $q$  and lasting longer than  $T$  will be detected, if the method is

working for  $T$  and  $q$ . Indeed, let  $\theta > 0$  and  $t_0$  be fixed. Then  $c(x, t; q) \geq \theta$  implies  $c(x, t; \bar{q}) \geq \theta$  for  $\bar{q} \geq q$ . Hence,  $U_{t_0, T}^{\max}(q, \theta) \subseteq U_{t_0, T}^{\max}(\bar{q}, \theta)$  for  $\bar{q} \geq q$  and  $\bar{T} \geq T$ , see Fig. 4(b).

Our goal is to detect leakages, which can go unnoticed for a rather long time. In particular, we use, what is referred in Blackford et al. (2008) to as a long term-diffuse seepage. That is, we assume a constant low-level seepage of  $CO_2$ , spread homogeneously across the area of one model box ( $0.64 \text{ km}^2$ ). We use the same seepage rates as in Blackford et al. (2008), namely, a high seepage rate of  $0.0953 \text{ kg s}$  and a low seepage rate of  $9.53 \text{ kg s}$ , both recalculated for the considered model environment (model box of  $0.64 \text{ km}^2$ ). That is, we have  $q_{\text{high}} = 1.94 \times 10^{-7} \text{ kg m}^{-3} \text{ s}^{-1}$  and  $q_{\text{low}} = 1.94 \times 10^{-9} \text{ kg m}^{-3} \text{ s}^{-1}$ , see Table 2.

Further on, we assume  $T$  and  $\theta$  being fixed. For this reason, we omit  $T$  and  $\theta$  in the notation and simply write  $U_{t_0}^{\max}(q)$ .

Here we consider 20 leak locations  $z \in \{z_1, \dots, z_{20}\}$  selected uniformly at random in  $\Omega_0 \subset \Omega$ , see the locations marked red in Fig. 5. We assume that these are the only possible leak locations, which is a simplification. In order to take into account different footprint topologies at the same location, we vary the starting time  $t_0$ . For illustrations we chose 10 different starting points  $t_0 \in \{t_{0,1}, \dots, t_{0,10}\}$ , with  $t_{0,k} - t_{0,k-1} = 126.57 \text{ h}$ ,  $k = 2, \dots, 10$ , where  $t_{0,1}$  was set to 02 February 2012 02:17:08. That is, for each leak location  $z_j$ ,  $j = 1, \dots, 20$  we run the mode, Eqs. (13)–(15), with 10 different starting times  $t_0$ . The time duration  $T = 134.29 \text{ h}$  was chosen such that the seep signal does not reach the boundary of the domain, i.e.,  $c(\partial\Omega, t) = 0$  at all times  $t \in [t_0, t_0 + T]$  for all 200 leak simulations. For each leak simulation we computed two corresponding footprints  $U_{t_0}^{\max}(q_{\text{high}})$  and  $U_{t_0}^{\max}(q_{\text{low}})$  with  $\theta, q_{\text{high}}$ , and  $q_{\text{low}}$  as in Table 2.

In Fig. 4 we give an example of  $c(x, t_0 + T)$  for one leak location at  $t_0 = t_{0,1}$  and the corresponding footprints. As expected, the footprint corresponding to the low rate is contained in the one corresponding to

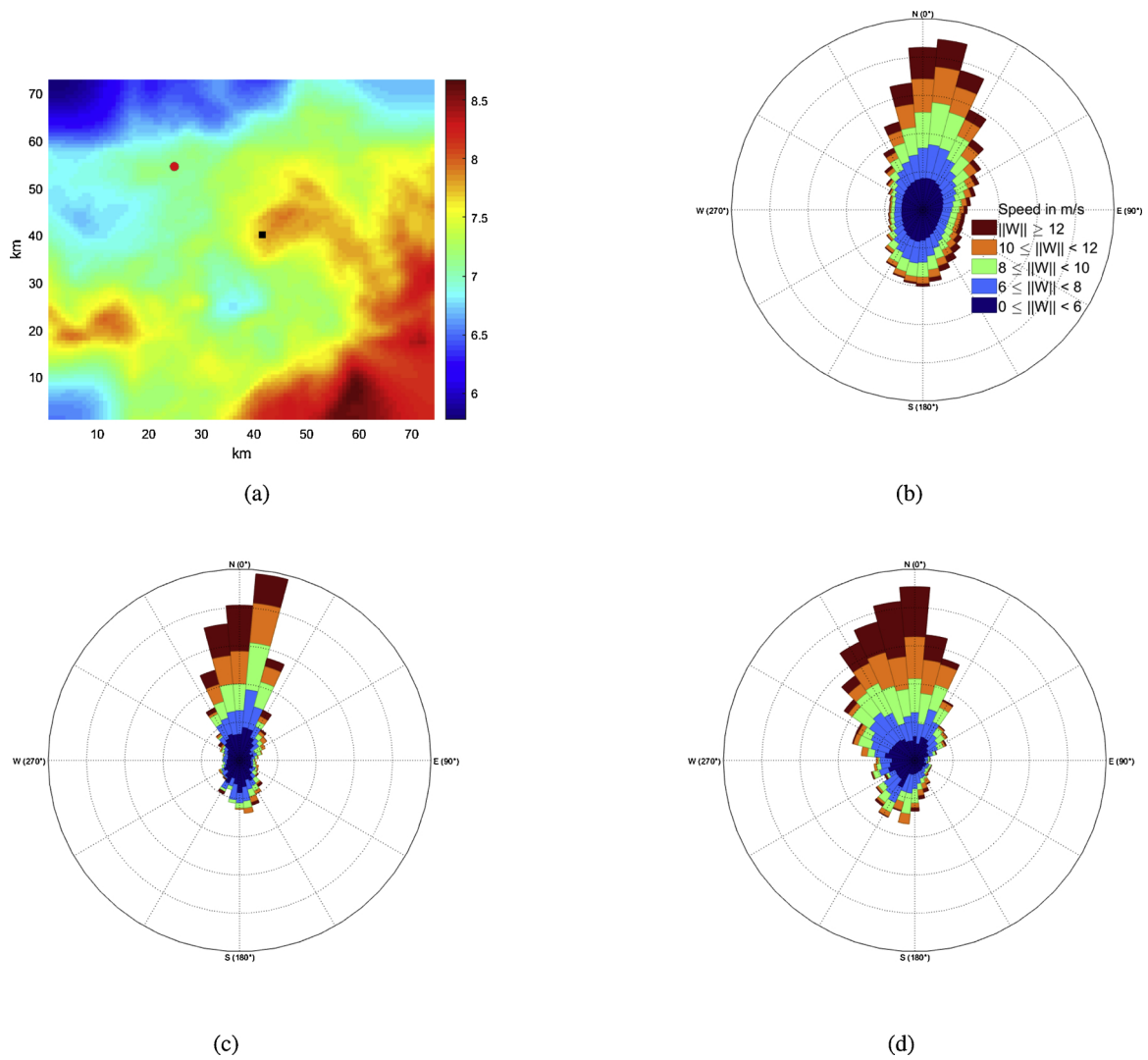


Fig. 3. Statistics for the bottom layer currents time-series from BOM simulations of  $72.8 \times 74.4 \text{ km}^2$  area centered at  $1.94 \text{ E } 58.36 \text{ N}$ , over a time span from 02 February 2012 to 4 April 2012. (a) Mean of the current speed (in cm/s) and two marked locations, (b) wind rose for the currents over the whole area, (c) wind rose for the currents from the location marked with the circle, and (d) wind rose for the currents from the location marked with the polygon.

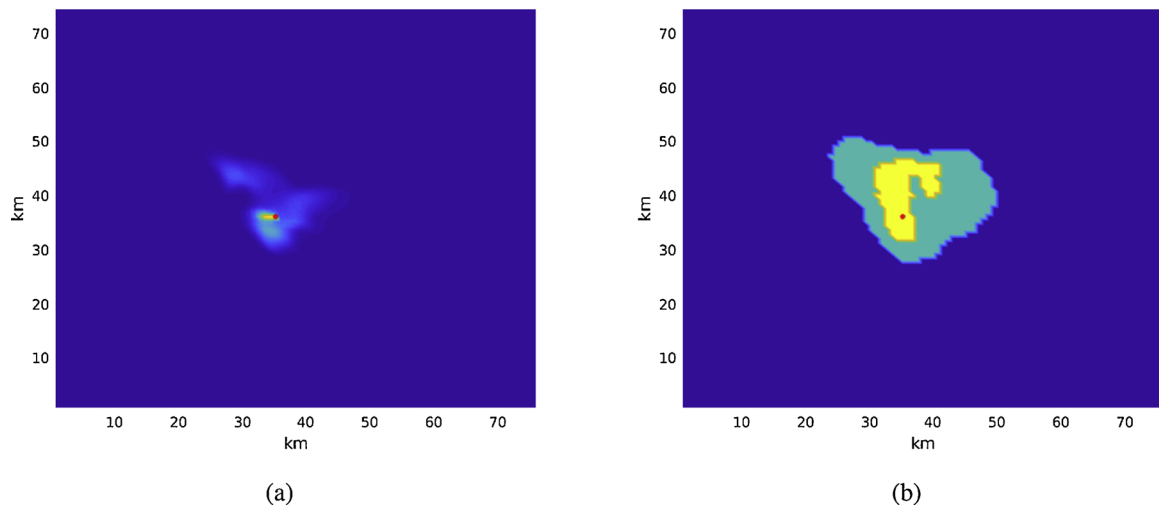


Fig. 4. (a) An example of  $c(x, t_0 + T)$  and (b) the corresponding footprints  $U_{t_0}^{\max}(q_{\text{high}})$  (green) and  $U_{t_0}^{\max}(q_{\text{low}})$  (yellow).

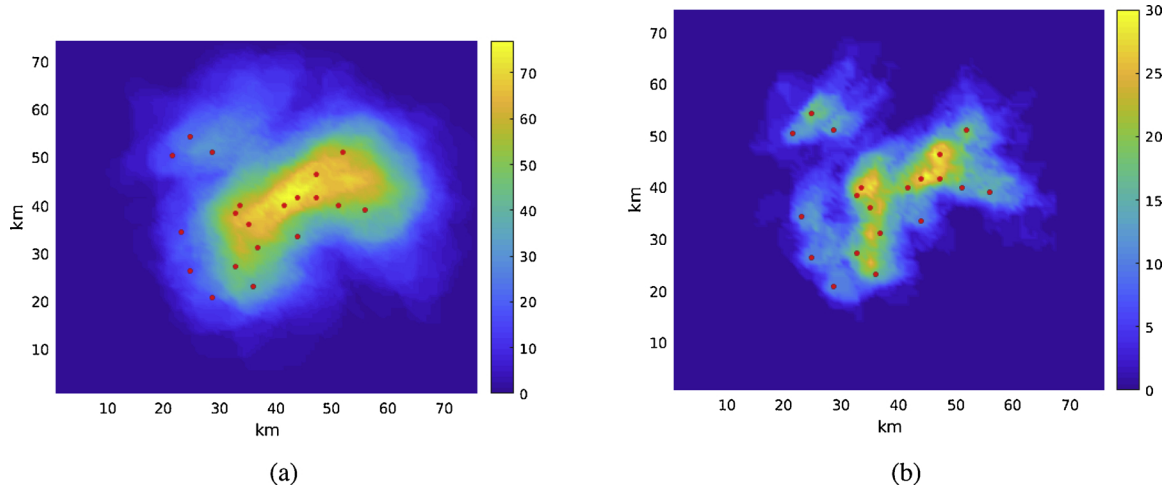


Fig. 5. (a) Overlap of 200 footprints  $\{U_{t_0}^{\max}(q_{\text{high}})\}$  and (b) overlap of 200 footprints  $\{U_{t_0}^{\max}(q_{\text{low}})\}$ , simulated for 20 different locations with 10 different starting points  $t_0$  for the duration of  $T = 134.29$  h.

the high rate, i.e.,  $U_{t_0}^{\max}(q_{\text{low}}) \subseteq U_{t_0}^{\max}(q_{\text{high}})$ .

To illustrate the overlap of 200 footprints with both seepage rates, we plot the color-map of the footprint's intersection in Fig. 5. The color code corresponds to the number of intersecting sets.

From Fig. 4(b), Fig. 5 and properties mentioned above, it is clear that a larger ratio  $\theta/q$  implies the larger number of intersecting sets, and thus, potentially, smaller number of sensors needed. We give examples for both  $q = q_{\text{high}}$  and  $q = q_{\text{low}}$  in the next section.

#### 4. Numerical experiments

In this section we illustrate the optimization problem, defined in Eq. (7), using different cost functions. The goal is to find the optimal sensors placement in the region  $\Omega$  to detect all potential leaks and compare the outcomes corresponding to different cost functions. To illustrate the method we used 200 maximal footprints, as described in Section 3, generated for two different seepage rates  $q$  and the detection threshold  $\theta$  as in Table 2, see Fig. 5. In the notation of Section 2, these footprints are the sets  $U_i$ ,  $i = 1, \dots, N$  with  $N = 200$ . First we consider the sets generated with the high  $q$  and then with the low. There is no reason treating these two cases together since the sets corresponding to the low  $q$  are contained in the sets with the high  $q$ , see the previous section for details.

To solve Eq. (7) we used the Matlab inbuilt function *intlinprog*. For all the test examples the program found an optimal solution which was indicated by the zero relative dual-primal gap.

In Fig. 6 we plot solutions to the problem with different choice of cost functions. Black crosses correspond to sensor locations, the red dots to the leak locations. The numbers next to the crosses indicate the number of leaks (out of total number of possible leaks) that can be detected per sensor. The color-plot indicates the sets  $U_i$  with the highest values corresponding to the place of maximum overlap. We specify the over-detection number and the approach name that corresponds to the choice of  $\gamma$ , see Table 3.

In particular, in Fig. 6(a) we plot a solution that gives the minimal number of sensor, which is equal to 7 in this case. In this case we solve Eq. (7) with  $\gamma(x) = 1$ , which we refer to the *minimal number* approach. The solutions in Fig. 6(b) and (c) are solutions of the optimization problem Eq. (7) with  $\gamma(x)$  chosen as Eqs. (4) and (5), respectively. The probabilities  $p_i$ , were set as in Fig. 7(a) with the normalization factor  $815 \times 10$ , so that  $\sum p_i = 1$ . One can see that both solutions are the solution of the minimal number approach (unweighted cover problem), that is, when  $\gamma(x) = 1$ . As the unweighted cover problem could have many solutions, introducing relevant costs could not only give minimal number of sensors but also allows to make the detection more robust.

That is, the solution in Fig. 6(b) maximizes the number of over-detected leaks while minimizing the number of sensors. Here all the leaks have equal probability of being active. Then the cost function Eq. (4) can be viewed as a particular case of Eq. (5) with  $p_i = 1/N$ ,  $i = 1, \dots, N$ . We call the methods corresponding to these cost functions as *maximal over-detection* and *maximal probability*, respectively.

Finally, to demonstrate the ability to include cost in the monitoring design, we consider the pointwise cost function

$$\gamma(x) = \alpha(\sqrt{(x - x_c)^2 + (y - y_c)^2} + 1), \tag{17}$$

where  $(x_c, y_c)$  are the coordinates of the center of domain and  $\alpha > 0$  is a scalar, see Fig. 7(b). This function aims to illustrate the operational cost which might be site dependent. The assumption is that there is some infrastructure in the center of the domain, e.g., a platform or the injection infrastructure, and that the cost of maintaining sensors increase with distance from this center point. We plot the solution to the optimization problem with this cost function in Fig. 6(d), and call the approach the *minimal operational cost*. The number of sensors in this case is equal to 12.

In order to compare the solutions  $\{x_1, \dots, x_n\}$  in Fig. 6 and motivate the choice of the method names, we calculate over-detection numbers, see Eqs. (2) and (6), and operational costs  $\Gamma_{\text{op, cost}}$  associated with Eq. (17) for each solution, see Table 3.

In addition, we ran 10 000 leak simulations initiated uniformly at random within available time range for currents, that is, between 2 February 2012 02:17:08 and 30 May 00:00:00. The leak locations were randomly drawn from the 20 leak locations. The seepage rate was fixed to high, see Table 2. Next, we have checked how many of the random leaks would be missed by the sensors placed as in Fig. 6. We report the results in the last row of Table 3. All the cases indicate less than 3% failure rate.

Next, we apply the same methods to the leaks generated with the low seepage rate. As shown in Fig. 8 and Table 4 the number of sensors has increased more than twice compared to the high seepage rate case. Analogous to the previous example, we test the sensors locations on the 10 000 leaks. The failure rate has increased, but still remains below 5%, which we consider acceptable. In order to decrease the failure rate without over-fitting, one requires longer time series of the current simulations. We do not pursue this task here.

We would like to point out that, even though using the minimal operational cost increased the total number of sensors in both examples, it did not improve the detection results for the randomly generated leaks.

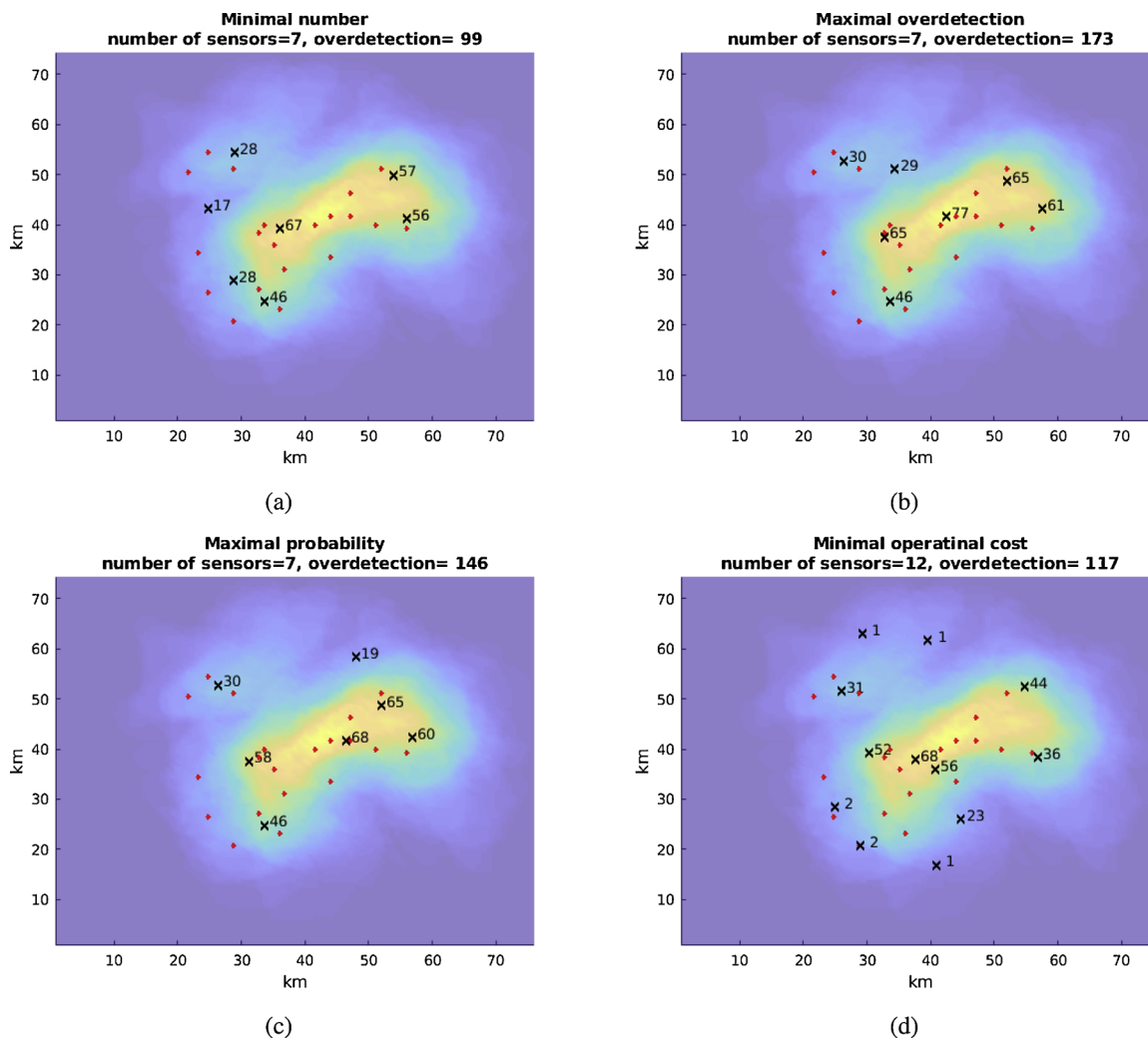


Fig. 6. Optimal sensor positions (black crosses) with the number of detected leaks for different choice of  $\Gamma$ . In particular, we used  $\gamma(x) = 1$  in (a),  $\gamma(x)$  given by Eq. (4) in (b), by Eq. (5) in (c) and by Eq. (17) in (d). For details see Table 3.

### 5. Concluding remarks

We have demonstrated how solving a classical mathematical problem could be used to design marine monitoring programs. To perform a more comprehensive design, valid for an actual storage site, will require a geological survey identifying potential leak locations and their relative probability, as for example in Fig. 7(a). In addition, a comprehensive environmental baseline is needed for establishing better detection limits, from for instance the C-seep method. Process models play a significant role in establishing the necessary baseline statistics (Blackford et al., 2018).

Transport models play important role in predicting the spatial and temporal signal of a tracer discharge to the water column and, thus, adequate current statistics will allow for better footprint predictions. Depending on the data available, it might be beneficial, however more computationally costly, to use three dimensional version of the advection–diffusion model. In addition, the footprints could be produced accounting for seasonality, the measuring frequency, and other factors and events, e.g., storm passages and fresh water run-off. Data from in-situ release experiments, e.g., QICS and STEMM-CCS (Blackford et al., 2014, 2018) are very useful for the required validation and quality assessment of these models.

Table 3

Comparison of the solutions with different cost functions, for the high seepage rate case. The value in boldface in rows 3–6 corresponds to the optimal solution in the sense of the cost function type indicated in the corresponding row. The last row reflects the percentage of the 10 000 random leaks missed by the sensors for each of the solutions.

Pointwise cost function Solution	Min. number $\gamma(x) = 1$ Fig. 6(a)–(c)	Max. overdetection Eq. (4) Fig. 6(b)	Max. probability Eq. (5) Fig. 6(c)	Min. op.cost Eq. (17) Fig. 6(d)
Number of sensors ( $n$ )	7	7	7	12
Overdetection $N_{over}$	99	<b>173</b>	146	117
Prob. overdetection $P_{over}$	0.51	1.06	<b>1.20</b>	0.53
Operational Cost ( $\Gamma_{op.cost}$ )	886.58	1144.60	1117.03	<b>456.75</b>
Missed leaks (%)	1.45	2.52	2.48	2.34



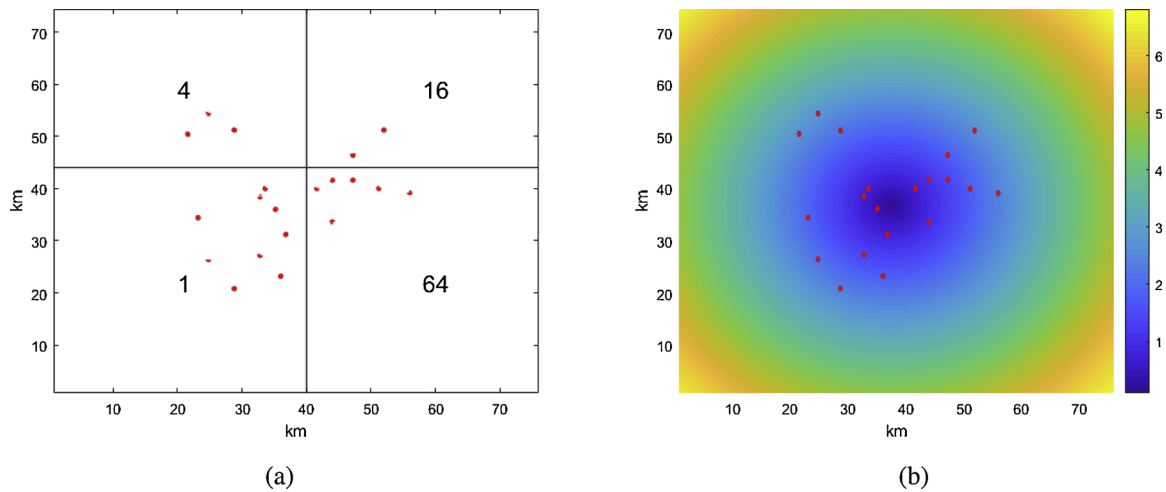


Fig. 7. (a) The probabilities  $p_i$  of a leak being active scaled by 8150 used in Eq. (17), and (b) the pointwise cost function equation (17).

Generally speaking, there might be several solutions to the optimization problem that corresponds to the minimal number of sensors, see, e.g., Fig. 6(a)–(c). Thus, it could be useful to design additional cost functions that would allow to select one solution that, at the same time, also optimizes this cost, e.g., see Fig. 6(b) and (c). This could be easily

done within the same mathematical framework. Costs associated with sensors placement and maintenance can be issues entering the cost function and, hence, in designing the monitoring program. These costs should be balanced with the needed confirmation that a leak will be detected and the imposed expenses caused by false alarms, i.e., the

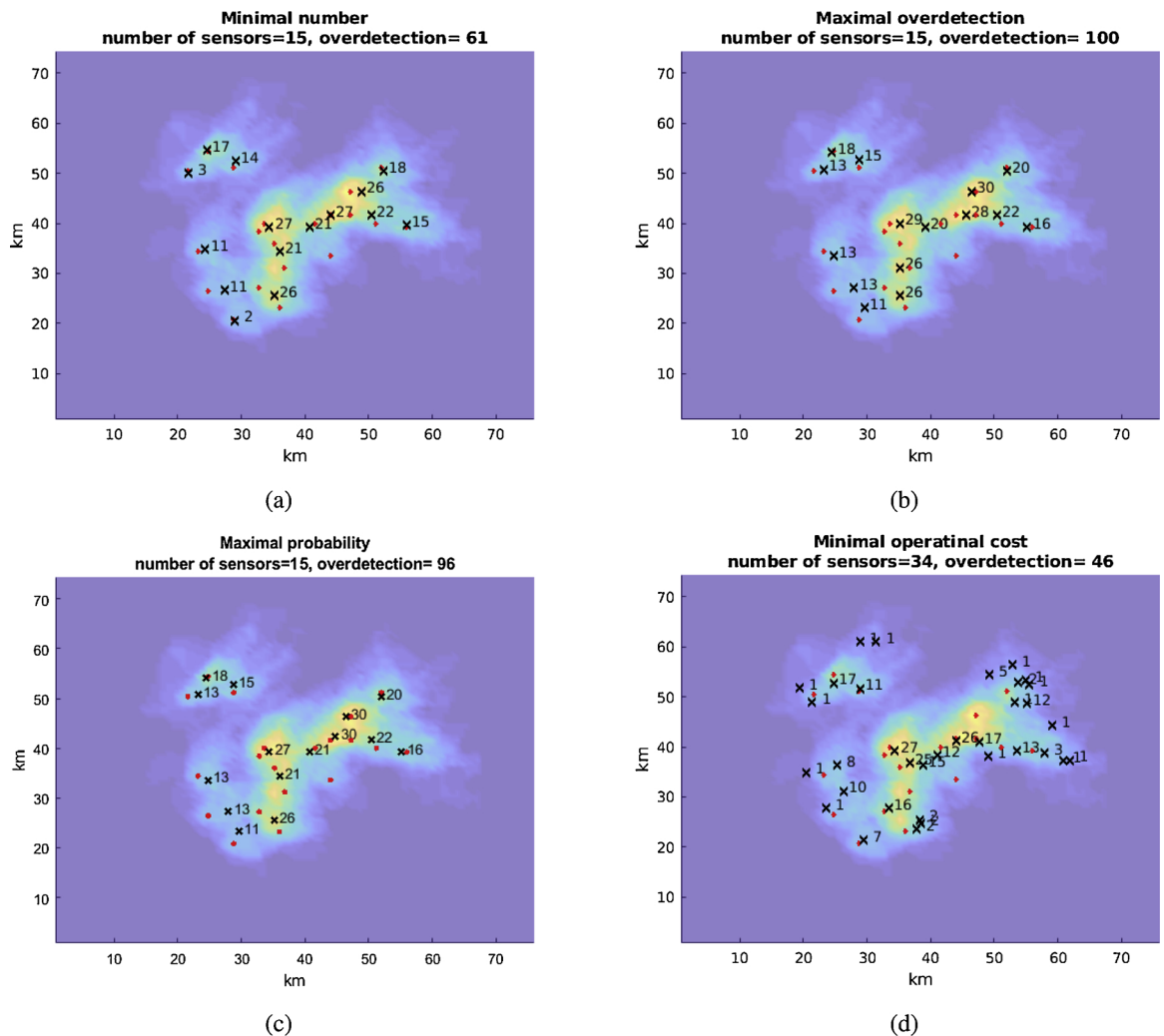


Fig. 8. Optimal sensor positions (black crosses) with the number of detected leaks for different choice of  $\Gamma$ . In particular, we used  $\gamma(x) = 1$  in (a),  $\gamma(x)$  given by Eq. (4) in (b), by Eq. (5) in (c) and by Eq. (17) in (d). For details see Table 4.

**Table 4**

Comparison of the solutions with different cost functions, for the low seepage rate case. The value in boldface in rows 3–6 corresponds to the optimal solution in the sense of the cost function type indicated in the corresponding row. The last row reflects the percentage of the 10 000 random leaks missed by the sensors for each of the solutions.

Pointwise cost function Solution	Min. number $\gamma(x) = 1$ Fig. 8(a)-(c)	Max. overdetection Eq. (4) Fig. 8(b)	Max. probability Eq. (5) Fig. 8(c)	Min. op.cost Eq. (17) Fig. 8(d)
Number of sensors ( $n$ )	<b>15</b>	<b>15</b>	<b>15</b>	34
Overdetection $N_{\text{over}}$	61	<b>100</b>	96	46
Prob. overdetection $P_{\text{over}}$	0.58	0.59	<b>0.72</b>	0.30
Operational Cost ( $\Gamma_{\text{op.cost}}$ )	703.24	836.73	811.68	<b>495.34</b>
Missed leaks (%)	4.63	4.83	5.55	4.74

probabilities of false positives and negatives (Alendal et al., 2017). It is possible to include these considerations into the cost function for the optimization problem. When dealing with fine grids, the cost function as in Eq. (2) could be of value, as it would secure the same detection outcome if a sensor is moving within the equivalence class.

Obviously, larger area of the footprints will lead to the smaller number of sensors needed, see Figs. 6 and 8. Thus, adding a tracer to the injected CO<sub>2</sub> may be cost efficient. In addition, footprints could be produced in a more sophisticated manner using, for example, machine learning methods (Gundersen et al., 2018).

It is possible to combine different instrumentations, e.g., acoustics, chemical and images, when designing footprints. The complexity of the problem however may increase drastically with the number of footprints and the grid size. Thus, the matrix  $\tilde{V}$  in Eq. (7) could be further simplified by removing some columns corresponding to sets that do not have desired properties, i.e., the sets with too small equivalence class.

Combining fixed measurements with moving platforms are not as straight forward and it would require some effort to establish routines for designing such monitoring platforms. One way would be to design the fixed platforms first, and then add moving platforms to increase our abilities to detect seeps in areas in which the fixed coverage is limited.

As mentioned initially, an important factor when designing the monitoring program is our ability to justify that a seep will be detected and to assist in communicating with stakeholders, governmental bodies and public at large. Collaborating with other offshore operators or activities through data sharing or collaborative surveys will be a win-win situation. The monitoring program can take advantage of existing infrastructure and, as a spin off, the storage project contribute to sustainable management of our oceans.

#### Authors' contribution

Anna Oleynik: conceptualization, methodology, data curation, software, formal analysis, writing – original draft, visualization. Maribel I. García-Ibáñez: methodology, software, writing – reviewing and editing. Nello Blaser: conceptualization, formal analysis, writing – reviewing and editing. Abdirahman Omar: methodology, data curation, writing – reviewing and editing. Guttorm Alendal: supervision, writing – reviewing and editing.

#### Conflict of interest

None declared.

#### Acknowledgement

This work has received funding from the Research Council of Norway, through the CLIMIT program (project 254711, BayMode) and the European Union Horizon 2020 research and innovation program under grant agreement 654462, STEMM-CCS, and the ACT programme (Accelerating CCS Technologies Project No 294766). Nello Blaser was

supported by the Research Council of Norway through grant 248840, dCod 1.0.

#### References

- Alendal, G., 2017. Cost efficient environmental survey paths for detecting continuous tracer discharges. *J. Geophys. Res.:Oceans* 122, 5458–5467.
- Alendal, G., Berntsen, J., Engum, E., Furnes, G.K., Kleiven, G., Eide, L.I., 2005. Influence from 'Ocean Weather' on near seabed currents and events at Ormen Lange. *Mar. Petrol. Geol.* 22, 21–31.
- Alendal, G., Blackford, J., Chen, B., Avlesen, H., Omar, A., 2017. Using Bayes theorem to quantify and reduce uncertainties when monitoring varying marine environments for indications of a leak. *Energy Proc.* 114, 3607–3612.
- Alendal, G., Drange, H., 2001. Two-phase, near-field modeling of purposefully released CO<sub>2</sub> in the ocean. *J. Geophys. Res.* 106.
- Ali, A., Frøysa, H.G., Avlesen, H., Alendal, G., 2016. Simulating spatial and temporal varying CO<sub>2</sub> signals from sources at the seafloor to help designing risk-based monitoring programs. *J. Geophys. Res.: Oceans* 121, 745–757. <https://doi.org/10.1002/2015JC011198>.
- Anderson, L.A., Sarmiento, J.L., 1994. Redfield ratios of remineralization determined by nutrient data analysis. *Glob. Biogeochem. Cycles* 8, 65–80. <https://doi.org/10.1029/93GB03318>.
- Artioli, Y., Blackford, J.C., Butenschön, M., Holt, J.T., Wakelin, S.L., Thomas, H., Borges, A.V., Allen, J.I., 2012. The carbonate system in the North Sea: sensitivity and model validation. *J. Marine Syst.* 102–104, 1–13.
- Blackford, J., Alendal, G., Artioli, Y., Avlesen, H., Cazenave, P., Chen, B., Dale, A., Dewar, M., Gundersen, K., Haeckel, M., Khajepor, S., Lessin, G., Oleynik, A., Saleem, U., García-Ibáñez, M., Omar, A., 2018. Ensuring efficient and robust offshore storage? The role of marine system modelling. *Proceedings for GHGT14, Melbourne, October 2018*.
- Blackford, J., Artioli, Y., Clark, J., de Mora, L., 2017. Monitoring of offshore geological carbon storage integrity: implications of natural variability in the marine system and the assessment of anomaly detection criteria. *Int. J. Greenhouse Gas Control* 64, 99–112.
- Blackford, J., Bull, J.M., Cevatoglu, M., Connelly, D., Hauton, C., James, R.H., Lichtschlag, A., Stahl, H., Widdicombe, S., Wright, I.C., 2015. Marine baseline and monitoring strategies for carbon dioxide capture and storage (CCS). *Int. J. Greenhouse Gas Control* 38, 221–229 CCS and the Marine Environment.
- Blackford, J., Jones, N., Proctor, R., Holt, J., 2008. Regional scale impacts of distinct CO<sub>2</sub> additions in the North Sea. *Mar. Poll. Bull.* 56, 1461–1468.
- Blackford, J., Stahl, H., Bull, J.M., Bergès, B.J.P., Cevatoglu, M., Lichtschlag, A., Connelly, D., James, R.H., Kita, J., Long, D., Naylor, M., Shitashima, K., Smith, D., Taylor, P., Wright, I., Akhurst, M., Chen, B., Gernon, T.M., Hauton, C., Hayashi, M., Kaieda, H., Leighton, T.G., Sato, T., Sayer, M.D.J., Suzumura, M., Tait, K., Vardy, M.E., White, P.R., Widdicombe, S., 2014. Detection and impacts of leakage from sub-seafloor deep geological carbon dioxide storage. *Nat. Clim. Change* 4, 1011–1016.
- Blackford, J.C., Widdicombe, S., Lowe, D., Chen, B., 2010. Environmental risks and performance assessment of carbon dioxide (CO<sub>2</sub>) leakage in marine ecosystems. *Developments and Innovation in Carbon Dioxide (CO<sub>2</sub>) Capture and Storage Technology, Vol. 2 – Carbon Dioxide (CO<sub>2</sub>) Storage and Utilisation* 344–373.
- Botnen, H., Omar, A., Thorseth, I., Johannessen, T., Alendal, G., 2015. The effect of submarine CO<sub>2</sub> vents on seawater: implications for detection of subsea Carbon sequestration leakage. *Limnol. Oceanogr.* 60.
- Boyd, A.D., Liu, Y., Stephens, J.C., Wilson, E.J., Pollak, M., Peterson, T.R., Einsiedel, E., Meadowcroft, J., 2013. Controversy in technology innovation: contrasting media and expert risk perceptions of the alleged leakage at the Weyburn carbon dioxide storage demonstration project. *Int. J. Greenhouse Gas Control* 14, 259–269.
- Brewer, P.G., Chen, B., Warzinski, R., Baggeroer, A., Peltzer, E.T., Dunk, R.M., Walz, P., 2006. Three-dimensional acoustic monitoring and modeling of a deep-sea CO<sub>2</sub> droplet cloud. *Geophys. Res. Lett.* 33, 5.
- Buhl-Mortensen, L., Buhl-Mortensen, P., Dolan, M.F.J., Holte, B., 2014. The MAREANO programme – a full coverage mapping of the Norwegian off-shore benthic environment and fauna. *Mar. Biol. Res.* 11, 4–17.
- Dixon, T., McCoy, S.T., Havercroft, I., 2015. Legal and regulatory developments on CCS. *Int. J. Greenhouse Gas Control* 40, 431–448.

- Domínguez-Tejo, E., Metternicht, G., Johnston, E., Hedge, L., 2016. Marine Spatial Planning advancing the Ecosystem-Based Approach to coastal zone management: a review. *Mar. Policy* 72, 115–130.
- Drange, H., Alendal, G., Johannessen, O., 2001. Ocean release of fossil fuel CO<sub>2</sub>: a case study. *Geophys. Res. Lett.* 28, 2637–2640.
- Friis, K., Kärtzinger, A., Wallace, D.W.R., 2003. The salinity normalization of marine inorganic carbon chemistry data. *Geophys. Res. Lett.* 30. <https://doi.org/10.1029/2002GL015898>.
- García, J.H., Torvanger, A., 2019. Carbon leakage from geological storage sites: implications for carbon trading. *Energy Policy* 127, 320–329.
- Greenwood, J., Craig, P., Hardman-Mountford, N., 2015. Coastal monitoring strategy for geochemical detection of fugitive CO<sub>2</sub> seeps from the seabed. *Int. J. Greenhouse Gas Control* 39, 74–78.
- Gruber, N., Sarmiento, J.L., Stocker, T.F., 1996. An improved method for detecting anthropogenic CO<sub>2</sub> in the oceans. *Global Biogeochem. Cycles* 10, 809–837. <https://doi.org/10.1029/96GB01608>.
- Gundersen, K., Oleynik, A., Alendal, G., Skaug, H., Avlesen, H., Berntsen, J., Blaser, N., Blackford, J., Cazenave, P., 2018. Ensuring efficient and robust offshore storage? Use of models and machine learning techniques to design leak detection monitoring. In: *Proceedings for GHGT14*. Melbourne, October 2018.
- Halland, E., Riis, F., Magnus, C., Johansen, W., Tappel, I., Gjeldvik, I., Solbakk, T., Pham, V., 2013. CO<sub>2</sub> storage atlas of the Norwegian part of the North Sea. *Energy Proc.* 37, 4919–4926.
- Halpern, B.S., Longo, C., Hardy, D., McLeod, K.L., Samhuri, J.F., Katona, S.K., Kleisner, K., Lester, S.E., O'Leary, J., Ranelletti, M., Rosenberg, A.A., Scarborough, C., Selig, E.R., Best, B.D., Brumbaugh, D.R., Chapin, F.S., Crowder, L.B., Daly, K.L., Doney, S.C., Elfes, C., Fogarty, M.J., Gaines, S.D., Jacobsen, K.I., Karrer, L.B., Leslie, H.M., Neeley, E., Pauly, D., Polasky, S., Ris, B., Martin, K.S., Stone, G.S., Sumaila, U.R., Zeller, D., 2012. An index to assess the health and benefits of the global ocean. *Nature* 488, 615–620.
- Harding, B., Tremblay, C., Cousineau, D., 2014. Standard errors: a review and evaluation of standard error estimators using Monte Carlo simulations. *Quant. Methods Psychol.* 10, 107–123.
- Haugan, P.M., Joos, F., 2004. Metrics to assess the mitigation of global warming by carbon capture and storage in the ocean and in geological reservoirs. *Geophys. Res. Lett.* 31. <https://doi.org/10.1029/2004GL020295>.
- Hochbaum, D., 1997. *Approximation Algorithms for NP-hard Problems*. PWS Series in Computer Science. PWS Publishing Company.
- Hvidevold, H.K., Alendal, G., Johannessen, T., Ali, A., 2016. Survey strategies to quantify and optimize detecting probability of a CO<sub>2</sub> seep in a varying marine environment. *Environ. Modell. Softw.* 83, 303–309.
- Hvidevold, H.K., Alendal, G., Johannessen, T., Ali, A., Mannseth, T., Avlesen, H., 2015. Layout of CCS monitoring infrastructure with highest probability of detecting a footprint of a CO<sub>2</sub> leak in a varying marine environment. *Int. J. Greenhouse Gas Control* 37, 274–279.
- Jenkins, C., Chadwick, A., Hovorka, S.D., 2015. The state of the art in monitoring and verification ten years on. *Int. J. Greenhouse Gas Control* 40, 312–349 Special Issue commemorating the 10th year anniversary of the publication of the Intergovernmental Panel on Climate Change Special Report on CO<sub>2</sub> Capture and Storage.
- Jones, D., Beaubien, S., Blackford, J., Foekema, E., Lions, J., Vittor, C.D., West, J., Widdicombe, S., Hauton, C., Queries, A., 2015. Developments since 2005 in understanding potential environmental impacts of CO<sub>2</sub> leakage from geological storage. *Int. J. Greenhouse Gas Control* 40, 350–377.
- Mabon, L., Shackley, S., Bower-Bir, N., 2014. Perceptions of sub-seabed carbon dioxide storage in Scotland and implications for policy: a qualitative study. *Mar. Policy* 45, 9–15.
- Maeda, Y., Shitashima, K., Sakamoto, A., 2015. Mapping observations using AUV and numerical simulations of leaked CO<sub>2</sub> diffusion in sub-seabed CO<sub>2</sub> release experiment at Ardmucknish Bay. *Int. J. Greenhouse Gas Control* 38, 143–152.
- Metz, B., Davidson, O., Coninck, H.C.d., Loos, M., Meyer, L.A., 2005. *Special Report on Carbon Dioxide Capture and Storage*. Miscellaneous.
- Noble, R.R.P., Stalker, L., Wakelin, S.A., Pejčić, B., Leybourne, M.I., Hortle, A.L., Michael, K., 2012. Biological monitoring for carbon capture and storage – a review and potential future developments. *Int. J. Greenhouse Gas Control* 10.
- Oldenbure, C.M., Lewicki, J.L., 2006. On leakage and seepage of CO<sub>2</sub> from geologic storage sites into surface water. *Environ. Geol.* 50, 691–705.
- Olsen, A., Key, R.M., van Heuven, S., Lauvset, S.K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F.F., Suzuki, T., 2016. The global ocean data analysis project version 2 (glodapv2) – an internally consistent data product for the world ocean. *Earth Syst. Sci. Data* 8, 297–323.
- Omar, A., Ibanez Garcia, M., Alendal, G., 2018. The stoichiometric CSEEP method as a tool to distinguish CO<sub>2</sub> seepage signal from the natural variability. In: *Proceedings for GHGT14*. Melbourne, October 2018.
- Omar, A.M., Olsen, A., Johannessen, T., Hoppema, M., Thomas, H., Borges, A.V., 2010. Spatiotemporal variations of fCO<sub>2</sub> in the north sea. *Ocean Sci* 6, 77–89.
- Redfield, A., 1934. On the Proportions of Organic Derivatives in Sea Water and Their Relation to the Composition of Plankton. University Press of Liverpool.
- Romanak, K., Sherk, G.W., Hovorka, S., Yang, C., 2013. Assessment of alleged CO<sub>2</sub> leakage at the Kerr farm using a simple process-based soil gas technique: implications for carbon capture, utilization, and storage (CCUS) monitoring. *Energy Proc.* 37, 4242–4248.
- Romanak, K.D., Bennett, P.C., Yang, C., Hovorka, S.D., 2012. Process-based approach to CO<sub>2</sub> leakage detection by vadose zone gas monitoring at geologic CO<sub>2</sub> storage sites. *Geophys. Res. Lett.* 39, 114.
- Schrijver, A., 1986. *Theory of Linear and Integer Programming*. John Wiley & Sons, Inc., New York, NY.
- Smith, L.M., Barth, J.A., Kelley, D.S., Plueddemann, A., Rodero, I., Ulses, G.A., Vardaro, M.F., Weller, R., 2018. The ocean observatories initiative. *Oceanography* 31, 16–35.
- Torvanger, A., Grimstad, A.A., Lindeberg, E., Rive, N., Rypdal, K., Skeie, R.B., Fuglestvedt, J., Tollefsen, P., 2012. Quality of geological CO<sub>2</sub> storage to avoid jeopardizing climate targets. *Clim. Change* 114, 245–260.
- Uchimoto, K., Nishimura, M., Kita, J., Xue, Z., 2018. Detecting CO<sub>2</sub> leakage at offshore storage sites using the covariance between the partial pressure of CO<sub>2</sub> and the saturation of dissolved oxygen in seawater. *Int. J. Greenhouse Gas Control* 72, 130–137.
- Vermeul, V.R., Amonette, J.E., Strickland, C.E., Williams, M.D., Bonneville, A., 2016. An overview of the monitoring program design for the FutureGen 2.0 CO<sub>2</sub> storage site. *Int. J. Greenhouse Gas Control* 51, 193–206.
- Vielstädte, L., Karstens, J., Haeckel, M., Schmidt, M., Linke, P., Reimann, S., Liebetrau, V., McGinnis, D.F., Wallmann, K., 2015. Quantification of methane emissions at abandoned gas wells in the Central North Sea. *Mar. Petrol. Geol.*
- Wegener, G., Shovitri, M., Knittel, K., Niemann, H., Hovland, M., Boetius, A., 2008. Biogeochemical processes and microbial diversity of the Gullfaks and Tommeliten methane seeps (Northern North Sea). *Biogeochem. Discuss.* 5, 971–1015.
- Yang, Y.M., Small, M.J., Junker, B., Bromhal, G.S., Strazisar, B., Wells, A., 2011. Bayesian hierarchical models for soil CO<sub>2</sub> flux and leak detection at geologic sequestration sites. *Environ. Earth Sci.* 64, 787–798.