Cost efficient environmental survey paths for detecting continuous tracer discharges

G. Alendal

1Department of Mathematics, University of Bergen, Bergen, Norway

Abstract Designing monitoring programs for detecting potential tracer discharges from unknown locations is challenging. The high variability of the environment may camouflage the anticipated anisotropic signal from a discharge, and there are a number of discharge scenarios. Monitoring operations may also be costly, constraining the number of measurements taken. By assuming that a discharge is active, and a prior belief on the most likely seep location, a method that uses Bayes’ theorem combined with discharge footprint predictions is used to update the probability map. Measurement locations with highest reduction in the overall probability of a discharge to be active can be identified. The relative cost between reallocating and measurements can be taken into account. Three different strategies are suggested to enable cost efficient paths for autonomous vessels.

Plain Language Summary Designing monitoring programs for detecting potential tracer discharges from unknown locations is challenging. The high variability of the environment may camouflage the anticipated anisotropic signal from a discharge, and there are a number of discharge scenarios. Monitoring operations may also be costly, constraining the number of measurements taken. By assuming that a discharge is active, and a prior belief on the most likely seep location, a method that uses Bayes’ theorem combined with discharge footprint predictions is used to update the probability map. Measurement locations with highest reduction in the overall probability of a discharge to be active can be identified. The relative cost between reallocating and measurements can be taken into account. Three different strategies are suggested to enable cost efficient paths for autonomous vessels.

1. Introduction

Monitoring with the purpose of detecting tracer discharges from an unknown location is challenging in many aspects. For instance, projects involved in subsurface gas storage, either to store energy to dampen fluctuations in renewable energy sources or CO₂ storage to mitigate the burden of elevated atmospheric concentrations, will have to detect any leaks through to the atmosphere or the ocean [Bauer et al., 2013; Blackford et al., 2015; Jones et al., 2015]. Further, with the number of fish farms present in Norwegian fjords, there are many potential sources for organic waste [Ali et al., 2011] and other contaminants that might have adverse effects on marine ecosystems [Hylland et al., 2017].

A monitoring program has three levels of modus operandi: (1) anomaly detection modus, (2) confirmation and localization modus, and (3) flux quantification and abatement modus. All modes will have different needs with regard to instrumentation and data. For example, location mode will require information about real time current/wind direction and speed to be able to move upstream toward an increasing signal.

In the detection phase, in which the monitoring program looks for environmental anomalies, baseline statistics, and predictions of discharge characteristics determine monitoring design. Three main questions will be: (1) where will a discharge most likely occur, (2) how will a discharge trail materialize, and (3) will it be possible to distinguish the signal from the background variability? In addition, what degree of anomaly will mobilize the more expensive confirmation and localization step? There is a need to balance between the need for confirmation and the cost of false alarms.

The aim here is to suggest three strategies to define survey paths for detecting anomalies using an autonomous vessel, such as an AUV in the ocean or an UAV in the atmosphere, capable of taking instant
measurements [Maeda et al., 2015]. The relative cost, e.g., use of limited battery capacity, between taking measurements and moving to another location can be taken into account. Bayes’ theorem lays the foundation for defining these paths, and the methodology is exemplified using CO₂ gas that seeps through the seafloor.

Environmental changes, e.g., changes in bottom fauna or in the pelagic ecosystem [Wegener et al., 2008; Blackford et al., 2010], detection of bubbles from ship sonars [Brewer et al., 2006; Noble et al., 2012], or elevated concentration of dissolved gases [Alendal and Drange, 2001; Drange et al., 2001; Vielstädte et al., 2015], can be used as indicators of marine gas releases. Here elevated CO₂ concentration is used as an indicator.

CO₂ is a natural occurring tracer in seawater, with seasonal concentration and variability dependency. Botnen et al. [2015] demonstrated a stoichiometric approach for detection of small CO₂ concentrations that might stem from seeps, lowering the concentration threshold for a signal to become statistically significant. This approach is used here to define a detection limit.

Another challenge is the high variability in ocean dynamics. Tides change current directions, wind alters the amount of mixing, and local topography change local current conditions [Alendal et al., 2005]. Footprints of leaks are thus varying and highly anisotropic signals, depending strongly on the local oceanic and atmospheric conditions [Ali et al., 2016]. Hvidevold et al. [2015, 2016] presented a procedure for optimizing placement of fixed sensors on the seafloor that accounts for footprint anisotropy. The same tracer footprint predictions, consisting of time series in an array of 51 × 51 grid points around the seep location from the GCM simulations presented in Ali et al. [2016], are used here. One challenge is the extent of the area being monitored and the unknown location of a potential discharge.

Bayes’ theorem has had a renaissance in the recent years, even though the technique has been used in many contexts over the years without being named as Bayesian [Mcgrayne, 2011]. The theorem offers an approach to update our belief about the presence of a discharge during a series of unsuccessful detections in an area, and to identify most likely locations should a leak be indicated.

2. Bayesian Search Strategies

The classical search algorithm when searching for and locating missing objects based on Bayes theorem reads [Breivik et al., 2012; Stone et al., 2014]:

\[
p_{\text{post}} = p_{\text{prior}} \frac{1 - q}{1 - p_{\text{prior}} q} < p_{\text{prior}}. \tag{1}
\]

A failed attempt in a location will reduce the probability of the object being there, and increase the probability in all other locations, including the probability of the object not being in the search area (A) at all. To shorten notation, \( p(\text{leak}) = p_{\text{prior}} \), \( p(\text{found} | \text{leak}) = q \), and \( p(\text{leak} | \sim \text{found}) = p_{\text{post}} \) have been introduced.

A tracer discharge in the search area can be detected remotely, hence a failed attempt will also reduce the probability of a discharge to be in neighboring locations. Denoting the probability of detecting a discharge at a location \( x \) by a measurement taken at \( y \) as \( q(x, y) \) the posterior probability field after an unsuccessful measurement at \( y \) will be

\[
p_{\text{post}}(x, y) = p_{\text{prior}}(x) \frac{1 - q(x, y)}{1 - \int_A p_{\text{prior}}(x') q(x', y) dx'}. \tag{2}
\]

The resulting change in overall belief, i.e., the probability of the discharge being present in the search area (A) from a measurement at \( y \in A \) failing to detect a discharge, will be

\[
\Delta p(y) = p_{\text{post}} - p_{\text{prior}} = \int_A p_{\text{prior}}(x) \left( \frac{1 - q(x, y)}{1 - \int_A p_{\text{prior}}(x') q(x', y) dx'} - 1 \right) dx. \tag{3}
\]

An initial overall prior belief at the start of the search is given by
in which \( p_{\text{prior}}(x) \) represents the probability of the leak to be at \( x \). Hence, the achievable \( \Delta p(y) \) to be obtained is dependent on our initial prior belief. Since the three strategies select the location with highest reduction, this dependency does alter the initial path chosen. This is demonstrated later.

The three search strategies based for selecting the next measurement location are:

1. **Highest**:
   The most straightforward method is to select the next location for measurement, \( y \), that will maximize \( |\Delta p(y)| \), i.e.,
   \[
   y = \arg \max_{y \in A} |\Delta p(y)|.
   \] (5)

2. **Distance**:
   It will take some time to move to the next location, so the location of highest \( |\Delta p(y)| \) might not give the highest reduction in \( p \) for a given time interval, i.e., reduction of probability of the leak with respect to time.
   
   Assuming that the vessel moves with a constant speed (\( V \)), and that the distance between present location \( z \) and a location \( y \) is \( d(y, z) \), then the changes in \( p \) in a given time interval \( \Delta t \) will be
   \[
   \frac{\Delta p(y)}{\Delta t} = \frac{\Delta p(y)}{d(y, z)} V.
   \] (6)
   Notice that in the limit \( \Delta t \to 0 \) this represents \( dp/dt \). The goal is to reduce probability as fast as possible, so the next location that will be most efficient with respect to time, \( y \), given that the vessel is at \( z \), will then be
   \[
   y(z) = \arg \max_{y \in A} \frac{|\Delta p(y)|}{d(y, z)}.
   \] (7)

3. **Continuous**:
   The third approach is to select next position similar to the previous strategy, but measurements are taken as the sensor moves toward it:
   \[
   y_n(z) = z + \rho_n(y - z), \quad y = \arg \max_{y \in A} \frac{|\Delta p(y)|}{d(y, z)}, \quad n = 1, 2, \ldots, N.
   \] (8)
   where \( \rho_n = n/N \leq 1 \), and \( N \) is the number of measurements to take while moving toward \( y \). If the probability field is not updated after each of these under way measurements this strategy equals the Distance strategy in the limit \( N = 1 \), i.e., choose to jump directly to \( y \) in one step. However, if the target location \( y \) is updated after each measurement, the direction might change under way and the sensor starts to move toward another location.

### 3. An Example Problem

To illustrate the method, and the three strategies, an artificial gas seep problem has been created. In an area of \( 80 \times 80 \) km\(^2 \) in extent, 15 possible high-risk discharge locations have been randomly placed. These locations might represent perforating wells, each assumed to have equally high probabilities of being the leak location. To account for the possibility of gas migrating horizontally as it moves toward the seafloor, the probability reduces exponentially toward a general background probability that is 100 times less than in the center. Initially the probability field is normalized with an initial prior probability of an ongoing discharge

\[
\int_A p(x) dx = p_{\text{prior}}.
\] (9)

i.e., the starting prior overall belief that a leak is active. The resulting prior probability field is shown in Figure 1 for \( p_{\text{prior}} = 0.99 \).

In a real case, this map will have to be based on a thorough site characterization, taking into account, e.g., how deep the wells penetrate the geology, presence of gas pockets in the subsurface, and other geological structures being present in the area.
Ali et al. [2016] showed that the footprints of discharges are highly dependent on seep locations and this will have to be accounted for in real case scenarios. Given the other simplifications taken in this example, the footprint is assumed to be independent of discharge location. With this assumption, $q(x, y) = q(x - y)$, i.e., the probability of detecting a seep at $x$ by a measurement taken at $y$ is only dependent on the relative distance vector. Here an anisotropic footprint prediction of a CO$_2$ seep, taken from Ali et al. [2016], is used for $q(x - y)$.

Concentration time series in a $51 \times 51$ grid surrounding a seep with flux 1 kg/s simulated by an 800 m resolution North Sea setup has been collected. Time series for 3 by 3 grid cells closest to the seep location is shown in Figure 2. The relative time the CO$_2$ concentration stays above a detection threshold of $c_t = 5 \mu$mol/kg$_{SW}$, based on a stoichiometric approach [Botnen et al., 2015], is shown in the left plot of Figure 3. This field represent the probability of detecting a leak at $y = (x, y)$ if the leak is in $(0, 0)$, i.e., the detectable region in the neighborhood of a seep.

To achieve $q(x - y)$, i.e., the area monitored from a location, the probability field will have to be inverted through a $180^\circ$ rotation [Ali et al., 2016]. The location with highest probability will now represent the location of the measurement, and the probability field represents the probability of detecting a leak from that position, as shown in the right plot of Figure 3. The probability of detecting a seep from a measurement location, $y$ is then the sum of these probabilities weighted by the probabilities of the individual locations to be the location of the discharge.

To illustrate the strategies, assume that the sensor is located at the red cross in the two plots of Figure 3. Using the probability field in Figure 1 as $p_{prior}(x)$ and the right plot in Figure 3 as $q(x - y)$ in equation (3) gives the resulting fields of $\Delta p(y)$ as shown in the left plot of Figure 4. The point of highest reduction in $\Delta p$ is in the red circle. It is independent of where the sensor is presently.

When accounting for the distance, i.e., $\Delta p(y)/d(y, z)$ in the Distance method, the location of the first measurement is shifted to the red circle in the right plot of Figure 3. In the Continuous strategy the sensor will start moving toward the red circle in the right plot taking measurements while traveling.

In the following calculations, all methods select the red circle in the left plot of Figure 4 as the location of the first measurement. Hence, the cost of traveling to the first measurement location is not accounted for in any strategies. This removes any dependence of the resulting path taken on the starting location.

Figure 5 shows measurement locations and paths taken for the different strategies using initial $p_{prior} = 0.99$. The left plot shows the unstructured path taken when using Highest method (equation (5)), to choose the next measurement point. The sensor will move to the next point with the highest reduction in $|\Delta p(x)|$ without considering the distance or time it takes to move there.

The middle plot in Figure 5 shows that the path becomes more organized when the distance is taken into account in the Distance method (equation (7)). The sensor tends to make new measurements close to the present locations, but occasionally there are discontinuities in the paths. These occur when the probability of the seep to be in the local area has been reduced to a level that makes it beneficial to move over a distance for the next measurement. As expected, the Continuous strategy results in a path that is continuous, as shown in the right plot.
Figure 6 shows the remaining overall probability of a discharge being present in the area as function of number of measurements taken (left plot) and distance traveled (right plot). If the number of measurements is the limiting factor, choosing the next measurement in the location that maximizes $D_p$, labeled $\text{Highest}$. The $\text{Continuous}$ strategy will then collect measurements that do not reduce the probability significantly.

Figure 2. Time series of the CO$_2$ concentration along the seabed at the (middle plot) leak grid cell and the eight neighboring grid cells. The red line indicates the threshold value $c_t = 5 \, \mu \text{mol/kg}_{\text{sw}}$ from Botnen et al. [2015]. Figure taken from Ali et al. [2016].

Figure 3. (left plot) The relative number of points in the time series in which the concentration is above detection threshold, hence the probability of drawing a random sample from the time series that will be above the threshold. (right plot) The resulting probability of the concentration to be above the threshold in the origin, i.e., in $(0, 0)$, if a leak is present in $(x, y)$. The latter is achieved through a 180° rotation of the probability field in the left plot.
If distance traveled comes with the highest cost, the situation is the opposite, as shown in the right plot of Figure 6. It is evident that the distance traveled is considerably larger when always selecting the location with maximum $|\Delta p|$, i.e., the Highest approach. The Continuous approach is now the most cost efficient strategy.

The example used here is related to chemical sensing in a marine environment using an autonomous vessel. In such a case, propulsion and measurements utilize the same energy source, the onboard batteries. A simple cost calculation for the different strategies, labeled $i$, might be

$$C_i = c_m (#m_i + \alpha d_i),$$

where $c_m$ is the cost of taking one measurement, $#m_i$ is the number of measurements taken, $d_i$ is the distance moved, and $\alpha = c_m / c_p$ is the proportionate cost factor between moving and taking a measurement. The resulting relative cost is shown in Figure 7, with the relative cost, $\alpha$, on a logarithmic scale along the $x$ axis and the remaining probability of an ongoing discharge along the $y$ axis. The contours are lines of constant $C_i / C_j$, i.e., the relative cost between the two methods.

The left plot in Figure 7 shows the relative cost between Distance and Highest. In the shaded area, the Distance approach will be more expensive than using the Highest strategy. Taking a measurement will have to be $\sim$300 times more energy consuming than moving for the Highest approach to be selected. The middle plot of Figure 7 shows the costs of Continuous relative to Highest strategy. The shaded area is where the Continuous approach will be the more expensive than Highest. This occurs when measuring costs are $\sim$166 times more costly than reallocating. The right plot shows the contours of cost of Continuous relative to Distance. In the shaded area, it will be more costly to perform measurements along transects while moving toward the next “hot spot.”
The number of measurements needed, and hence distance traveled, to bring $p_{post}$ down to 0.1 is reduced with reduced initial $p_{prior}$, as shown in Figure 8. Hence, the prior is important when quantifying the overall probability of a leak in the area. However, the initial part of the paths generated will not depend on the choice of initial $p_{prior}$.

Figure 9 shows that the first 50 measurement locations for the three strategies for initial $p_{prior}$ equal to 0.99 and 0.5 are the same. Two of the strategies, High and Distance, will cover the hot spots early, with Distance covering the local area before jumping to other locations. Both Distance and Continuous use the first eight measurements to cover the three wells close to the starting point, before moving toward the single well to the north, denoting north in positive $y$ direction and east in positive $x$ direction. When reaching this location, the Continuous strategy has collected measurements and route and after two measurements at the same locations in the vicinity of this well, the paths for the two strategies diverge. The Continuous measurements favor the single well to west as opposed to the double wells to the south selected by the Distance case. This demonstrates that continuous measurements might alter the paths taken.

4. Discussion

The capability of Bayes’ theorem to design search paths for detecting a tracer discharge is demonstrated here for a marine gas tracer’s discharge, using an autonomous vessel capable of doing instant measurements. The idea is that each measurement updates the probability field, and the vessel decides where to go next, i.e., a version of the greedy algorithm.

Figure 7. Contours of the relative cost, $C_i/C_j$, between the different strategies for a given target belief that a leak is ongoing along the vertical axis and the relative cost between distance traveled and measurements, $\pi$, on logarithmic scale along the $x$ axis. Notice the logarithmic scale along the $x$ axis. Contours are (left) Distance/Highest, (middle) Continuous/Highest, and (right) Continuous/Distance. Initial used is $p_{prior} = 0.99$. 
The initial path is not depending on the initial prior \( q_{\text{prior}} \) in equation (9), but will become important when quantifying the overall belief after a survey. Subsequent surveys will also benefit from the reduction in belief achieved. How frequent surveys should be executed depends on the underlying site characterization,

Figure 8. The reduction in probability of a leak for different initial \( p_{\text{prior}} \) as function of number of measurement taken (top row) and distance traveled (bottom row) for Highest (left), Distance (middle), and Continuous (right) strategies.

The initial path is not depending on the initial prior \( q_{\text{prior}} \) in equation (9), but will become important when quantifying the overall belief after a survey. Subsequent surveys will also benefit from the reduction in belief achieved. How frequent surveys should be executed depends on the underlying site characterization,

Figure 9. The overlapping paths taken for the 50 first measurements for (left) Highest, (middle) Distance, and (right) Continuous strategies for initial \( p_{\text{prior}} = 0.99 \) (circles) and \( p_{\text{prior}} = 0.5 \) (stars). The top plots show the whole area, while the bottom plots focus on the initial part of the paths.
the need for assurance, e.g., the vulnerability of the area in questions, and the cost of performing the surveys.

In other approaches where measurements take some time (e.g., water sampling) or there is a delay between sampling and results, it might be beneficial to define an optimal route during the cruise planning. The cost function to be minimized will then differ. However, model predictions of discharge footprints and prior identification of potential discharge locations can still be combined with Bayes’ theorem in solving such “traveling salesman” problems as demonstrated in Hvidevold et al. [2015]. Also, presence of obstacles and off limit areas that constrain the paths taken can be accounted for. Here algorithms developed within robotics can be worth reviewing [Karahan and Frazzoli, 2011; Hollinger and Sukhatme, 2014].

The method can also be extended to a monitoring program that combines different sensing techniques and supplement with fixed installations. Also, noncontinuous, e.g., bursting, discharges can be implemented through the footprint probability field.

The method and strategies are dependent on the initial belief of where a discharge will most probably occur, exemplified through Figure 1, and the footprint predictions as exemplified in Figure 3. The former will have to come from a thorough site characterization, identifying potential seep or discharge locations. The latter will also be highly site specific and depend on the processes that govern the environment and the distribution of the discharged tracer. Such footprints will have to rely on transport modeling, and to some extent experiments. All will have to be based on a proper environmental baseline.

It will be necessary to establish routines in the event that a measurement indicates an increased probability of an ongoing discharge. The confirmation mode will require real time current/wind directions and speed to move the sensor upstream toward an increasing signal. Bayes’ theorem can be useful in this context as well, the belief of an ongoing discharge is updated as new measurements are taken and false alarm might be concluded, should the probability show a consistent negative trend.

Acknowledgments
I will thank the anonymous reviewers for very constructive suggestions and corrections, you have definitely improved the quality of the manuscript. This work has received funding from the Research Council of Norway, through the CLIMIT program (project 254711) and the SUCCESS center for CO2 storage (grant 193825/560), as well as the European Unions Horizon 2020 research and innovation program under grant agreement 654462, STEMM-CCS. Results from the BOM simulations used in this study are available at zenodo.org with doi:10.5281/zenodo.806088.

References
Erratum

In the originally published version of this article, the website link containing the funding information was incomplete. This has since been corrected and this version may be considered the authoritative version of record.