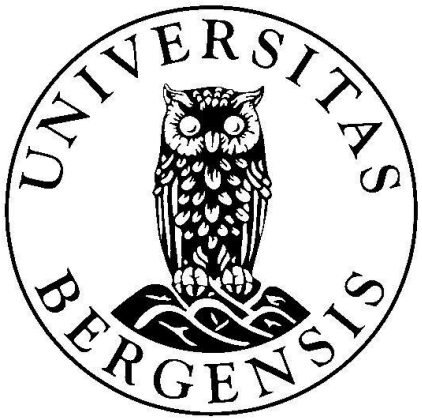
Design of a monitoring program in a varying marine environment

Master thesis in Applied and
Computational Mathematics

Håvard Guldbrandsen Frøysa



Department of Mathematics
University of Bergen
September 2015

Acknowledgements

First of all, I would like to thank my supervisor professor Guttorm Alendal for introducing me to an interesting problem and being really helpful throughout the process. You have been sincerely interested in my work and always been willing to discuss new thoughts and ideas, which I'm extremely grateful for.

I would also like to thank Alfatih Ali for providing test data and Hilde Kristine Hvidevold for allowing me to use some of her work before it was published. Thanks to my fellow students for great friendship, lunch breaks and discussions. In particular thanks to Sondre for help with parallel computing.

Finally, I would like to thank all family and friends that have supported me during my studies. Especially I would like to thank my parents for inspiring me to pursue a degree in mathematics. You have been of great support and have provided valuable insight in the academic world of mathematics.

Håvard G. Frøysa
September 2015

Give thanks to the Lord, for he is good;
his love endures forever.
Psalm 107,1

Abstract

In this thesis we develop methods for optimal design of a monitoring program for offshore geological CO_2 storage. The goal is to find the layout of fixed chemical sensors at the seafloor that maximizes the probability of detecting a leakage. Numerical simulations of leakage scenarios are used as origin to predict the regions that sensors monitor. Based on leakage scenarios, this gives the detection probability. All methods are tested using test cases. The methods could be applied to other problems involving monitoring of potential pollutants into the ocean. The main results are inclusion of spatial variability in the estimated leakage footprint and an exact inversion of the resulting footprint.

Contents

1 Introduction **1**
- 1.1 CCS as a greenhouse gas technology . . . 1
- 1.2 Monitoring of a storage site . . . 3

2 Monitoring design: Mathematical framework **7**
- 2.1 Site characteristics . . . 8
  - 2.1.1 Potential leakage locations . . . 8
  - 2.1.2 Footprint . . . 8
  - 2.1.3 Synthetic map . . . 8
- 2.2 Detection calculations . . . 9
  - 2.2.1 Detection limit . . . 9
  - 2.2.2 Detection function . . . 10
  - 2.2.3 Monitoring function . . . 11
  - 2.2.4 Detection probability . . . 11
  - 2.2.5 Optimization problem . . . 14

3 Monitoring design: Approximation of an average concentration footprint **17**
- 3.1 Detection function for average . . . 17
- 3.2 Monitoring function for average . . . 18
- 3.3 Previous work: Approximation of footprint . . . 18
- 3.4 Detection function for approximation . . . 19
- 3.5 Monitoring function for approximation . . . 20
- 3.6 Optimal design . . . 21
- 3.7 Discussion of the method . . . 22

4 Monitoring design: Translation of an average concentration footprint **27**
- 4.1 Estimation of the leakage footprint . . . 27
- 4.2 Detection function for translated average . . . 28
- 4.3 Monitoring function for translated average . . . 29
- 4.4 Inversion by rotation . . . 30
- 4.5 Extended application of the inversion . . . 32
- 4.6 Optimal design . . . 34
- 4.7 Discussion of the method . . . 35

5 Monitoring design: Interpolation of average concentration footprints **37**
5.1 Estimation of the leakage footprint 37
5.2 Detection function for interpolation 39
5.3 Monitoring function for interpolation 40
5.4 Extended application of the spatial variation 42
5.5 Optimal design 43
5.6 Discussion of the method 44

6 Monitoring design: Translation of a probability footprint **47**
6.1 Probability footprint 48
6.2 Detection function for event 50
6.3 Monitoring function for event 51
6.4 Optimal design 52
6.5 Discussion of the method 53

7 Monitoring design: Interpolation of probability footprints **57**
7.1 Estimation of the probability footprint 58
7.2 Detection function for interpolated event 58
7.3 Monitoring function for interpolated event 60
7.4 Optimal design 62
7.5 Discussion of the method 64

8 Discussion and further work **67**
8.1 Developed methods for monitoring design 67
8.2 Footprint prediction and detection function 68
8.3 Spatial dependency on leakage location in footprint estimation . 70
8.4 Inversion of the footprint 72
8.5 The leakage scenarios 73
8.6 Computational considerations 74
8.7 Closing remarks 75

Bibliography **77**

A Symbols used **79**

B Computational considerations **83**

C Prediction of leakage scenarios **89**

Chapter 1

Introduction

1.1 CCS as a greenhouse gas technology

Release of the greenhouse gas CO_2 from human activities has over the last hundreds of years increased the concentration of CO_2 in the atmosphere with about 40% [1]. This increases the greenhouse effect, and is one of the main reasons for global warming.

About 30% of the CO_2 released by humans into the atmosphere is dissolved into the ocean [2]. This leads to an increased concentration of CO_2 in the ocean. To obtain chemical equilibrium, some CO_2 reacts with the water. This increases the concentration of H^+, reducing the pH and making the ocean more acid.

There are several options to mitigate the increasing level of CO_2 in the atmosphere, where five different methods are suggested in Metz et al. [10]. The most obvious thing to do is to reduce the energy consumption by improved energy efficiency. However, people in developed countries don't want to lower their living standard. Together with an increasing living standard in developing countries, this makes it hard to reduce the energy consumption.

The second possibility is to replace fossil fuels with less carbon intensive fuels. An example of this is to use natural gas instead of coal. The third option presented is to increase the use of renewable energy sources or nuclear power. These energy sources contribute with little or no net CO_2. The next possibility is by sequestering CO_2 by improved biological absorption capacity in forests and soils. The fifth and final method presented is Carbon Capture and Storage (CCS), either chemically or physically. The different methods are of course not disjoint, but should be performed together to give a synergy in the mitigation of the increased CO_2 level.

In order to make a significant contribution to the mitigation of climate changes, CO_2 storage needs to be done in quantities of gigatonnes of CO_2 per year [10]. The large scale storage sites being operative today store about one megatonne. The CO_2 must be stored and kept away from the atmosphere for a long time, at the scale of hundreds to thousands of years [9]. If not, the extra energy needed to perform CCS may lead to a net increase in the CO_2 level in the atmosphere and in the ocean.

Capture

Systems for CCS can be installed in power plants to capture CO_2 processed in the plant. The available technology is capable of capturing about 85-95% of the CO_2 in a plant with capture system. However, such plants require approximately 10-40% more energy than their equivalent plants without CCS technology. This implies that the CO_2 emissions from plants with CCS could be reduced by about 80-90% compared to regular plants without CCS [9]. The capture of CO_2 is costly and the main bottleneck of CCS.

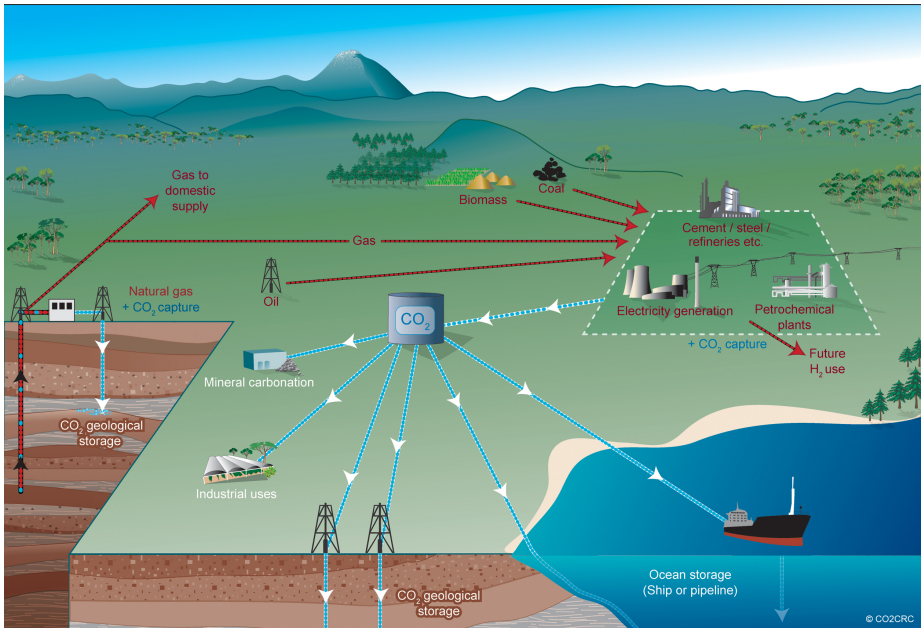


Figure 1.1: Schematic diagram of possible CCS systems. Courtesy of CO2CRC.

Transport

The most common method of transportation of CO_2 from the capture to storage site is by pipelines. High pressure gaseous CO_2 can be transported over long distances ranging several thousand kilometres using mature technology [9]. The CO_2 may also be transported as liquid in insulated tanks. If the CO_2 is to be transported over large distances, transportation by ship often is the most economically attractive.

Storage

The only type of CO_2 storage having a mature technology is geological storage [9]. CO_2 is injected into underground geological formations and trapped. The formations may be onshore or offshore where oil and gas reservoirs are typical formations used. The CO_2 is injected in dense form into porous rock formations that hold or previously have held fluids, typically natural gas or oil reservoirs or aquifers. For the injection of CO_2, much of the technology developed by the oil and gas industry can be used.

Storage of CO_2 in reservoirs usually takes place at a depth of more than 800 m below the seafloor such as at the Sleipner storage site. At this depth, the CO_2 will be in liquid or supercritical state with a density 50-80% of the density of water. With this density, the CO_2 becomes buoyant and will experience forces driving it upwards [9]. For this reason, a cap rock sealing the storage reservoir is needed to keep the CO_2 trapped. Injected CO_2 will fill the pore space by displacing fluids already present called the "in situ fluids". For oil and gas reservoirs, most of the pore volume can be made available for storage by displacement of in situ fluids [9].

The cap rock above the storage formation trapping the CO_2 is important to avoid upward migration and usually consists of shale and clay rock. In addition, capillary forces help keeping the CO_2 in the pores. However, the storage formation is often open for lateral migration below the cap rock. Independently of the efforts made to prevent leakages, we can never completely rule out the possibility of a leakage.

Impact of a leakage

The impact of a leakage on the marine ecosystem is studied in ECO_2 [12]. It is found that the CO_2 is mixed over a small spatial area. This limits the impact of a leakage to be mainly local. One of the main factors determining the severity of a leak, is the decrease in pH due to the leakage. A decrease of less than approximately 0.5 pH units is found to have minimal impact. A larger decrease in pH may also have a minimal impact in short term. However, a large decrease in pH has large long term impact. If the decrease is above approximately 1 pH unit, the decrease in biomass will be large, and *"after 1 to 3 growing seasons the biomass loss for the macrofauna is near complete"* [12]. An even larger pH decrease, estimated to approximately 1.3 pH units, one reaches the stage of general mortality. The impact of a leakage will also depend on other factors as other pollutants, salinity and natural seasonal events.

Even though a potential leakage has relatively small and local impact, it is important that leakages are detected. To fulfil the trading scheme, it must be verified that injected CO_2 is kept in the storage formation. To do so, one must ensure that a potential leakage is detected.

Regulations

The EU directive 2009/31/EC on geological storage of CO_2 establishes the legal framework for storage. When performing offshore storage, these regulations are aligned with the amendments to the 1996 London Protocol and to the OSPAR Convention. Common for the regulations, is to minimize negative effects and environmental risks due to the storage. An important part of this is an adequate monitoring program.

1.2 Monitoring of a storage site

A monitoring system should be created such that the probability of detecting a leakage from the storage formation is maximized. When creating such a monitoring system, there are several challenges.

Storage sites are typically offshore. Marine operations are costly and the environment is hostile to instruments. In addition, the area of potential leakage locations is usually large, for the Sleipner storage site at the scale of thousands of square kilometres. Another problem is the long time frame involved. Leakages with low flux rates are hard to detect, but the total leaked CO_2 may be large since leakages may last for years.

Monitoring techniques

There are several ways to monitor a storage site. Five different activities that should be repeated regularly are suggested by ECO_2 [12]:

- **3-D seismic:** By using seismic surveys, potential ascent of CO_2 from the storage formation is detected as changes in the seismic signatures. When using this technique, not only the storage reservoir should be monitored, but also the overlying sequences.
- **Bathymetry/backscatter:** Using acoustic backscattering techniques, new seabed structures created at the seabed can be detected.
- **Hydro-acoustic:** Shallow gas accumulations and gas bubbles seeps at the seafloor may be detected using hydro-acoustic methods.
- **Video/photo:** Biological indicators for leakage may be visible on video or pictures. Mats of bacteria is one such potential indicator.
- **Chemical:** Sensors are used to measure the concentration of CO_2 and other parameters. A CO_2 leakage results in increased CO_2 concentration at the seafloor in a region about the leakage. The other parameters measured may help to distinguish a leakage from natural variability.

Chemical sensors

We will study the use of chemical sensors to monitor the CCS site. These sensors measure the concentration of CO_2 and other relevant parameters in the water at the seafloor. Concentrations above a detection limit (Sec. 2.2.1) will trigger the leakage alarm. Different methods for estimating when this happens are developed in Ch. 3 to 7.

Fixed sensors

One way to implement a monitoring system using chemical sensors is to make a fixed array of sensors on the seafloor. This will provide time series of the concentration at each of the sensor locations, making statistical studies possible. Due to the cost of installing sensors, it is desirable to find the sensor layout monitoring the area using as few sensors as possible.

In Hvidevold et al. [8] it is shown that for a fixed number of sensors, the probability of detecting a leakage can nearly be doubled by an optimal sensor layout compared to using an equally spaced array of sensors. We will use fixed sensors in our work, seeking to find optimal sensor layouts. This is a relatively new field of study, and we will base our work on the framework of Hvidevold et al. in [8] and [7].

Surveys
Instead of installing fixed sensors on the seafloor, surveys can be performed to check for indications of leakages. Measurements of the concentration are taken at specific locations. This will only give single measurements at a given time, not concentration time series. Due to the operational cost of the vessel used to perform the measurements, we still need to limit the number of measurements. To maximize the probability of detecting a potential leakage, these measurement locations should be optimized similar to the sensor layout for the fixed sensors.

Taking into account the potential large distances between the measurement locations, an optimal cruiseplan for the measurements is necessary to minimize transit time and operational cost. In Hvidevold et al. [7], the survey method is compared to a fixed sensor grid. An optimal cruise plan for a synthetic problem is also presented. We will not study the use of surveys, as we are interested in optimal layouts of fixed sensors.

Alarm response

In case of an alarm by the monitoring system indicating an ongoing leakage, investigation must be initiated. The first thing that needs to be done is to localize the potential leakage location. Dependent on the monitoring system, the potential leakage area of an alarm may be large. Surveys to identify the source of the increased concentration may then be a costly process. In addition, the leakage may be large and not only a point source. This implies that the fluxes measured will be smaller, making the source harder to detect.

If we are certain that there is a leakage, one should stop the injection. Further action to stop the leakage should also be performed, but since CCS is relatively new technology, there is no established procedure to stop leakages.

Chapter 2

Monitoring design: Mathematical framework

In this chapter we will develop the mathematical framework needed to estimate and compare the ability of different sensor layouts to detect a leakage. This leads to the problem of finding the optimal sensor layout having the highest possible detection probability. Monitoring design can be divided in two parts: 1) The site dependent studies, localizing leakage pathways and predicting leakage footprints. 2) The detection calculations leading to an optimal sensor layout and a detection probability. We will focus on the second part by developing the method for the calculations shown in grey in Fig. 2.1. In the following chapters, different methods will be developed and studied using test cases.

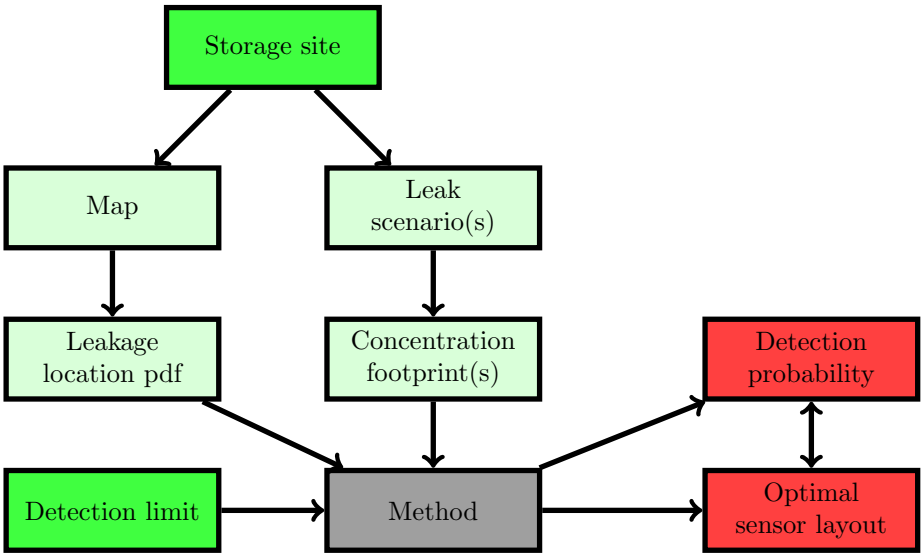


Figure 2.1: Schematic presentation of monitoring design. The green boxes are inputs to the method and the red boxes are outputs.

2.1 Site characteristics

We will now discuss what information is needed to perform a site study. Being interested in the method used for detection calculations, and not specific site studies, a complete site study is not performed here.

2.1.1 Potential leakage locations

We assume that we have the area A to be monitored for leakages. An important question when designing a monitoring program, is where a leakage is likely to occur. Places of a high leakage probability is more important to monitor than those of a low leakage probability. We assume that there is exactly one leakage ongoing. For this leakage, we create a probability density function $f(\mathbf{x})$ for the location. In other words, $f(\mathbf{x})$ is the probability density of $\mathbf{x}$ being the leakage location given that there is a leakage ongoing in A, hence

$$f(\mathbf{x}) \geq 0 \qquad \forall \mathbf{x} \in A \qquad \text{and} \tag{2.1}$$

$$\int\int_A f(\mathbf{x}) dxdy = 1. \tag{2.2}$$

The function $f(\mathbf{x})$ will be site dependent. To make a realistic probability density function, pathways from the storage formation to the seafloor should be identified and given a leakage probability. This is a cumbersome process and requires much information about the storage formation and the overburden. A simplified approach used in Hvidevold et al. [8] is to use a map of faults and wells. In our work, we think of the leakage location probability density function as given from site surveys. For this reason, we will use synthetic test cases.

2.1.2 Footprint

To predict the footprints of potential leakages, the General Circulation Model Bergen Ocean Model is used to simulate leakage scenarios as described in App. C. Some methods only use the average concentration, giving a leakage footprint. If all leakage locations are assumed to have the same footprint, only one simulation is needed. If we on the other hand only have access to one simulation, we have to assume that all leakage locations have the same footprint. Fig. 2.2 shows the average concentration footprint from [8] used as test case in Ch. 3 and 4.

2.1.3 Synthetic map

We will adopt the test case of [8], where the Sleipner gas field is to be monitored. Sleipner was the first large scale storage site in the world used for CCS, being operative since 1996. Sleipner is located west of Stavanger in the North Sea, by the border between Norway and Great Britain. A public map by the Norwegian Petroleum Directorate is used to locate faults and wells in an area of 77x77 km^2. The western part of the region considered is British, and the map does not include the wells in this area making the test case less authentic. The map of faults and wells used is shown in Fig. 2.3.

It is assumed that all faults and wells have the same probability of being the location of a leak, with wells having a higher probability than faults. This

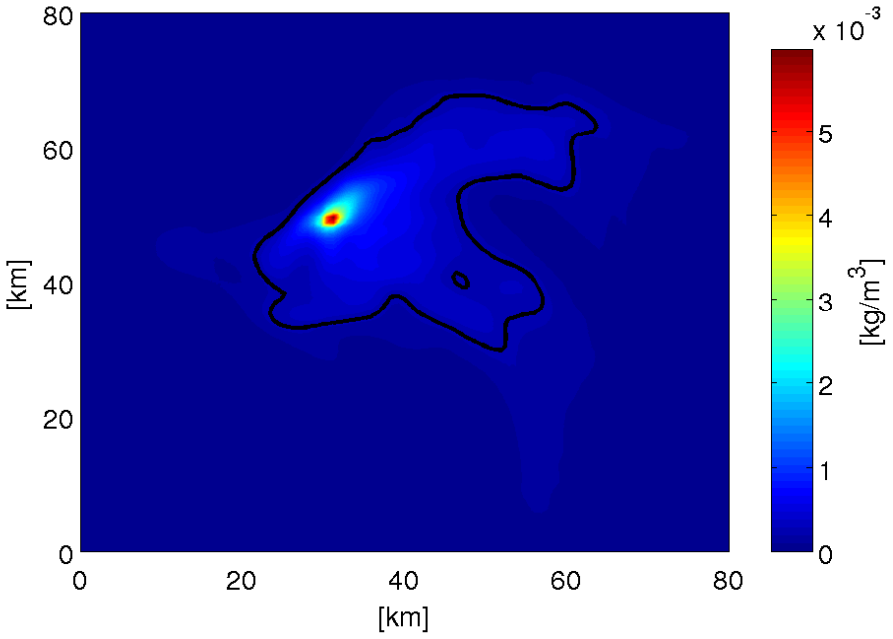


Figure 2.2: The simulated seafloor CO_2 concentration used in [8]. The black contour line shows where the concentration is above $c_t = 2.26 \cdot 10^{-4}$ kg/m^3.

simplification implies that it is not distinguished between faults and wells reaching and not reaching the storage formation. To do such a distinction would require a thorough analysis of the overburden.

Due to potential unknown pathways from the storage formation, every location can be given a small background probability of being the location of a leak.

Assuming that a leakage is ongoing inside the map, a probability density function $f(\mathbf{x})$ for the location is obtained by normalizing the probabilities such that the function integrates to 1 according to Eq. 2.2.

2.2 Detection calculations

We will now discuss the mathematical framework of calculating detection probabilities and optimizing sensor layouts. This section is site independent, and will be used in the following chapters to develop different methods that can be applied to any storage site.

2.2.1 Detection limit

A leakage of CO_2 will result in an increased CO_2 concentration in a region around the leakage location. This increase in CO_2 can be detected by a sensor if the sensor is sufficiently close to the leakage. Due to natural variability in the CO_2 concentration in the ocean and sensor precision, the increase in the CO_2 concentration from the leakage must be above a threshold value c_t before

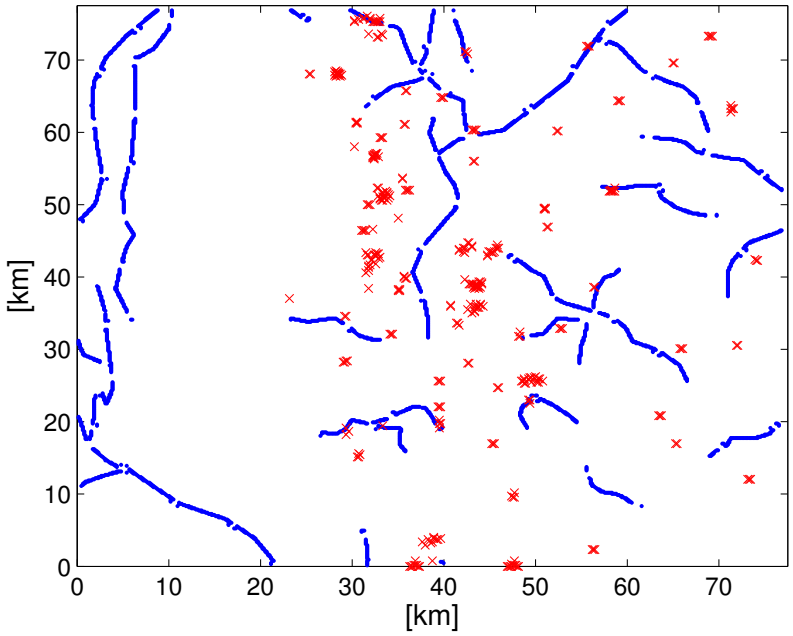


Figure 2.3: Map of faults (blue) and wells (red) used in [8].

it is detected by the sensor. Several different threshold values are discussed in Hvidevold et al. [8].

The theoretical precision for measuring total inorganic carbon is approximately $9.0 \cdot 10^{-5}$ kg/m^3. The natural variability in the north sea is between $2.260 \cdot 10^{-3}$ kg/m^3 and $4.520 \cdot 10^{-3}$ kg/m^3. If we don't have statistics for this variability, a high threshold concentration should be used to ensure that the measured value is statistically significant. Having statistics of the variability, a threshold value $c_t = 2.26 \cdot 10^{-4}$ kg/m^3 is suggested by Hvidevold et al. [8], which we will adopt.

2.2.2 Detection function

Assuming that there is a leakage at a given known location, an important question when is what the probability of detecting this leakage is for different sensor locations. This probability will be frequently used, and we will call it the detection function as in Hvidevold et al. [7]. We assume that the leakage occurs at $\mathbf{x}_0$ and a threshold concentration c_t of the sensors. The detection function $D(\mathbf{x}; \mathbf{x}_0; c_t)$ is the probability that the leakage is detected by a sensor at $\mathbf{x}$.

The methods of the different chapters will give different detection functions $D(\mathbf{x}; \mathbf{x}_0; c_t)$, which we will emphasize by using different subscripts and superscripts. To be a probability, D must satisfy

$$0 \leq D(\mathbf{x}, \mathbf{x}_0; c_t) \leq 1 \qquad \forall \mathbf{x}, \mathbf{x}_0 \in A \text{ and } \forall c_t > 0. \tag{2.3}$$

Independently of the method used, we expect D to be close to 1 near the leakage, decreasing away from the leakage and close to 0 far from the leakage.

2.2.3 Monitoring function

We now assume that we have a sensor at a given location $\mathbf{x}_s$ with threshold concentration c_t. What area is monitored by this sensor? To answer this, we define the monitoring function similar to the detection function introduced in Sec. 2.2.2. The monitoring function $M(\mathbf{x}; \mathbf{x}_s; c_t)$ is the probability that a leakage at $\mathbf{x}$ is detected by the sensor at $\mathbf{x}_s$.

For the detection function in Sec. 2.2.2, we fixed the leakage location and let the sensor location be the independent variable. For the monitoring function, we fix the sensor location and let the leakage location be the independent variable. This implies that we can look at the monitoring function as an inversion of the detection function. Thus, different methods of creating the detection function will give different monitoring functions.

An important relation between the detection and monitoring function showing their inversion property follows directly from their definitions:

$$M(\mathbf{x}; \mathbf{x}_s; c_t) = D(\mathbf{x}_s; \mathbf{x}; c_t). \tag{2.4}$$

Most of the time, D will be known and we want to find M. To be a probability, M must satisfy the same condition (Eq. 2.3) as D:

$$0 \leq M(\mathbf{x}, \mathbf{x}_s; c_t) \leq 1 \qquad \forall \mathbf{x}, \mathbf{x}_s \in A \text{ and } \forall c_t > 0. \tag{2.5}$$

Independently of the method used, we expect M to be close to 1 near the sensor, decreasing away from the sensor and close to 0 far from the sensor.

2.2.4 Detection probability

We now want to estimate the probability that an arbitrary leakage is detected for a given storage site (Sec. 2.1) and sensor layout. This probability will be called the detection probability.

We assume that we have N sensors in a given layout and want to estimate their total detection probability. We assume that the monitoring functions for the sensors are independent of each other and are calculated as before. Let $\mathbf{x}_s^n$ be the location of sensor n. The monitoring function for this sensor will be $M(\mathbf{x}; \mathbf{x}_s^n; c_t)$. We introduce the short notation M_n for the following calculations:

$$M_n = M(\mathbf{x}; \mathbf{x}_s^n; c_t). \tag{2.6}$$

The probability that sensor n does not detect a leakage at $\mathbf{x}$ is given by $(1 - M_n)$. Having independent monitoring functions, the total probability of none of the sensors detecting a leakage at a given location $\mathbf{x}$ is then given by

$$p(\text{not detect}|\mathbf{x}) = \prod_{n=1}^{N} (1 - M_n) = (1 - M_1) \dots (1 - M_N). \tag{2.7}$$

The probability of detecting a leakage at a given location $\mathbf{x}$ is then given by the complementary probability

$$p(\text{detect}|\mathbf{x}) = 1 - p(\text{not detect}|\mathbf{x}) \tag{2.8}$$

$$p(\text{detect}|\mathbf{x}) = 1 - \prod_{n=1}^{N}(1 - M_n) = 1 - (1 - M_1)\dots(1 - M_N). \tag{2.9}$$

Since the probability density function for the leakage location is given by $f(\mathbf{x})$, the conditional probability density $f(\mathbf{x})p(\text{detect}|\mathbf{x})$ will be the non-proper probability density function for the leakage being located in $\mathbf{x}$ and detected by the sensors. It will be non-proper since it does not necessarily integrate to 1. In fact, the probability p of detecting an arbitrary leakage is given by the integral of this non-proper probability density function:

$$p = p(\mathbf{x}_s^1; \dots; \mathbf{x}_s^N; f; c_t) = \int\int_A f(\mathbf{x})p(\text{detect}|\mathbf{x})dxdy. \tag{2.10}$$

The only time time this will be 1 is if all leakages always are detected independently of the leakage location implying that we have a perfect monitoring program. Most of the time, this will certainly not be the case. We will now look at different ways to compute the detecting probability p. It can of course be computed directly by

$$p(\mathbf{x}_s^1; \dots; \mathbf{x}_s^N; f; c_t) = \int\int_A f(\mathbf{x}) \left(1 - \prod_{n=1}^{N}(1 - M_n)\right) dxdy. \tag{2.11}$$

By rewriting the integrand of Eq. 2.11 we obtain

$$f(\mathbf{x}) \left(1 - \prod_{n=1}^{N}(1 - M_n)\right) = \tag{2.12}$$

$$f(\mathbf{x}) \left(1 - (1 - M_N)\prod_{n=1}^{N-1}(1 - M_n)\right) = \tag{2.13}$$

$$f(\mathbf{x}) \left(1 - \prod_{n=1}^{N-1}(1 - M_n) + M_N \prod_{n=1}^{N-1}(1 - M_n)\right) = \tag{2.14}$$

$$f(\mathbf{x}) \left(1 - \prod_{n=1}^{N-1}(1 - M_n)\right) + M_N f(\mathbf{x}) \prod_{n=1}^{N-1}(1 - M_n). \tag{2.15}$$

We see that the integrand can be divided in two terms. The first term is what the integrand would look like without the last sensor. The second term corresponds to the extra probability contribution from adding the last sensor to a sensor layout of the $N - 1$ first sensors. This suggests a recursive calculation placing one sensor at the time. We can continue to rewrite to obtain

$$f(\mathbf{x})\left(1-\prod_{n=1}^{1}(1-M_n)\right)+\sum_{k=2}^{N}M_k f(\mathbf{x})\prod_{n=1}^{k-1}(1-M_n)= \tag{2.16}$$

$$M_1 f(\mathbf{x})+\sum_{k=2}^{N}M_k f(\mathbf{x})\prod_{n=1}^{k-1}(1-M_n)= \tag{2.17}$$

$$\sum_{k=1}^{N}M_k f(\mathbf{x})\prod_{n=1}^{k-1}(1-M_n). \tag{2.18}$$

Here, we have used the convention of an empty product being 1, hence

$$\prod_{n=1}^{0}(1-M_n)=1. \tag{2.19}$$

We define the residual leakage location probability density functions

$$f_k(\mathbf{x})=f(\mathbf{x})\prod_{n=1}^{k}(1-M_n)\qquad k=0,\ldots,N. \tag{2.20}$$

As suggested by the name, $f_k(\mathbf{x})$ will be the residual probability left to be monitored after k of the sensors are placed. The special case $k=0$ will be the original leakage location probability density function. For $k\neq 0$, the residual probability density function will usually not integrate to 1 and thus not be a proper probability density function. If it does, the sensor placed will not monitor anything and no leakages will be detected independently of the leakage location.

The integrand of Eq. 2.11 can now be written

$$\sum_{k=1}^{N}M_k f_{k-1}(\mathbf{x}), \tag{2.21}$$

meaning that the detection probability can be written

$$p(\mathbf{x}_s^1;\ldots;\mathbf{x}_s^N;f;c_t)=\int\int_A\sum_{k=1}^{N}M_k f_{k-1}(\mathbf{x})dxdy= \tag{2.22}$$

$$\sum_{k=1}^{N}\int\int_A M_k f_{k-1}(\mathbf{x})dxdy. \tag{2.23}$$

This emphasises the process of placing one sensor at the time and adding the extra probability contribution of adding the sensor. We observe that

$$f_N(\mathbf{x})=f(\mathbf{x})\prod_{n=1}^{N}(1-M_n)\qquad\text{and} \tag{2.24}$$

$$f_k(\mathbf{x})=f_{k-1}(\mathbf{x})(1-M_k)\qquad k=1,\ldots,N. \tag{2.25}$$

We see that the residual probability density functions f_k can be calculated recursively. Since f_N is the residual probability density after placing all the sensors, we can use a recursive process to calculate the detection probability p:

$$f_0(\mathbf{x}) = f(\mathbf{x}) \tag{2.26}$$

$$f_n(\mathbf{x}) = f_{n-1}(\mathbf{x})\,(1 - M_n) \qquad n = 1, \ldots, N \tag{2.27}$$

$$p = p(\mathbf{x}_s^1; \ldots; \mathbf{x}_s^N; f; c_t) = 1 - \int\int_A f_N(\mathbf{x}) dxdy. \tag{2.28}$$

Note that the numbering of the sensors is irrelevant, as is easily seen using Eq. 2.11. The sensor layout will look the same and have the same detection properties independently of the sensor numbering, implying that all permutations of given sensor locations are the same layout.

2.2.5 Optimization problem

For a given storage site (Sec. 2.1), the detection probability $p(\mathbf{x}_s^1; \ldots; \mathbf{x}_s^N; c_t)$ given by Eq. 2.28 will be a function of the sensor locations and the threshold concentration. We will consider problems with fixed threshold concentration. The detection probability will then be a function only of the sensor locations.

When designing a monitoring program, we want the detection probability to be as high as possible. Finding the optimal sensor layout can then be viewed as a non-linear optimization problem. The objective function to maximize is the detection probability p given by Eq. 2.28. The optimization parameters to be found are the sensor locations $\mathbf{x}_s^1; \ldots; \mathbf{x}_s^N$. For each sensor, the location is given by two parameters, the x- and the y-coordinate. Having N sensors, the system has $2N$ degrees of freedom to be determined. Using $\mathbf{z}_n$ as optimization parameter vector for sensor n, the optimization problem can be written

$$\left[\mathbf{x}_s^1; \ldots; \mathbf{x}_s^N\right] = \text{argmax}\left(p(\mathbf{z}_1; \ldots; \mathbf{z}_N; f; c_t)\right) \qquad f, c_t \text{ fixed.} \tag{2.29}$$

Since the numbering of the sensors is irrelevant for the computation of p, any permutation of the best sensor locations is a solution. There may be several layouts that have almost the same detection probability. It is suggested in Hvidevold et al. [8] that other factors as installation and operational cost can be included in the cost function to choose between the different layouts, but this is not done here. Due to the complex structure of the cost function, a numerical optimization toolbox must be applied to solve the problem. We will use the built-in Genetic Algorithm *ga* in MATLAB.

As shown in Sec. 2.2.4, the detection probability for a given sensor layout can be calculated by placing one sensor at the time and consecutively adding the extra probability contributions from each sensor.

It can be tempting to apply a similar consecutive approach to the optimization problem by finding the optimal placement of one sensor at the time. This splits the optimization problem in N optimization problems of two degrees of freedom instead of one with $2N$ degrees of freedom. Since the total degrees of freedom are equal, we expect the N small problems to be easier to solve than the large.

Using the notation of residual leakage location probability density functions f_n, we can write the N smaller optimization problems

$$\mathbf{x}_s^n = \text{argmax}\left(p(\mathbf{z}_n; f_{n-1}; c_t)\right) \qquad n = 1, \ldots, N \qquad f, c_t \text{ fixed.} \tag{2.30}$$

This procedure will place the first sensor at the location having the highest detection probability for one sensor. The subsequent sensors will then always be placed at the location having the highest detection probability for one sensor using the residual leakage location probability density function resulting from the previous sensors. In other words, this is the intuitive approach of always placing the next sensor at the best available location.

This does not take into account the interaction between the sensors, which as noted in Hvidevold et al. [8] is the reason that the best layout of N sensors does not necessarily have a sensor at the location with the highest detection probability. A result of this, is that we can only split the optimization problem into the smaller problems if the sensors of the optimal sensor layout are independent. Mathematically, this corresponds to the monitoring functions of the sensors of the optimal sensor layout having disjoint supports. Since we don't know the optimal sensor layout before we perform the optimization, we can not assume the independence of the sensors needed to split the optimization problem.

If all possible sensor locations in the region to be monitored have monitoring functions with small supports, the splitting of the optimization will typically be close to unbiased.

Genetic Algorithm (GA)

We use the Genetic Algorithm *ga* in MATLAB to solve the optimization problem of Eq. 2.29. A Genetic Algorithm (GA) is an evolutionary algorithm which uses properties of evolution to solve optimization problems. GAs are stochastic and are typically applied to hard optimization problems where it is difficult to find an efficient probabilistic algorithm [11]. We have chosen to use a GA to solve Eq. 2.29 because of the complex structure of the cost function and previous successful use on Eq. 2.29 by Hvidevold et al. in [8] and [7].

A GA solves optimization problems by evolving solutions towards the optimal solution. Some of the key words in this process are the evolutionary terms generation, population, individual and mutation. A generation is a population of solutions called individuals. Each individual is a complete solution of the optimization problem. In our case, this implies that an individual consists of a sensor location for all the sensors. The problem of finding the optimal solution can be described by finding the fittest individual, i.e. the individual having the best value of the cost function. A population consists of a fixed number of individuals called the population size, usually 500 in our calculations.

The initial population, i.e. the first generation, must be created stochastically. This should be done in a way such that a wide variety of potential solutions in the solution space is represented. The fitness, i.e. the value of the cost function, is calculated for each individual together with an average fitness value of the population. Based on this, a new generation is created that hopefully has better fitness than the previous. This is done by mutating individuals and favouring individuals having better fitness. In addition, the fittest individuals of a generation always survive to the next generation. We use a rate of 5%, implying that the 25 fittest individuals always survive having a population size of 500. However, it can also be beneficial to keep individuals with lower fitness as the optimal solution may evolve from individuals with low fitness.

Following generations are then created until a termination condition is met. Some possible conditions are a fitness limit, a maximum number of generations,

a fitness change limit, or a combination. Since we are maximizing a probability, we have a natural fitness limit of 1 implying that a solution reaching this value per definition is optimal. We also apply a maximum of 100 generations, implying that the optimization will terminate regardless of the fitness if the algorithm reaches 100 generations. Finally, we terminate the process if the average relative change of fitness of the fittest individual over 50 generations is below some limit, usually between 10^{-6} and 10^{-3}. In the optimizations run, it will typically be the tolerance on the average relative change of fitness that terminates the process.

It is found that it is beneficial to apply a large rate of mutation when solving Eq. 2.29, or the algorithm will typically converge to suboptimal solutions. This is probably since the fittest solutions have a large "distance" between them and the solutions between them have low fitness. In other words, one has to perform a large mutation (move the sensors far) to get from a good solution to an even better solution. We use the option *mutationuniform* in MATLAB with a mutation rate of 0.4. This performs the same mutation for later generations as for the first, implying that the average fitness of the populations is lower, but the ability to find optimal solutions is better.

To ensure the results and account for the stochastic element of the algorithm, we usually run the optimization at least 10 times and use the best solution obtained. The current implementation *ga* in MATLAB does not allow for parallel computing when solving an optimization problem. However, we are typically interested in solving Eq. 2.29 for different number of sensors, giving a separate optimization problem for each. This allows for parallel computing, solving each of the optimization problems on a separate core. Alternatively, the same optimization problem can be solved independently on each of the available cores, and then the best of the obtained solutions is chosen. We use this approach since it instantaneously gives the desired verification of the solution.

Chapter 3

Monitoring design: Approximation of an average concentration footprint

In this chapter, the average CO_2 concentration resulting from a leakage is used in the design of the monitoring program. We assume that there is a leakage ongoing, and that we want to find the optimal layout of sensors. If the average concentration over some given time at a sensor location is above a threshold concentration c_t, the sensor is assumed to always detect the leakage. If it is below, the sensor is assumed to never detect the leakage. We will present the previous work done in [8] based on an approximation of average concentration footprint, mainly to compare and check later results. The approximation and assumptions made are discussed and potential improvements suggested.

3.1 Detection function for average

Having a leakage at $\mathbf{x}_0$, the average concentration at $\mathbf{x}$ is denoted by $\bar{C}(\mathbf{x}, \mathbf{x}_0)$. If the concentration at $\mathbf{x}$ is above the threshold concentration c_t, a sensor at $\mathbf{x}$ is assumed to always detect the leakage. If the concentration is below c_t, the sensor is assumed to never detect the leakage. The detection function for this approach is denoted $D_A(\mathbf{x}; \mathbf{x}_0; c_t)$ and fulfils Eq. 2.3:

$$D_A(\mathbf{x}; \mathbf{x}_0; c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x}; \mathbf{x}_0) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x}; \mathbf{x}_0) < c_t \end{cases}. \tag{3.1}$$

This will give one or several regions where the concentration is above c_t and the detection function is 1, meaning that any sensor in one of these regions is assumed to detect the leakage. This will be called the detectable region:

$$\boxed{\text{Detectable region average}} = \left\{ \mathbf{x} \,\middle|\, \bar{C}(\mathbf{x}; \mathbf{x}_0) \geq c_t \right\}. \tag{3.2}$$

3.2 Monitoring function for average

The monitoring function is implicitly defined by the detection function using

$$M(\mathbf{x};\mathbf{x}_s;c_t) = D(\mathbf{x}_s;\mathbf{x};c_t). \tag{2.4}$$

We denote the monitoring function for the average concentration method $M_A(\mathbf{x};\mathbf{x}_s;c_t)$. Using Eq. 2.4 and Eq. 3.1, we obtain for a sensor at $\mathbf{x}_s$

$$M_A(\mathbf{x};\mathbf{x}_s;c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x}_s;\mathbf{x}) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x}_s;\mathbf{x}) < c_t \end{cases} \tag{3.3}$$

where $\bar{C}(\mathbf{x}_s;\mathbf{x})$ is the average concentration at $\mathbf{x}_s$ caused by a leakage at $\mathbf{x}$.

We see that a sensor located at $\mathbf{x}_s$ is assumed to always detect leakages at all locations $\mathbf{x}$ resulting in an average concentration above c_t at $\mathbf{x}_s$. The sensor is assumed to never detect leakages at locations $\mathbf{x}$ resulting in an average concentration below c_t at $\mathbf{x}_s$. As for the detection function, we get one or several regions where the monitoring function is 1 and leakages always will be detected. This will be called the monitored region and is

$$\boxed{\text{Monitored region average}} = \left\{\mathbf{x}\,\middle|\,\bar{C}(\mathbf{x}_s;\mathbf{x}) \geq c_t\right\}. \tag{3.4}$$

3.3 Previous work: Approximation of footprint

We will present the method used in Hvidevold et al. [8]. All leakages are assumed to have the same average concentration footprint which is predicted by simulating a leakage scenario. The simulated footprint is then approximated by a family G of functions on the form

$$G(\mathbf{x};\mathbf{x}_0;\mathbf{z}) = \bar{C}_0 \exp\left\{-\left(\mathbf{x}-\mathbf{x}_0\right)^T A(\mathbf{z})\left(\mathbf{x}-\mathbf{x}_0\right)\right\} \tag{3.5}$$

with the matrix

$$A(\mathbf{z}) = \begin{bmatrix} \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2} & -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2} \\ -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2} & \frac{\sin^2\theta}{2\sigma_x^2} + \frac{\cos^2\theta}{2\sigma_y^2} \end{bmatrix}, \tag{3.6}$$

a parameter vector $\mathbf{z} = [\theta, \sigma_x, \sigma_y]$, the maximum concentration $\bar{C}_0$ at the leakage point $\mathbf{x}_0$ and the position $\mathbf{x}$. Let $\bar{C}_i$ be the average simulated CO_2 concentration and G_i the approximated CO_2 concentration in grid point i. Further, let Σ_i be the empirical standard deviation of the time series of the CO_2 concentration in grid point i. The parameters in $\mathbf{z}$ are obtained by applying the Levenberg-Marquardt method to the non-linear least squares problem

$$\mathbf{z} = \operatorname{argmin}\left\{\sum_{i=1}^{I}\left(\frac{\bar{C}_i - G_i}{\Sigma_i}\right)^2\right\} \tag{3.7}$$

where I is the number of grid points. Fig. 3.1 shows the approximation of the footprint in Fig. 2.2 used in Hvidevold et al. [8]. The parameter vector obtained by minimizing Eq. 3.7 is $\mathbf{z} = [\theta, \sigma_x, \sigma_y] = [0.9125 \text{ rad}, 1.6207 \text{ km}, 2.6462 \text{ km}]$.

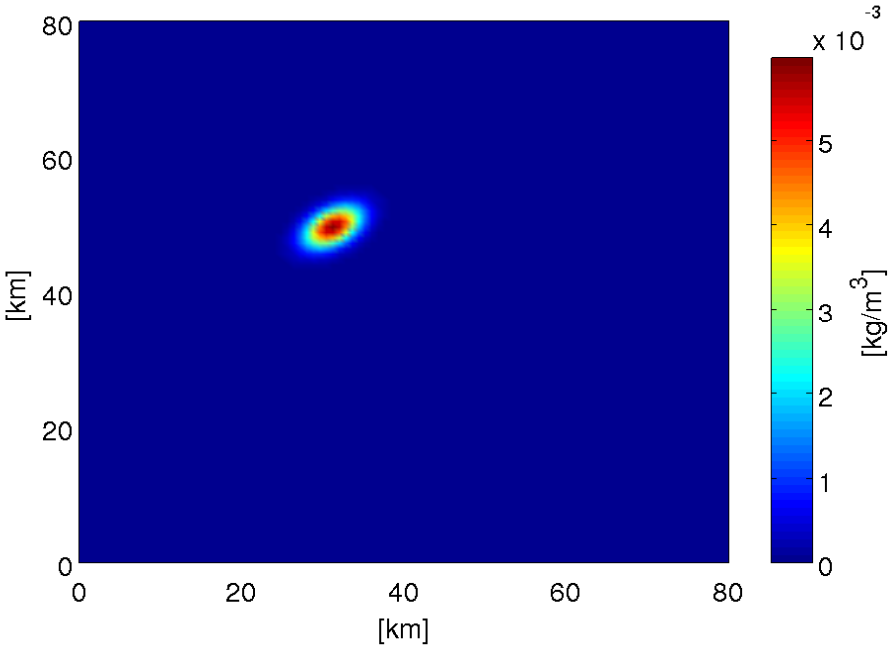


Figure 3.1: The approximation of the footprint in Fig. 2.2.

3.4 Detection function for approximation

Using the approximation $G(\mathbf{x};\mathbf{x}_0;\mathbf{z})$ of Eq. 3.5 for the average concentration footprint from a leakage at $\mathbf{x}_0$, the detection function of Eq. 3.1 becomes

$$D_A^A(\mathbf{x};\mathbf{x}_0;c_t) = \begin{cases} 1 & \text{if } G(\mathbf{x};\mathbf{x}_0;\mathbf{z}) \geq c_t \\ 0 & \text{if } G(\mathbf{x};\mathbf{x}_0;\mathbf{z}) < c_t \end{cases}. \tag{3.8}$$

To determine the detectable region where the approximation is above c_t, we need to find the level contours of $G(\mathbf{x};\mathbf{x}_0;\mathbf{z})$. For a given c_t, we have

$$c_t = G(\mathbf{x};\mathbf{x}_0;\mathbf{z}) = \bar{C}_0 \exp\left\{-(\mathbf{x}-\mathbf{x}_0)^T A(\mathbf{z})(\mathbf{x}-\mathbf{x}_0)\right\}, \tag{3.9}$$

such that the level contours of $G(\mathbf{x};\mathbf{z})$ are given by

$$\ln\frac{\bar{C}_0}{c_t} = (\mathbf{x}-\mathbf{x}_0)^T A(\mathbf{z})(\mathbf{x}-\mathbf{x}_0) \tag{3.10}$$

giving rise to ellipses with semi axes of lengths $\hat{\sigma}_x$ and $\hat{\sigma}_y$ rotated a clockwise angle θ (Eq. 3.6) with respect to the x-axis. The lengths of the semi axes can be found by setting $\theta = 0$. The level contours are then given by

$$\ln\frac{\bar{C}_0}{c_t} = \frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}. \tag{3.11}$$

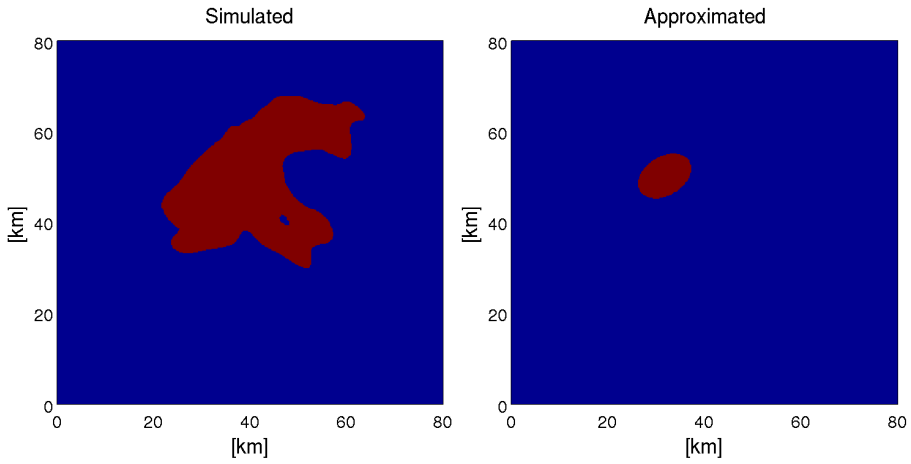


Figure 3.2: Detectable region (red) for the simulated and the approximated leakage footprint used in [8]. Threshold concentration $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}.

$\hat{\sigma}_x$ is obtained by setting $y = 0$, and $\hat{\sigma}_y$ by setting $x = 0$ giving

$$\hat{\sigma}_x = \sigma_x \sqrt{2 \ln \frac{\bar{C}_0}{c_t}} \qquad \text{and} \qquad \hat{\sigma}_y = \sigma_y \sqrt{2 \ln \frac{\bar{C}_0}{c_t}}. \tag{3.12}$$

Having a leakage at $\mathbf{x}_0$ and threshold concentration c_t for the sensors, the ellipse given by Eq. 3.10 defines the detectable region. All locations $\mathbf{x}$ inside this ellipse will have an average concentration above c_t, implying that sensors at these locations will detect the leak. The detectable region is thus given by

$$\boxed{\begin{array}{c}\text{Detectable region} \\ \text{approximated average}\end{array}} = \left\{ \mathbf{x} \,\middle|\, (\mathbf{x} - \mathbf{x}_0)^T A(\mathbf{z}) (\mathbf{x} - \mathbf{x}_0) \leq \ln \frac{\bar{C}_0}{c_t} \right\}. \tag{3.13}$$

Equivalently, the detection function for the approximation is given by

$$D_A^A(\mathbf{x}; \mathbf{x}_0; c_t) = \begin{cases} 1 & \text{if } (\mathbf{x} - \mathbf{x}_0)^T A(\mathbf{z}) (\mathbf{x} - \mathbf{x}_0) \leq \ln \frac{\bar{C}_0}{c_t} \\ 0 & \text{if } (\mathbf{x} - \mathbf{x}_0)^T A(\mathbf{z}) (\mathbf{x} - \mathbf{x}_0) > \ln \frac{\bar{C}_0}{c_t} \end{cases}. \tag{3.14}$$

Fig. 3.2 shows the detectable region of the simulated footprint used in Hvidevold et al. [8] and its approximation for $c_t = 2.26 \cdot 10^{-4}$ kg/m^3. The detection ellipse has semi axes $\hat{\sigma}_x = 4.15$ km and $\hat{\sigma}_y = 6.78$ km making a detection area for the approximation of $\hat{\sigma}_x \hat{\sigma}_y \pi \approx 88$ km^2.

3.5 Monitoring function for approximation

Using Eq. 3.8 for the detection function for approximated average, the monitoring function will be given by

$$M(\mathbf{x}; \mathbf{x}_s; c_t) = D(\mathbf{x}_s; \mathbf{x}; c_t). \tag{2.4}$$

In terms of the approximation function (Eq. 3.5) for the concentration, the monitoring function for the approximated average will then be

$$M_A^A(\mathbf{x};\mathbf{x}_s;c_t) = \begin{cases} 1 & \text{if } G(\mathbf{x}_s;\mathbf{x};\mathbf{z}) \geq c_t \\ 0 & \text{if } G(\mathbf{x}_s;\mathbf{x};\mathbf{z}) < c_t \end{cases}. \tag{3.15}$$

Due to symmetry of the assumed approximation profile, we can interchange the arguments $\mathbf{x}$ and $\mathbf{x}_s$ of the approximation function (Eq. 3.5):

$$G(\mathbf{x}_s;\mathbf{x};\mathbf{z}) = G(\mathbf{x};\mathbf{x}_s;\mathbf{z}). \tag{3.16}$$

This gives us the monitoring function for the approximated average:

$$M_A^A(\mathbf{x};\mathbf{x}_s;c_t) = \begin{cases} 1 & \text{if } G(\mathbf{x};\mathbf{x}_s;\mathbf{z}) \geq c_t \\ 0 & \text{if } G(\mathbf{x};\mathbf{x}_s;\mathbf{z}) < c_t \end{cases}. \tag{3.17}$$

We see that the monitoring function is the same as the detection function of Eq. 3.8, just with the sensor location $\mathbf{x}_s$ taking the role of the leakage location $\mathbf{x}_0$. The work done in Sec. 3.4 for the detection function will then be the same for the monitoring function. For a given threshold concentration c_t, the monitored region will be the same ellipse as we obtained for the detectable region in Sec. 3.4. Leakages at all locations $\mathbf{x}$ inside this ellipse will give a concentration above c_t at $\mathbf{x}_s$ and is assumed to be detected giving

$$\boxed{\begin{array}{c}\text{Monitored region} \\ \text{approximated average}\end{array}} = \left\{ \mathbf{x} \,\middle|\, (\mathbf{x}-\mathbf{x}_s)^T A(\mathbf{z}) (\mathbf{x}-\mathbf{x}_s) \leq \ln \frac{\bar{C}_0}{c_t} \right\}. \tag{3.18}$$

Equivalently, the monitoring function for the approximation is given by

$$M_A^A(\mathbf{x};\mathbf{x}_s;c_t) = \begin{cases} 1 & \text{if } (\mathbf{x}-\mathbf{x}_s)^T A(\mathbf{z}) (\mathbf{x}-\mathbf{x}_s) \leq \ln \frac{\bar{C}_0}{c_t} \\ 0 & \text{if } (\mathbf{x}-\mathbf{x}_s)^T A(\mathbf{z}) (\mathbf{x}-\mathbf{x}_s) > \ln \frac{\bar{C}_0}{c_t} \end{cases}. \tag{3.19}$$

3.6 Optimal design

We now want to maximize the detection probability to find the optimal sensor layout using the approximation of this chapter.

Independently of the approximation, the formulas for the detection probability in Sec. 2.2.4 for one sensor located at $\mathbf{x}_s$ can be written

$$p(\mathbf{x}_s;f;c_t) = \int\int_A f(\mathbf{x}) M(\mathbf{x};\mathbf{x}_s;c_t) dxdy. \tag{3.20}$$

Using the approximation of [8], this reduces to integrating the leakage location probability density function over the monitored ellipse of the sensor:

$$p(\mathbf{x}_s;f;c_t) = \int\int_{\text{ellipse}} f(\mathbf{x}) dxdy. \tag{3.21}$$

Having one sensor, the detection probability is the probability contribution inside the monitored ellipse. Having several sensors, the detection probability is the total probability inside the monitored ellipses of all the sensors:

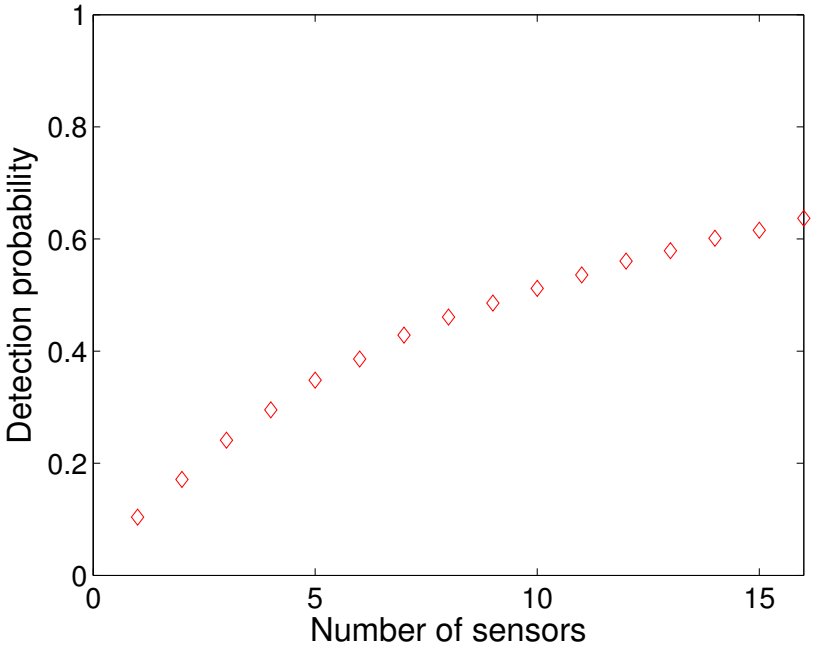


Figure 3.3: Detection probability as function of number of sensors for the test case in [8]. Concentration threshold $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}.

$$p(\mathbf{x}_s^1; \ldots; \mathbf{x}_s^N; f; c_t) = \int\int_{\text{all ellipses}} f(\mathbf{x}) dx dy. \tag{3.22}$$

Fig. 3.3 shows the detection probability for the optimal sensor layout as a function of the number of sensors for the test case. Fig. 3.4 shows the optimal layout of two sensors and Fig. 3.5 of four sensors. A threshold concentration of $c_t = 2.26 \cdot 10^{-4}$ kg/m^3 is used.

3.7 Discussion of the method

In this chapter we have presented the method of an approximated average concentration footprint from [8]. We will now look at the assumptions made, and discuss their impact on the resulting sensor layout and detection probability. Finally, improvements of the method are suggested.

Leakage location

When creating a monitoring program, the first thing that should be done is to identify the potential locations for leakages to be monitored and use this information to construct a leakage location probability density function (Sec. 2.1.1). To do this, a careful study of the storage formation and the overburden needs to be done to identify potential leakage pathways from the storage formation to the seafloor. This is not done in [8] where a probability

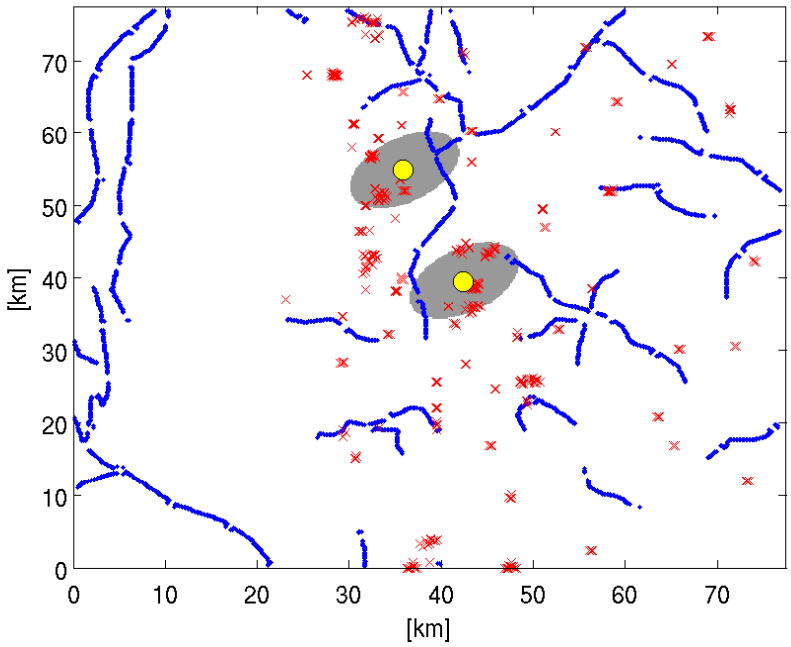


Figure 3.4: Optimal layout of two sensors (yellow circles) for the test case in [8]. The monitored region is shown in grey. Red crosses are wells and blue lines are faults. A threshold concentration of $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3} *is used. The detection probability is* $p = 0.17$.

density field for the leakage location is made based only on information about locations of faults and wells. There could be other pathways to the storage formation than faults and wells. On the other hand, all faults and wells may not reach the storage formation as discussed in [8]. In addition, it is natural to assume that faults and wells further away from the storage formation should have a lower probability of being the leakage location than those close to the formation. In [8], all faults and wells are given the same probability.

A study to find the leakage location probability density function is completely site dependent. We are not interested in site studies of specific storage sites, as we are studying methods to find the optimal sensor layout for any given leakage location probability density function. We will thus not study the leakage location probability density function in the following. In the further work, we will take the leakage location probability density functions as given and only use them as a test cases.

Approximation and inversion of footprint

Instead of using the simulated footprint directly and do an exact inversion of it, an approximation of the footprint is used for easy inversion. It is discussed in [8] that this approximation does not capture the anisotropy of the simulated footprint. Along the axes of the ellipses, the approximated concentration will be Gaussian. This implies that the approximation will impose symmetry about

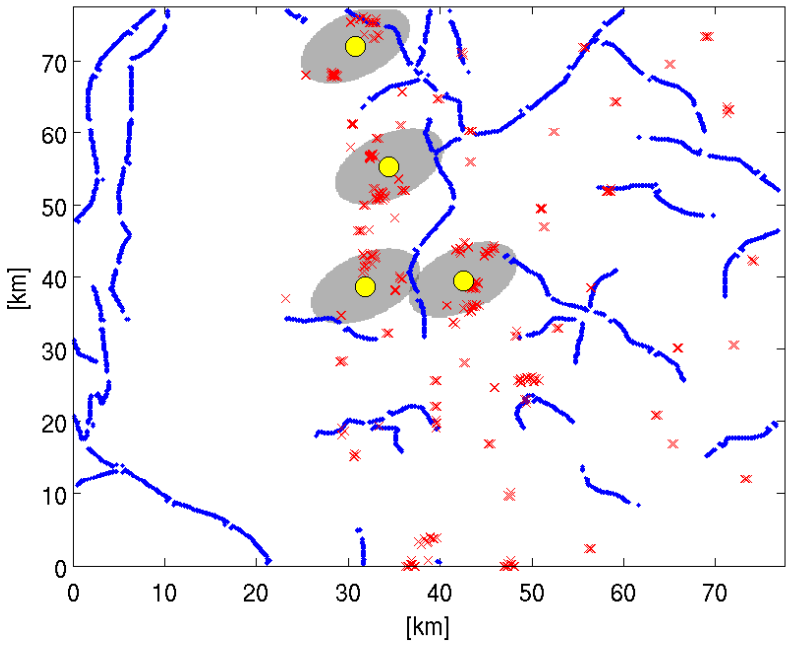


Figure 3.5: Optimal layout of four sensors (yellow circles) for the test case in [8]. The monitored region is shown in grey. Red crosses are wells and blue lines are faults. A threshold concentration of $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3} *is used. The detection probability is* $p = 0.30$.

both axes of the ellipses, which is a major simplification. For the test case, it is found that the Gaussian approximation will be too large close to the leakage, but too small further away. In Hvidevold et al. [8] it is suggested to try other approximation functions with skewness and slower decay. This has been done, but are not found to approximate the footprint significantly better.

For the method considered in [8], we are only interested in whether the concentration is below or above the threshold concentration c_t. The exact value of the approximated footprint at all locations may then not be too important. The most interesting question is how well the detectable region is approximated, since this is what is actually used in the computations. The detectable region for the footprint of Fig. 2.2 using the threshold concentration $c_t = 2.26 \cdot 10^{-4}$ kg/m^3 is shown in Fig. 3.2. We clearly see that the detectable region for the simulated footprint is much larger than for the approximated. In fact, the simulated footprint has a detectable region of 851 km^2, while the approximated has a detectable region of only 88 km^2. This makes the detectable region for the simulated footprint almost 10 times as large as for the approximated.

Having a detectable region so much larger for the simulated than for the approximated footprint, the approximation will have great effect on the resulting sensor layout and detection probability. A leakage is assumed detected in a much larger area than the approximation indicates. This implies that we can monitor the same area using fewer sensors than the approximation method

used in Hvidevold et al. [8] indicates. To obtain plausible results, we should use the simulated footprint directly without approximation. A method using exact inversion of the footprint is presented in Ch. 4.

Spatial independence of footprint

The leakage footprint is assumed to be the same independently of the leakage location. In real life, there will be spatial dependence on the leakage location in the footprint due to topography and current variations. Since the shape of the footprint is one of the main factors determining the sensor locations that will detect leakages, the simplification of spatial independence could alter the result much. However, including spatial dependence requires much more calculations. First of all, we would need to do simulations for many leakage locations, not only one as in [8]. Further, the inversion is much harder since the easy inversion used in [8] is based on the assumption of no spatial variation. A method incorporating spatial dependence is presented in Ch. 5.

Average concentration

The method presented in [8] uses average concentration, for the test case over a period of two months. If the average concentration is above the threshold concentration c_t, we assume that there is a leakage. If it is below, we assume that there is no leakage. This is a very conservative approach. In real life, we would not wait until the average concentration over two months is above the threshold before we assume there is a leakage. In addition, locations having an average concentration below the threshold concentration may have periods of time where the concentration is high enough that the leakage is detectable at that location. To get a more realistic detectable region, we can use concentration time series and extract other information than just the average. In Ch. 6, a method using events suggested in [7] is presented and discussed.

Chapter 4

Monitoring design: Translation of an average concentration footprint

Ch. 3 presented the work done in Hvidevold et al. [8] based on approximation of the average concentration footprint of a leakage. The approximation was discussed in Sec. 3.7 where it was argued that the exact footprint should be used directly without approximations. In this chapter, we develop a method based on an exact inversion of the footprint. We will test the method on the same test case as in [8] and compare with the approximation of Ch. 3.

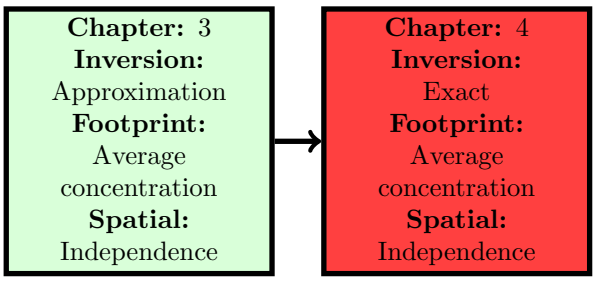


Figure 4.1: Summary of the different methods for monitoring design, the red is the one to be developed in this chapter.

4.1 Estimation of the leakage footprint

As in Ch. 3, we will use the average concentration footprint from a leakage. Let $\bar{C}(\mathbf{x};\mathbf{x}_0)$ be the average concentration at $\mathbf{x}$ resulting from a leakage at $\mathbf{x}_0$. We assume as before that the average concentration footprint is independent of the leakage location. This implies that we only need one predicted footprint. Let $\mathbf{x}^*$ be the leakage location of the predicted footprint such that $\bar{C}(\mathbf{x};\mathbf{x}^*)$ is the average concentration footprint predicted from a simulation. An arbitrary leakage location $\mathbf{x}_0$ is assumed to have the same footprint, only translated to

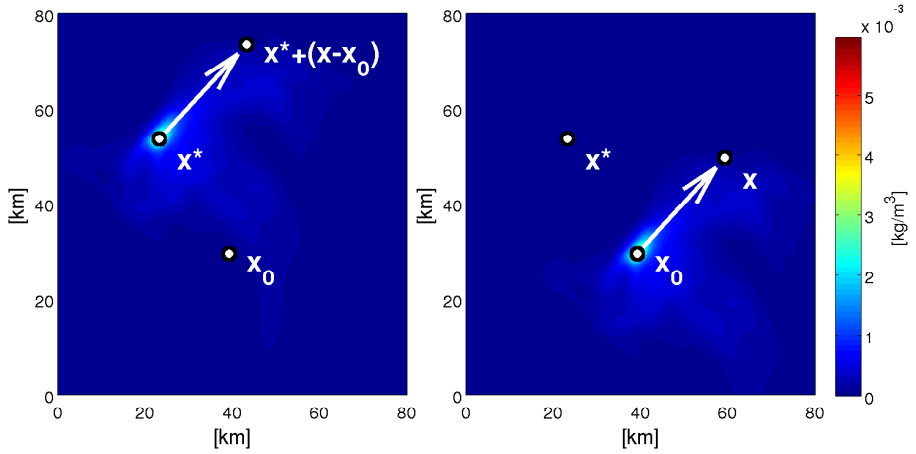


Figure 4.2: The predicted concentration $\bar{C}(\mathbf{x};\mathbf{x}^*)$ from a leakage at $\mathbf{x}^*$ to the left. The translated concentration field $\bar{C}(\mathbf{x};\mathbf{x}_0)$ from a leakage at $\mathbf{x}_0$ to the right. Since the footprints are equal, we see that $\bar{C}(\mathbf{x};\mathbf{x}_0) = \bar{C}(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*)$.

the correct leakage location. The translated average concentration field $\bar{C}(\mathbf{x};\mathbf{x}_0)$ resulting from a leakage at $\mathbf{x}_0$ is then given by

$$\bar{C}(\mathbf{x};\mathbf{x}_0) = \bar{C}(\mathbf{x} - (\mathbf{x}_0 - \mathbf{x}^*);\mathbf{x}^*) = \bar{C}(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*). \tag{4.1}$$

The translation property is illustrated in Fig. 4.2.

4.2 Detection function for translated average

We have the same detection function as in Sec. 3.1. Having a leakage at $\mathbf{x}_0$, a sensor at $\mathbf{x}$ will experience the average concentration $\bar{C}(\mathbf{x};\mathbf{x}_0)$. If the concentration is above the threshold concentration c_t, the sensor is assumed to always detect the leakage. If it is below, the sensor is assumed to never detect the leakage. The detection function for a leakage at $\mathbf{x}_0$ is thus given by

$$D_A(\mathbf{x};\mathbf{x_0};c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x};\mathbf{x}_0) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x};\mathbf{x}_0) < c_t \end{cases}. \tag{3.1}$$

Using the translation relation of Eq. 4.1, we can express the detection function $D_A^T(\mathbf{x};\mathbf{x}_0;c_t)$ for the translated average concentration in terms of the detection function for the predicted footprint:

$$D_A^T(\mathbf{x};\mathbf{x}_0;c_t) = D_A(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*;c_t), \tag{4.2}$$

$$D_A^T(\mathbf{x};\mathbf{x}_0;c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*) < c_t \end{cases}. \tag{4.3}$$

The corresponding detectable region will be

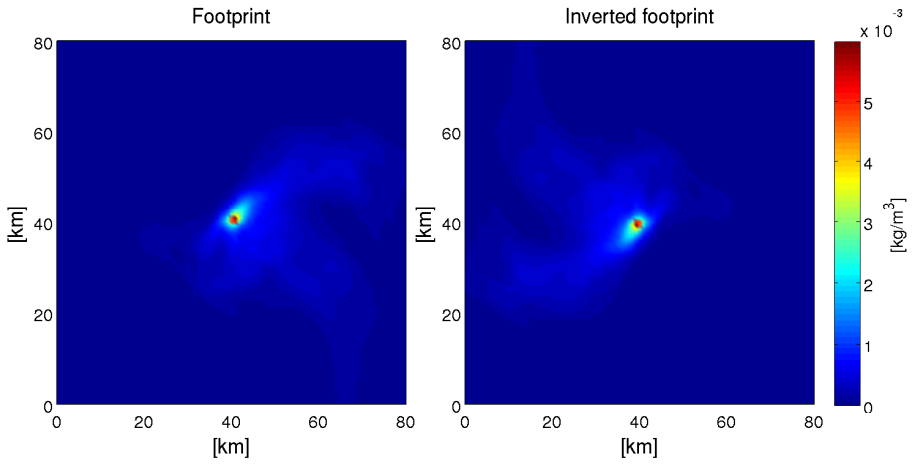


Figure 4.3: The average concentration footprint used in [8] to the left. Its corresponding inverse to the right, illustrating the inversion by rotating 180°.

$$\boxed{\text{Detectable region translated average}} = \{\mathbf{x} \,|\bar{C}(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}^*) \geq c_t\} \tag{4.4}$$

4.3 Monitoring function for translated average

As earlier, the monitoring function is given by the detection function:

$$M(\mathbf{x}; \mathbf{x}_s; c_t) = D(\mathbf{x}_s; \mathbf{x}; c_t) \tag{2.4}$$

This implies that the monitoring function will be the same as in Sec. 3.2:

$$M_A(\mathbf{x}; \mathbf{x}_s; c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x}_s; \mathbf{x}) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x}_s; \mathbf{x}) < c_t \end{cases}. \tag{3.3}$$

The monitoring function corresponds to the assumption that a sensor at $\mathbf{x}_s$ always will detect a leakage at a location giving an average concentration above the threshold concentration c_t at $\mathbf{x}_s$. Leakages at locations giving an average concentration below c_t at $\mathbf{x}_s$ is assumed to never be detected.

As for the detection function, we express the monitoring function in terms of the predicted footprint. Using Eq. 2.4, 4.2 and 4.3, we obtain

$$M_A^T(\mathbf{x}; \mathbf{x}_s; c_t) = D_A(\mathbf{x}^* + (\mathbf{x}_s - \mathbf{x}); \mathbf{x}^*; c_t) = D_A(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*; c_t) \tag{4.5}$$

$$M_A^T(\mathbf{x}; \mathbf{x}_s; c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*) < c_t \end{cases}. \tag{4.6}$$

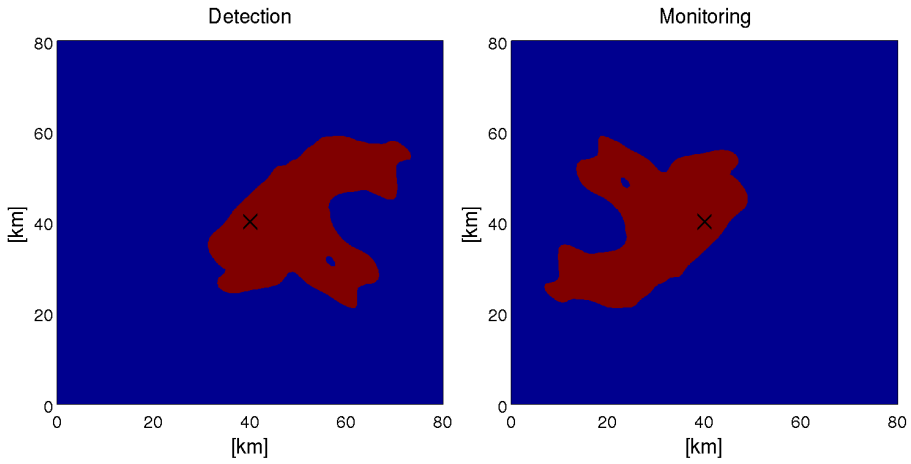


Figure 4.4: Detection function/detectable region for the footprint used in [8] to the left ($c_t = 2.26 \cdot 10^{-4}$ kg/m^3). The corresponding monitoring function/monitored region to the right. The leakage/sensor location is marked with a black cross. It is easy to see that the monitoring function is obtained by rotating the detection function 180° *about the leakage/sensor location.*

The corresponding monitored region will be

$$\boxed{\begin{matrix}\text{Monitored region} \\ \text{translated average}\end{matrix}} = \left\{ \mathbf{x} \,\middle|\, \bar{C}(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*) \geq c_t \right\}. \tag{4.7}$$

4.4 Inversion by rotation

Calculating the monitoring function can be looked at as inverting the detection function. The inversion property can be seen by the much used relation

$$M(\mathbf{x}; \mathbf{x}_s; c_t) = D(\mathbf{x}_s; \mathbf{x}; c_t). \tag{2.4}$$

In the calculations of the monitoring function, we are really inverting the average concentration footprint. We know the average concentration $\bar{C}(\mathbf{x}; \mathbf{x}_0)$ at $\mathbf{x}$ from a leakage at $\mathbf{x}_0$ for all $\mathbf{x}$ and $\mathbf{x}_0$. Due to the assumption of all footprints being equal, the average concentration field can be expressed in terms of the predicted footprint having leakage location at $\mathbf{x}^*$ as shown by Eq. 4.1:

$$\bar{C}(\mathbf{x}; \mathbf{x}_0) = \bar{C}(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}^*) \tag{4.8}$$

When calculating the monitoring function we need the average concentration $\bar{C}(\mathbf{x}_s; \mathbf{x})$ at $\mathbf{x}_s$ from a leakage at $\mathbf{x}$ for all $\mathbf{x}_s$ and $\mathbf{x}$. As used for the monitoring function (Eq. 4.6), this is found by plugging $\mathbf{x}_s$ and $\mathbf{x}$ into Eq. 4.8:

$$\bar{C}(\mathbf{x}_s; \mathbf{x}) = \bar{C}(\mathbf{x}^* + (\mathbf{x}_s - \mathbf{x}); \mathbf{x}^*) = \bar{C}(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*). \tag{4.9}$$

Similarly, the detection function for arbitrary $\mathbf{x}$ and $\mathbf{x}_0$ can be expressed

$$D_A^T(\mathbf{x}; \mathbf{x}_0; c_t) = D_A(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}^*; c_t). \tag{4.2}$$

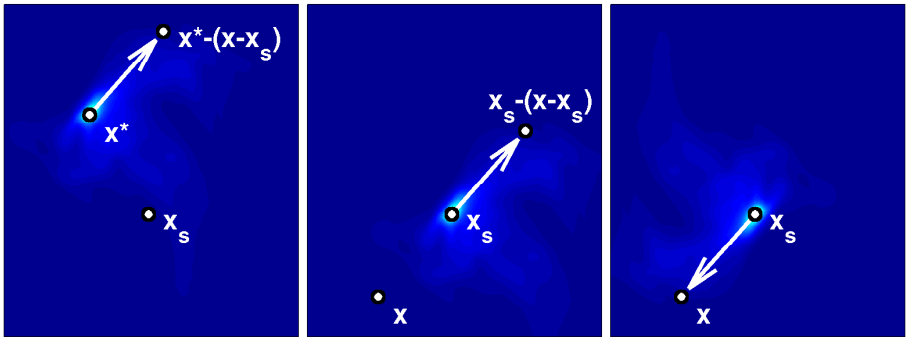


Figure 4.5: Showing the steps in obtaining the inverted footprint $\bar{C}(\mathbf{x}_s;\mathbf{x})$. The predicted footprint $\bar{C}(\mathbf{x};\mathbf{x}^*)$ is shown to the left. In the middle, the predicted footprint is translated to $\mathbf{x}_s$ to obtain the footprint $\bar{C}(\mathbf{x};\mathbf{x}_s)$. Finally, the footprint is rotated 180° to obtain the inverted footprint $\bar{C}(\mathbf{x}_s;\mathbf{x})$ to the right. We see that $\bar{C}(\mathbf{x}_s;\mathbf{x}) = \bar{C}(\mathbf{x}_s - (\mathbf{x}-\mathbf{x}_s);\mathbf{x}_s) = \bar{C}(\mathbf{x}^* - (\mathbf{x}-\mathbf{x}_s);\mathbf{x}^*)$.

This gives the monitoring function

$$M_A^T(\mathbf{x};\mathbf{x}_s;c_t) = D_A(\mathbf{x}^* - (\mathbf{x}-\mathbf{x}_s);\mathbf{x}^*;c_t) \tag{4.10}$$

We now want to compare the concentration field $\bar{C}(\mathbf{x};\mathbf{x}_0)$ with the inverted concentration field $\bar{C}(\mathbf{x}_s;\mathbf{x})$. Similarly, we want to compare the detection function $D_A^T(\mathbf{x};\mathbf{x}_0;c_t)$ with the monitoring function $M_A^T(\mathbf{x};\mathbf{x}_s;c_t)$. To do this, we assume that $\mathbf{x}_0 = \mathbf{x}_s$. We define the displacement vector $\mathbf{r}$ by

$$\mathbf{r} = \mathbf{x} - \mathbf{x}_0 = \mathbf{x} - \mathbf{x}_s. \tag{4.11}$$

We can then express the concentrations and the functions:

$$\bar{C}(\mathbf{x};\mathbf{x}_0) = \bar{C}(\mathbf{x}^* + \mathbf{r};\mathbf{x}^*) \tag{4.12}$$
$$\bar{C}(\mathbf{x}_s;\mathbf{x}) = \bar{C}(\mathbf{x}^* - \mathbf{r};\mathbf{x}^*) \tag{4.13}$$
$$D_A^T(\mathbf{x};\mathbf{x}_0;c_t) = D_A(\mathbf{x}^* + \mathbf{r};\mathbf{x}^*;c_t) \tag{4.14}$$
$$M_A^T(\mathbf{x};\mathbf{x}_s;c_t) = D_A(\mathbf{x}^* - \mathbf{r};\mathbf{x}^*;c_t). \tag{4.15}$$

We see that the sign of $\mathbf{r}$ is the only difference between the concentration field and the inverted concentration field. In the same way, the sign of $\mathbf{r}$ is the only difference between the detection function and the monitoring function. The concentration field can thus be inverted by changing the sign of the displacement vector $\mathbf{r}$, and the detection function can be inverted into the monitoring function by the same sign change. Changing the sign of $\mathbf{r}$ corresponds to a reflection about $\mathbf{x}^*$, which can be expressed using a matrix R:

$$\bar{C}(\mathbf{x};\mathbf{x}_0) = \bar{C}(\mathbf{x}^* + \mathbf{r};\mathbf{x}^*) \tag{4.16}$$
$$\bar{C}(\mathbf{x}_s;\mathbf{x}) = \bar{C}(\mathbf{x}^* + R\mathbf{r};\mathbf{x}^*) \tag{4.17}$$
$$D_A^T(\mathbf{x};\mathbf{x}_0;c_t) = D_A(\mathbf{x}^* + \mathbf{r};\mathbf{x}^*;c_t) \tag{4.18}$$
$$M_A^T(\mathbf{x};\mathbf{x}_s;c_t) = D_A(\mathbf{x}^* + R\mathbf{r};\mathbf{x}^*;c_t). \tag{4.19}$$

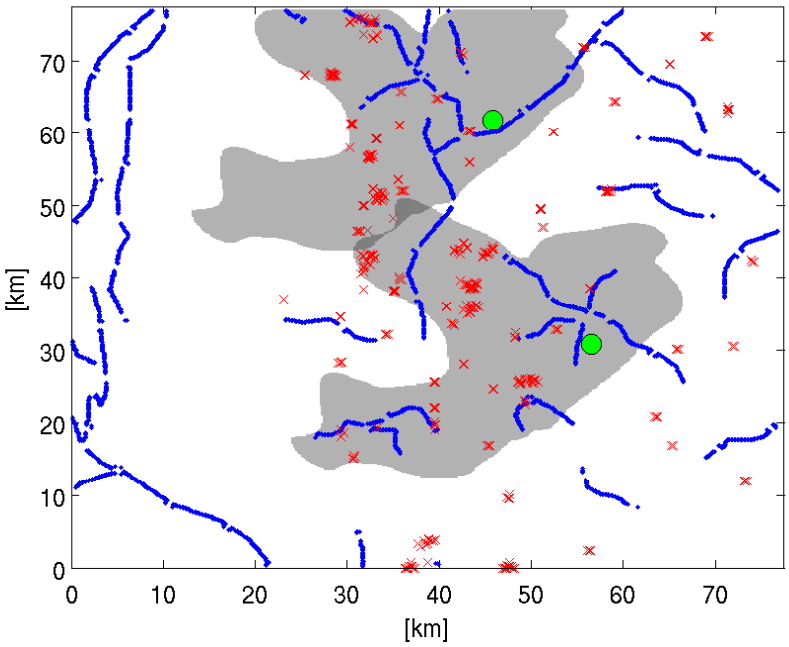


Figure 4.6: Optimal layout of two sensors (green circles) for the test case in [8]. The monitored region is shown in grey, darker areas are monitored by both sensors. Red crosses are wells and blue lines are faults. Concentration threshold $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}*. The detection probability is* $p = 0.57$*.*

Since the vectors are in $\mathbb{R}^2$, R is the negative 2x2 identity matrix

$$R = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.20}$$

A counter-clockwise rotation of α degrees is given by the matrix

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}. \tag{4.21}$$

We see that the matrix R is the rotation matrix for $\alpha = \pi = 180°$. The concentration field and the detection function can thus be inverted by a rotation of 180°. It is important to note that this property is completely dependent on the assumption of spatial independence of the footprint, and thus spatial independence of the detection function. Fig. 4.3 and 4.4 shows the inverting by rotating 180° property for the test test case of [8]. The steps of inverting the footprint and the resulting coordinate mapping is illustrated in Fig. 4.5.

4.5 Extended application of the inversion

In this chapter, we have developed an exact inversion using an average concentration footprint from a leakage. The method can of course be applied

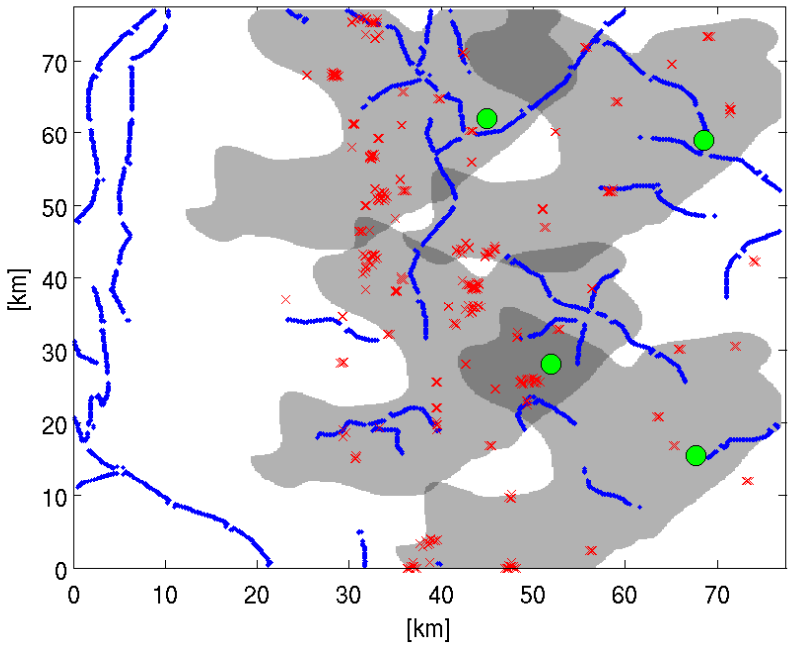


Figure 4.7: Optimal layout of four sensors (green circles) for the test case in [8]. The monitored region is shown in grey, darker areas are monitored by several sensors. Red crosses are wells and blue lines are faults. Concentration threshold $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}*. The detection probability is* $p = 0.79$*.*

to other input files, as will be done in later chapters where the theory of this chapter is used to develop other methods for monitoring design.

The inversion by rotation developed in this chapter can be applied to any kind of footprint, concentration or not. The detection function used in this chapter is an example that the method is applied to. The only assumption that must be made is that the footprint is independent of the leakage location and is given by a single predicted footprint.

Let $g(\mathbf{x};\mathbf{x}_0)$ be the value at $\mathbf{x}$ of an arbitrary footprint having leakage location $\mathbf{x}_0$. Further, let $\mathbf{x}^*$ be the leakage location of the predicted footprint giving $g(\mathbf{x};\mathbf{x}^*)$. Due to the assumption of spatial independence, we can express $g(\mathbf{x};\mathbf{x}_0)$ in terms of the predicted footprint as we did for the average concentration footprint in Eq. 4.1:

$$g(\mathbf{x};\mathbf{x}_0) = g(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*). \tag{4.22}$$

Here, we think of $\mathbf{x}_0$ as a fixed variable such that we get the footprint value $g(\mathbf{x};\mathbf{x}_0)$ at all locations $\mathbf{x}$ for a given leakage location. When inverting the footprint, we are only changing the fixed and the independent variable. Having a sensor at $\mathbf{x}_s$, we are interested in $g(\mathbf{x}_s;\mathbf{x})$. This is the footprint value at $\mathbf{x}_s$ for all leakage locations $\mathbf{x}$. As indicated, $\mathbf{x}_s$ is now a fixed variable. Using Eq. 4.22, we obtain the inverted footprint similarly to Eq. 4.9:

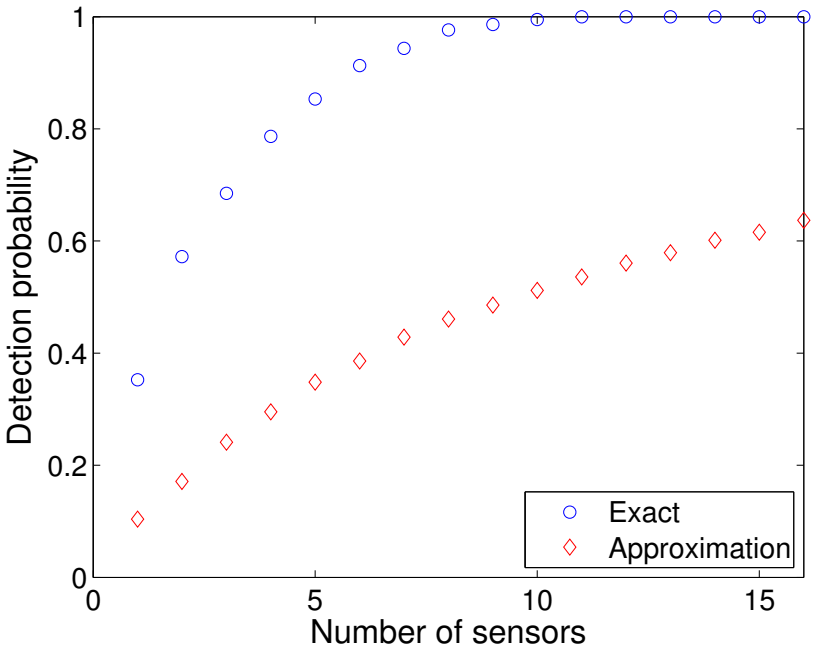


Figure 4.8: Detection probability for the optimal layout as function of number of sensors for the test case in [8] comparing the approximation of Ch. 3 and the exact inversion of this chapter. Concentration threshold $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}.

$$g(\mathbf{x}_s; \mathbf{x}) = g(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*). \tag{4.23}$$

As seen from Eq. 4.22 and 4.23, the footprint and the inverted footprint have the same form as what we obtained for the average concentration footprints. In these calculations however, average concentration was never used. This implies that the average concentration is just a special case where the inversion method can be used. As stated before, we will in the next chapters use the inversion with other inputs than in this chapter. For numerical implementations of the methods developed, the interpretation of inverting a footprint by rotating 180° is extremely useful.

4.6 Optimal design

We test the method of this chapter on the same test case as in Ch. 3. This is the test case used in Hvidevold et al. [8] and presented in Sec. 2.1.3. The average concentration footprint and its inverse are shown in Fig. 4.3. The corresponding detection and monitoring function using $c_t = 2.26 \cdot 10^{-4}$ kg/m^3 are shown in Fig. 4.4. We see that the shape of the monitored region is far from the ellipse approximation used in Ch. 3.

Optimal placements and monitored regions for two and four sensors using the approximation of [8] are shown in Fig. 3.4 and 3.5. The corresponding

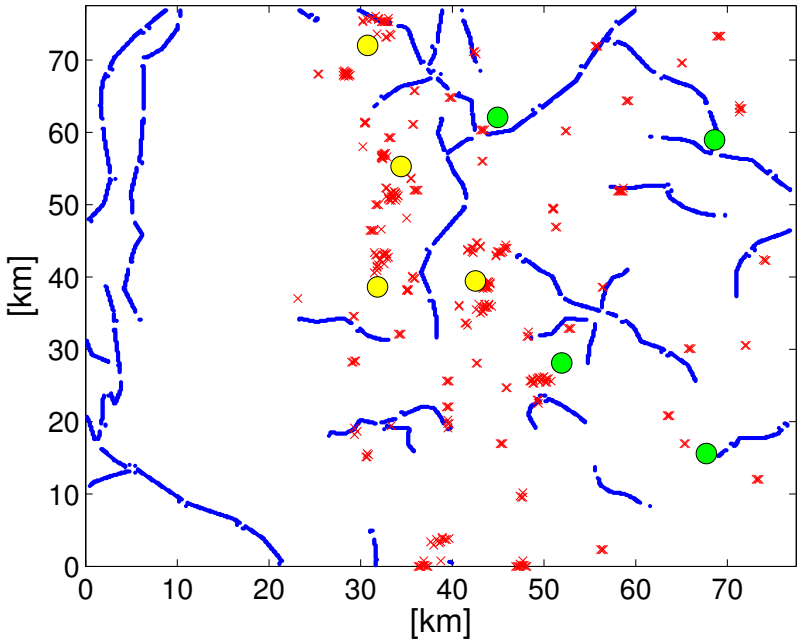


Figure 4.9: Optimal layout of four sensors for the test case in [8] using the approximation (yellow) and the exact (green) inversion. Red crosses are wells and blue lines are faults. Threshold concentration of $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}.

optimal placements and monitored regions for two and four sensors using the exact inversion are shown in Fig. 4.6 and 4.7. Fig. 4.9 shows the placement of four sensors for both methods for comparison. Fig. 4.8 shows the detection probability for the optimal layout as a function of the number of sensors for both methods.

4.7 Discussion of the method

The purpose of this chapter has been to develop a method inverting the average concentration footprint exactly instead of using the approximation of Ch. 3. The exact inversion can be used for other footprints too, which will be done in later chapters. We will now do a short discussion of the assumptions addressed in Sec. 3.7 for the method of this chapter.

Leakage location

As discussed in Sec. 3.7, we think of the leakage location probability density function only as a test case. We are not interested in special site studies, implying that when developing the method we can think of the leakage location probability density function as given.

Spatial independence of footprint

The method developed in this chapter is only using one concentration footprint as in Ch. 3. This leads to the assumption of spatial independence of the leakage location in the footprint, which the method of this chapter is based on. As discussed in Sec. 3.7, we would like to include spatial dependence in the footprint. A method for this is presented in Ch. 5.

Approximation and inversion of footprint

The method developed in this chapter enables us to use the predicted footprint directly without the approximation of Ch. 3. In addition, the method is just as easy to implement as the approximation method. This is mainly due to the easy inversion of a footprint by rotating it 180°. We see from Fig. 4.9 that the optimal sensor layout for the test case becomes significantly different using the exact footprint instead of the approximation. At the same time, Fig. 4.8 shows that we get a much higher detection probability using the exact method of this chapter. In the following work, we will only use the exact inversion.

Average concentration

We are still using an average concentration footprint as in Ch. 3, only without the approximation. The discussion on average concentration and alternative approaches of Sec. 3.7 is then still valid for the method of this chapter. An alternative approach using events will be presented in Ch. 6.

Chapter 5

Monitoring design: Interpolation of average concentration footprints

This far we have assumed the average concentration footprint of a leakage to be independent of the leakage location. Due to the topography of the seafloor and current variations, there will in real life be variations in the footprint for different leakage locations as discussed in Sec. 3.7.

We now want to include spatial dependence in our calculations such that all leakage locations have different footprints. Ideally we would like to do simulations for every grid point as leakage location, but this is not possible due to calculation and memory limitations. Instead we use a smaller number of simulated footprints and interpolate. This leads to the red method in Fig. 5.1.

Figure 5.1: Summary of the different methods for monitoring design, the red is the one to be developed in this chapter.

5.1 Estimation of the leakage footprint

As before, let $\bar{C}(\mathbf{x};\mathbf{x}_0)$ be the average concentration at $\mathbf{x}$ resulting from a leakage at $\mathbf{x}_0$. Assume that we have K predicted footprints. For each footprint k, let $\mathbf{x}_k^*$ be the leakage location such that $\bar{C}(\mathbf{x};\mathbf{x}_k^*)$ is predicted from a simulation. We want to interpolate between these predicted footprints to obtain

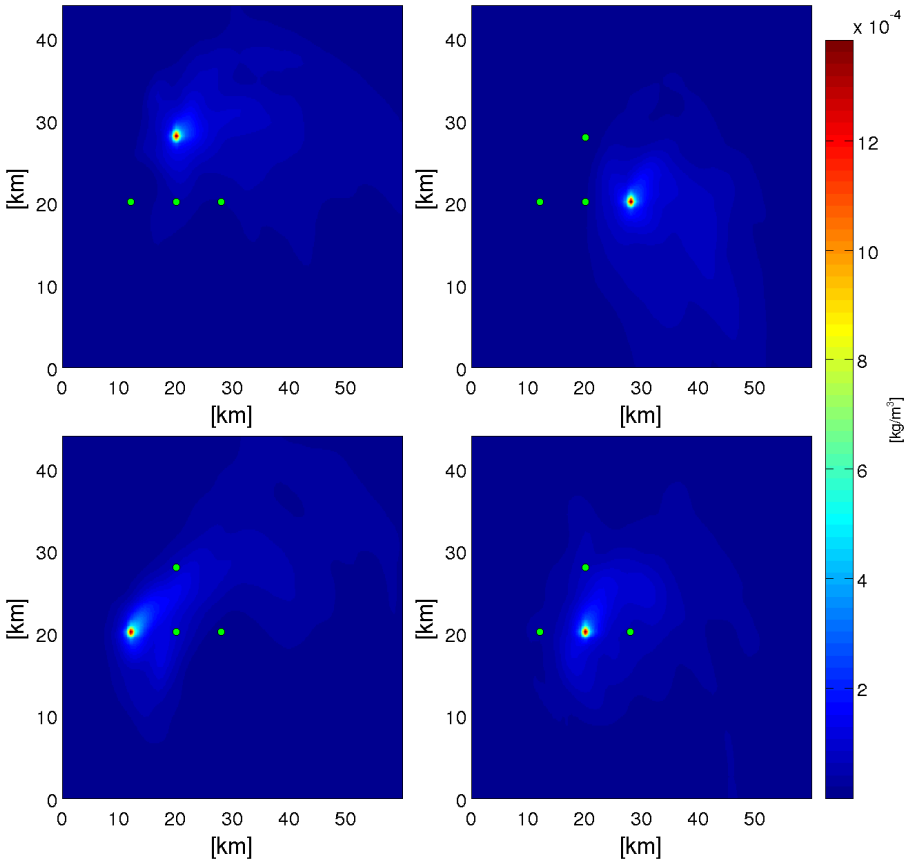


Figure 5.2: The four test case footprints showing the spatial dependence. The leakage locations of the other footprints are marked with green circles.

a footprint for an arbitrary leakage location $\mathbf{x}_0$. The first thing we need to do is to translate the footprints to have leakage location in $\mathbf{x}_0$ as in Ch. 4. The average concentration at $\mathbf{x}$ for a leakage at $\mathbf{x}_0$ with footprint k is given by

$$\bar{C}(\mathbf{x};\mathbf{x}_0) = \bar{C}(\mathbf{x}_k^* + (\mathbf{x}-\mathbf{x}_0);\mathbf{x}_k^*). \tag{5.1}$$

For each possible leakage position $\mathbf{x}_0$, we use a weighted sum of these translated footprints to obtain a concentration field:

$$\bar{C}(\mathbf{x};\mathbf{x}_0) = \sum_{k=1}^{K} w_k(\mathbf{x}_0)\bar{C}(\mathbf{x}_k^* + (\mathbf{x}-\mathbf{x}_0);\mathbf{x}_k^*) \tag{5.2}$$

where $w_k(\mathbf{x}_0)$ is the weight used for footprint k when approximating the footprint of a leakage in $\mathbf{x}_0$. For consistency, the weights must satisfy the normalization

$$\sum_{k=1}^{K} w_k(\mathbf{x}_0) = 1 \qquad \forall \mathbf{x}_0. \tag{5.3}$$

We assume that the weight functions only depend on the distances from $\mathbf{x}_0$ to the leakage locations $\{\mathbf{x}_k^*\}$ of the predicted footprints. Let $d_k(\mathbf{x}_0)$ be the

distance from $\mathbf{x}_0$ to the leakage point $\mathbf{x}_k^*$ of the predicted footprint k:

$$d_k(\mathbf{x}_0) = ||\mathbf{x}_0 - \mathbf{x}_k^*||_2 = \sqrt{(x_0 - x_k^*)^2 + (y_0 - y_k^*)^2}. \tag{5.4}$$

Further, assume that the weights are inversely proportional to the distances:

$$w_k(\mathbf{x}_0) = \frac{B(\mathbf{x}_0)}{d_k(\mathbf{x}_0)} \tag{5.5}$$

where $B(\mathbf{x}_0)$ is used for normalization to satisfy Eq. 5.3. This gives us

$$B(\mathbf{x}_0) = \frac{1}{\sum_{k=1}^{K} \frac{1}{d_k(\mathbf{x}_0)}}. \tag{5.6}$$

If one of the distances is zero, we are located in the leakage point of one of the predicted footprints. The weights are then set such that only the corresponding footprint is used, avoiding division by zero. This gives us

$$w_k(\mathbf{x}_0) = \begin{cases} 0 & \text{if } d_k(\mathbf{x}_0) > 0 \text{ and } \min(d_1(\mathbf{x}_0), \ldots, d_K(\mathbf{x}_0)) = 0 \\ 1 & \text{if } d_k(\mathbf{x}_0) = 0 \\ \frac{B(\mathbf{x}_0)}{d_k(\mathbf{x}_0)} & \text{if } \min(d_1(\mathbf{x}_0), \ldots, d_K(\mathbf{x}_0)) > 0 \end{cases}. \tag{5.7}$$

Fig. 5.2 shows the simulated average concentration footprint for four different leakage locations at the test site Sleipner, showing the great spatial variation between the four locations. For the test case of this chapter, the footprint of any location will be estimated by interpolation between these four footprints. The resulting interpolation weights from Eq. 5.7 are shown in Fig. 5.3. The resulting footprint for an arbitrary leakage location is shown in Fig. 5.4. Note that the footprints of Fig. 5.2 are from a different leakage scenario (Tab. C.1) than the footprint in Fig. 2.2 used in Ch. 3 and 4.

5.2 Detection function for interpolation

Using the average concentration resulting from leakages, we have the same detection function and corresponding detectable region as before:

$$D_A(\mathbf{x}; \mathbf{x}_0; c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x}; \mathbf{x}_0) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x}; \mathbf{x}_0) < c_t \end{cases} \tag{3.1}$$

$$\boxed{\text{Detectable region average}} = \left\{ \mathbf{x} \,\middle|\, \bar{C}(\mathbf{x}; \mathbf{x}_0) \geq c_t \right\}. \tag{3.2}$$

By using the interpolated concentration of Eq. 5.2 and Eq. 3.1, we obtain

$$D_A^I(\mathbf{x}; \mathbf{x}_0; c_t) = \begin{cases} 1 & \text{if } \sum_{k=1}^{K} w_k(\mathbf{x}_0) \bar{C}(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*) \geq c_t \\ 0 & \text{if } \sum_{k=1}^{K} w_k(\mathbf{x}_0) \bar{C}(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*) < c_t \end{cases} \tag{5.8}$$

$$\boxed{\begin{array}{c}\text{Detectable region} \\ \text{interpolated average}\end{array}} = \left\{ \mathbf{x} \,\middle|\, \sum_{k=1}^{K} w_k(\mathbf{x}_0) \bar{C}(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*) \geq c_t \right\}. \tag{5.9}$$

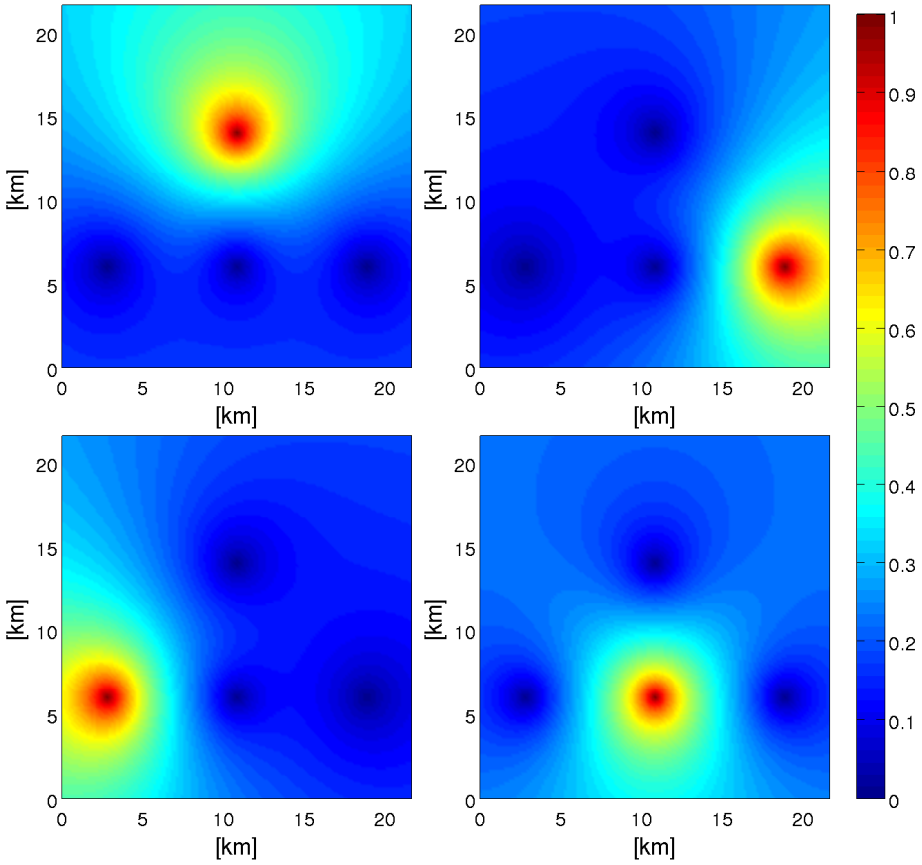


Figure 5.3: The interpolation weights (Eq. 5.7) when predicting the footprint of an arbitrary leakage location using the four footprints in Fig. 5.2.

Due to the spatial dependence in the footprint, the detectable region will have different shape and size for different leakage locations. This implies that we don't have the simple translation property of Eq. 4.2 where the detection function is expressed using the detection function for the predicted footprint. Fig. 5.5 shows the detectable regions of the footprints in Fig. 5.2 for concentration thresholds $c_t = 2.26 \cdot 10^{-4}$ kg/m^3 and $c_t = 9 \cdot 10^{-5}$ kg/m^3.

5.3 Monitoring function for interpolation

Since the monitoring function is implicitly defined by the detection function using Eq. 2.4, we also have the same monitoring function and monitored region as before for the average concentration:

$$M_A(\mathbf{x}; \mathbf{x}_s; c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x}_s; \mathbf{x}) \geq c_t \\ 0 & \text{if } \bar{C}(\mathbf{x}_s; \mathbf{x}) < c_t \end{cases} \tag{3.3}$$

$$\boxed{\text{Monitored region average}} = \left\{ \mathbf{x} \,\middle|\, \bar{C}(\mathbf{x}_s; \mathbf{x}) \geq c_t \right\}. \tag{3.4}$$

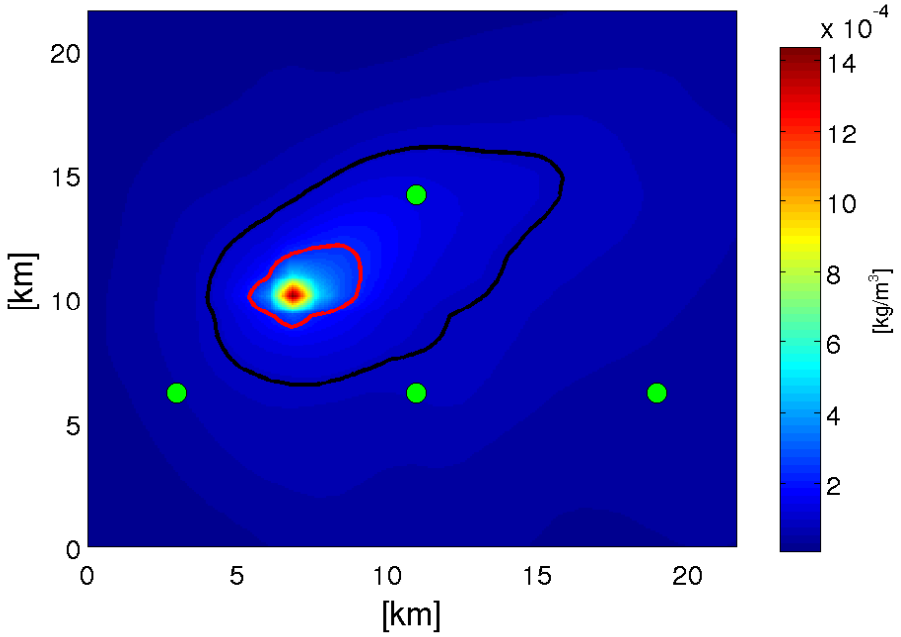


Figure 5.4: Footprint of an arbitrary leakage location using the footprints in Fig. 5.2 and the weights in Fig. 5.3. The leakage locations of the predicted footprints are marked with green circles. The black contour is the level curve for $c_t = 9 \cdot 10^{-5}$ *kg/m*^{3} *and the red contour for* $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}*, giving the boundary of the detectable regions.*

Using the interpolated concentration of Eq. 5.2, we obtain the following expression for the inverted concentration field $\bar{C}(\mathbf{x}_s; \mathbf{x})$:

$$\bar{C}(\mathbf{x}_s; \mathbf{x}) = \sum_{k=1}^{K} w_k(\mathbf{x}) \bar{C}(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*). \tag{5.10}$$

By comparing with Eq. 4.9, we see that this is a weighted sum of footprints inverted the same way as in Ch. 4. Note that while the weight functions had the fixed values $w_k(\mathbf{x}_0)$ for the interpolated concentration of Eq. 5.2, this is not the case here. The independent variable $\mathbf{x}$ is present in both the weight functions and the predicted footprints, making the calculation of the inverted footprint more complicated. The monitoring function and its corresponding monitored region using Eq. 5.10 become

$$M_A^I(\mathbf{x}; \mathbf{x}_s; c_t) = \begin{cases} 1 & \text{if } \sum_{k=1}^{K} w_k(\mathbf{x}) \bar{C}(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*) \geq c_t \\ 0 & \text{if } \sum_{k=1}^{K} w_k(\mathbf{x}) \bar{C}(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*) < c_t \end{cases} \tag{5.11}$$

$$\boxed{\begin{array}{c}\text{Monitored region} \\ \text{interpolated average}\end{array}} = \left\{ \mathbf{x} \,\middle|\, \sum_{k=1}^{K} w_k(\mathbf{x}) \bar{C}(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*) \geq c_t \right\}. \tag{5.12}$$

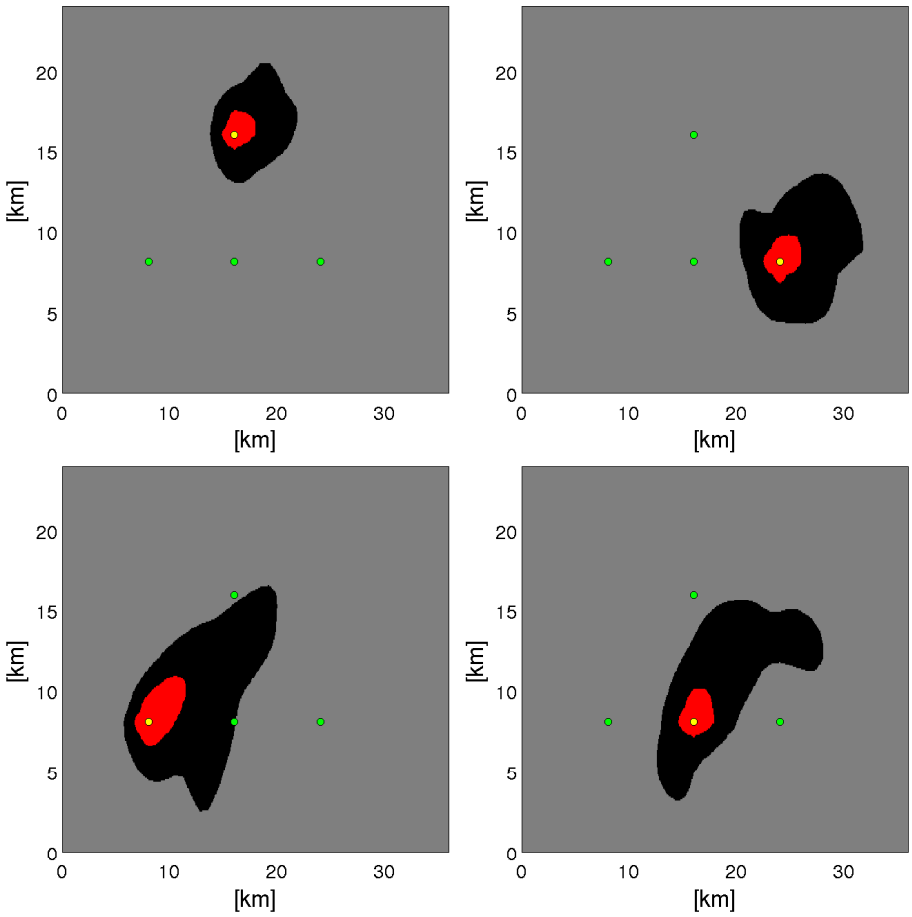


Figure 5.5: Detectable regions for the four footprints in Fig. 5.2. The leakage location is marked with a yellow circle and the leakage points of the other footprints with green circles. The black (including the red) is for $c_t = 9 \cdot 10^{-5}$ *kg/m*^{3} *and the red for* $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}*. Note the large spatial variation.*

As for the detectable region, the spatial dependence in the footprint will cause the monitored region to have different shape and size for different sensor locations. Fig. 5.6 shows the monitored region for an arbitrary sensor location using the footprints of Fig. 5.2.

5.4 Extended application of the spatial variation

The spatial dependence can be applied to any footprint, just as the inversion of Ch. 4. We will now briefly show how to do this for other inputs than average concentration, as we did for the inversion in Sec. 4.5. This method will be used for a different footprint in Ch. 7.

Assume as in Sec. 4.5 that $g(\mathbf{x}; \mathbf{x}_0)$ is the value at $\mathbf{x}$ of some arbitrary footprint having leakage location $\mathbf{x}_0$. We want to estimate the footprint of an

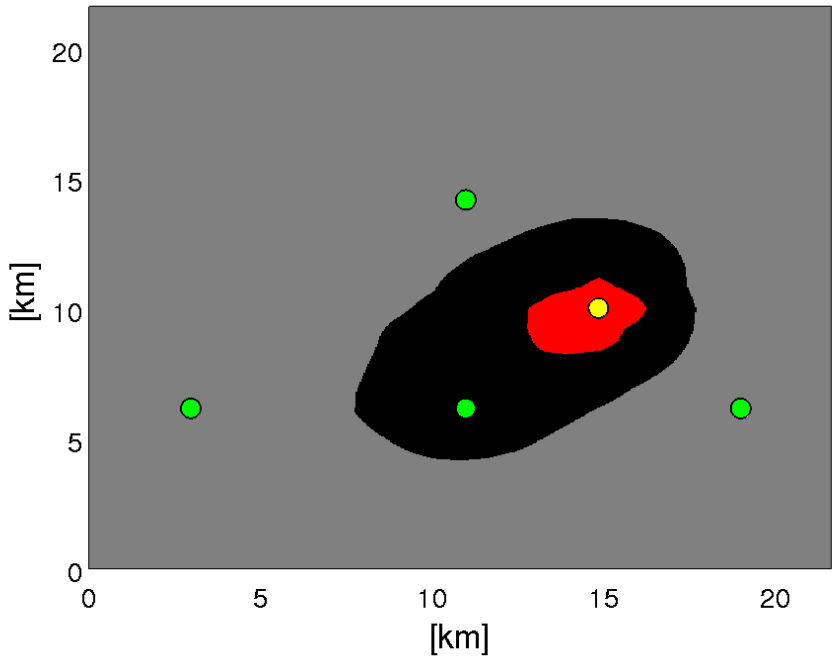


Figure 5.6: Monitored regions for an arbitrary sensor location (yellow circle) using the footprint in Fig. 5.2. The green circles show the leakage location of the four predicted footprints. The black (including the red) is for $c_t = 9 \cdot 10^{-5}$ *kg/m*^{3} *and the red for* $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}.

arbitrary leakage location based on interpolation between K predicted footprints $g(\mathbf{x}; \mathbf{x}_k^*)$ having leakage locations $\mathbf{x}_1^*; \ldots; \mathbf{x}_K^*$. The footprint of an arbitrary leakage location will then be given by the weighted sum equivalent to Eq. 5.2:

$$g(\mathbf{x}; \mathbf{x}_0) = \sum_{k=1}^{K} w_k(\mathbf{x}_0) g(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*) \tag{5.13}$$

where the weight functions w_k are given by Eq. 5.7. The inverted footprint $g(\mathbf{x}_s; \mathbf{x})$ for a fixed sensor at $\mathbf{x}_s$ will be given by the equivalent of Eq. 5.10:

$$g(\mathbf{x}_s; \mathbf{x}) = \sum_{k=1}^{K} w_k(\mathbf{x}) g(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*). \tag{5.14}$$

We see that Eq. 5.13 and 5.14 are completely equivalent to Eq. 5.2 and 5.10 for the average concentration footprint, implying that the average concentration footprint of this chapter only is an example of a footprint with spatial dependence.

5.5 Optimal design

Footprints are simulated for four different leakage positions. The footprints are shown in Fig. 5.2 and the corresponding detectable regions in Fig. 5.5.

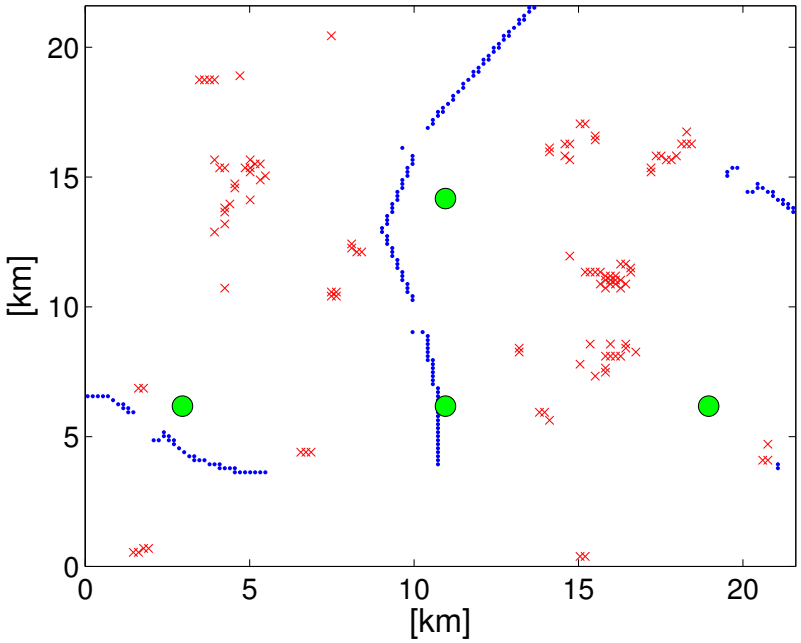


Figure 5.7: Storage site to be monitored, part of the site in Fig. 2.3. The green circles are the leakage locations of the predicted footprints used. Red crosses are wells and blue lines faults.

We want to use the same test case as in the previous chapters. However, the detectable regions (Fig. 5.5) of the footprints in Fig. 5.2 are much smaller than the detectable region of the previous used footprint of Fig. 2.2. For this reason, we choose to use a smaller part of the test site in Fig. 2.3. The part to be used is shown in Fig. 5.7 together with the leakage location of the four predicted footprints of Fig. 5.2.

The detection probability for the optimal sensor layout as function of the number of sensors is shown in Fig. 5.8. The optimal layout for four sensors is shown in Fig. 5.9 together with the monitored regions of the sensors.

5.6 Discussion of the method

The purpose of this chapter has been to incorporate spatial dependence in the average concentration footprint instead of using the same footprint for all leakage locations as in Ch. 4. The method for spatial dependence can also be applied to other footprints than average concentration footprints, which will be done in Ch. 7. We will now do a short discussion on the assumptions of the method as we have done in the previous chapters.

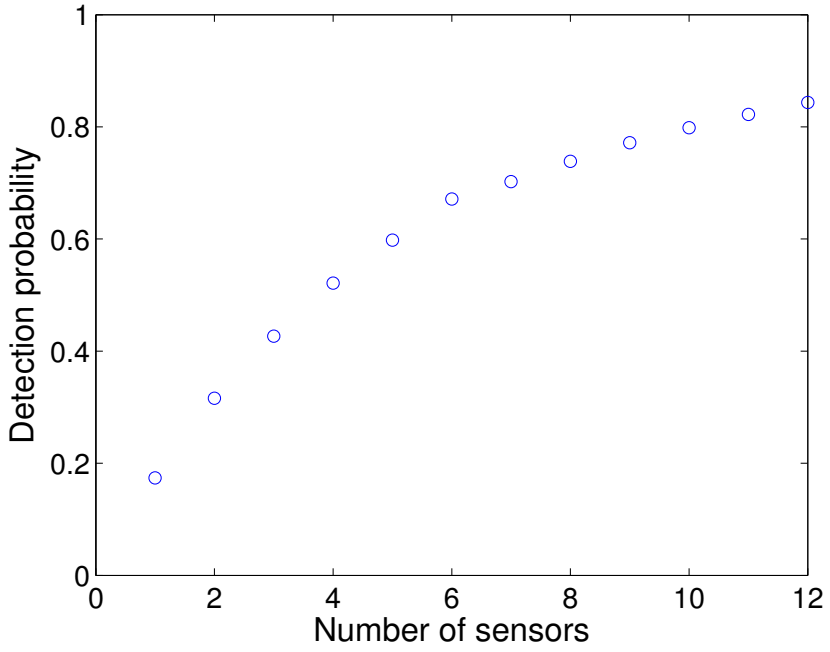


Figure 5.8: Detection probability for the optimal layout as function of the number of sensors. Detection threshold $c_t = 2.26 \cdot 10^{-4}$ kg/m^3.

Leakage location

As in the previous chapters, we have only used the leakage location probability density function as test case. The discussion of Sec. 3.7 on the leakage location then still applies.

Spatial dependence of footprint

The previous chapters have used the same footprint for all leakage locations. As discussed in Sec. 3.7, the average concentration footprint will in the real world not be the same for all leakage locations. The method of this chapter solves this problem by using an unique footprint for each possible leakage location. Ideally, we should have simulations for all possible locations. However, as mentioned before this requires too much computation and memory to be a reasonable option. The interpolation approach of this chapter requires a relatively small number of simulations, but will still give a good estimation for most leakage locations if the predicted footprints are wisely placed.

For the test case, only four predicted footprints are used, but the ideal number is probably a bit larger than that. This discussion on spatial variation applies also to other footprints when applying Sec. 5.4 to them.

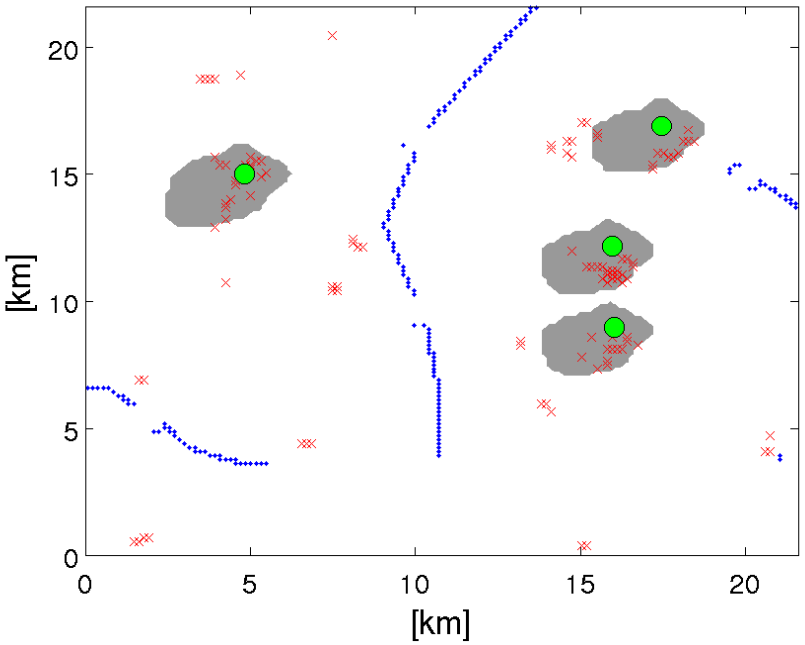


Figure 5.9: Sensor placements (green circles) and corresponding monitored regions (grey) for four sensors. Red crosses are wells and blue lines faults. Detection threshold $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}*. Detection probability* $p = 0.28$*.*

Approximation of footprint

Ch. 4 presented a method to invert a footprint exactly given that all leakage locations have the same footprint. The techniques of that chapter has been applied to the predicted footprints that are used for interpolation to predict the leakage footprint. This implies that even though we have introduced spatial dependence in the footprint, the inversion of the footprint will be exact once the footprint is estimated.

Average concentration

The method of this chapter is based on the average concentration from a leakage the same way as the methods of the previous chapters. The discussion on the use of average concentration in Sec. 3.7 then still applies. The next chapter will introduce a method that is not using the average concentration.

Chapter 6

Monitoring design: Translation of a probability footprint

This far we have only used the average concentration footprint from a leakage. As discussed in Sec. 3.7, the use of average concentration may be too conservative. Even though the average concentration over some time always is below the concentration threshold value c_t, the concentration may be above c_t a significant amount of the time, making the leakage possible to detect. A method based on the fraction of time the concentration is above c_t is suggested by Hvidevold et al. [7]. We will present and further develop this method. In this work we will use theory of Sec. 4.5 giving the red method in Fig. 6.1.

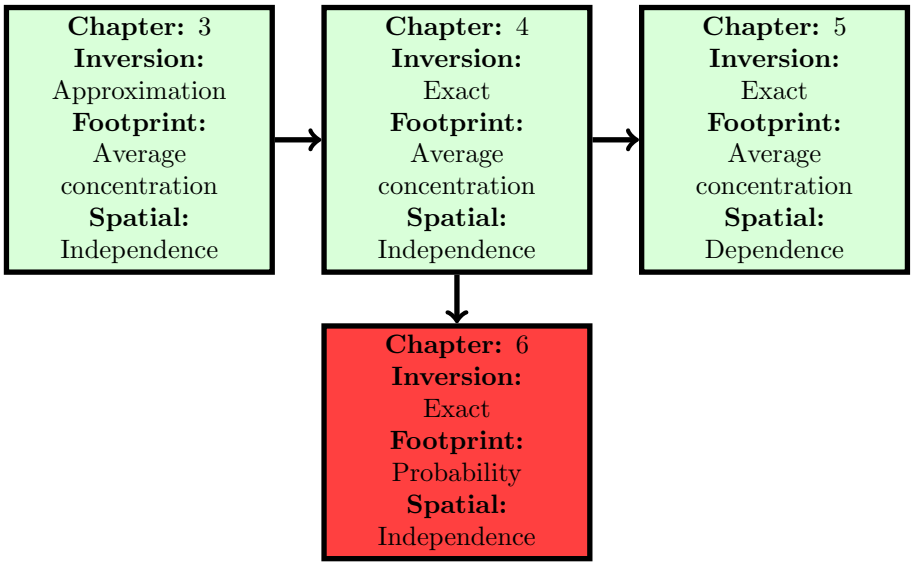


Figure 6.1: Summary of the different methods for monitoring design, the red is the one to be developed in this chapter.

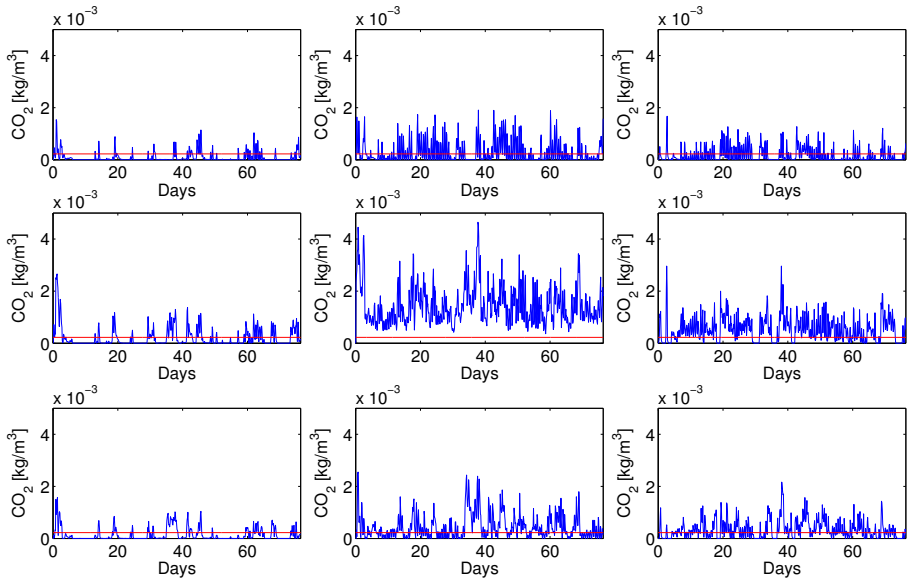


Figure 6.2: Time series (blue) for seafloor CO_2 concentration in the leakage point (center) and the eight surrounding grid points. The red lines show the threshold concentration $c_t = 2.26 \cdot 10^{-4}$ kg/m^3.

6.1 Probability footprint

Instead of using the average concentration footprint of a leakage, [7] suggests to use a probability footprint. The value of this footprint is the probability that the concentration at a single randomly chosen time is above the threshold value c_t. E.g. if the concentration is above c_t 20% of the time at a given location, the probability footprint will have a value of 0.2 there.

We can also think of the probability footprint as a detection function. If we perform a single measurement at a location, the value of the probability footprint is the probability that the concentration is above c_t. We assume that a leakage always is detected if we measure a concentration above c_t and never detected if we measure a concentration below c_t. The probability footprint will then be the detection function for a single measurement. For this reason, we will denote the value of the probability footprint at $\mathbf{x}$ when having a leakage at $\mathbf{x}_0$ by the detection function notation $D_S(\mathbf{x}; \mathbf{x}_0; c_t)$. The single measurement method will however only be used in the development of the event method. This is because we are interested in fixed sensors, not surveys.

To calculate the probability footprint, we need the concentration time series of the grid points surrounding the leakage. In Hvidevold et al. [7], the time series used are from a leakage scenario at Sleipner similar to the test cases used in the previous chapters. Fig. 6.2 shows the concentration time series for the leakage location and its surrounding grid points. The time series are plotted together with the standard threshold concentration $c_t = 2.26 \cdot 10^{-4}$ kg/m^3, showing when the concentration is above the threshold value.

Let now $C_l(\mathbf{x}; \mathbf{x}_0)$ be the value at time step l of the concentration time series at $\mathbf{x}$ when having a leakage at $\mathbf{x}_0$. Further, let L be the number of time steps

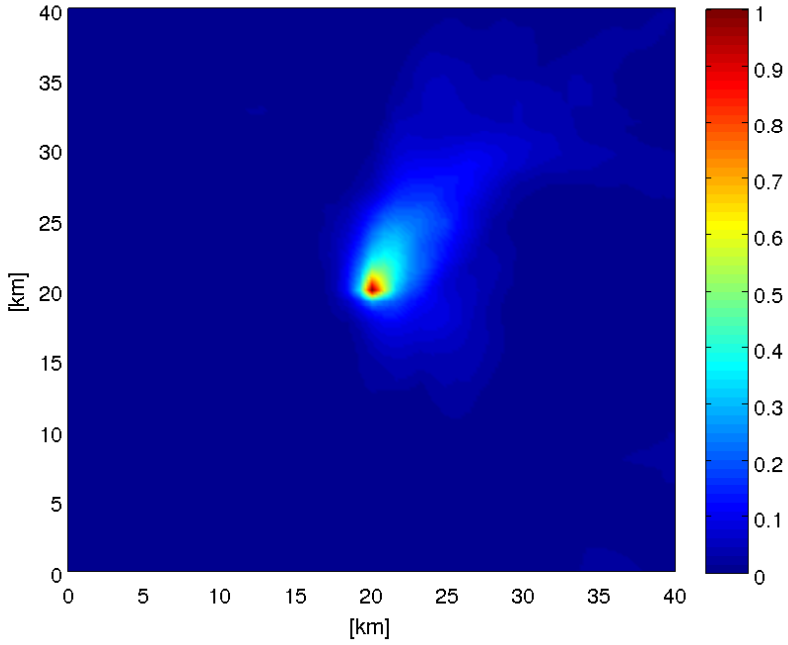


Figure 6.3: Probability footprint $D_S(\mathbf{x};\mathbf{x}^*;c_t)$ for $c_t = 2.26 \cdot 10^{-4}$ kg/m^3 used in [7] and as a test case in this chapter.

such that $l \in \{1, 2, \ldots, L\}$. The probability footprint, or the single measurement detection function, is then given by

$$D_S(\mathbf{x};\mathbf{x}_0;c_t) = \frac{\sum_{l=1}^{L} I_D(C_l(\mathbf{x};\mathbf{x}_0);c_t)}{L} \tag{6.1}$$

where $I_D(C_l(\mathbf{x};\mathbf{x}_0;c_t)$ is the detection indicator function

$$I_D(C_l(\mathbf{x};\mathbf{x}_0);c_t) = \begin{cases} 1 & \text{if } C_l(\mathbf{x};\mathbf{x}_0) \geq c_t \\ 0 & \text{if } C_l(\mathbf{x};\mathbf{x}_0) < c_t \end{cases}. \tag{6.2}$$

We assume that the probability footprint is independent of the leakage location. This makes it possible to apply the theory of Sec. 4.5. Let as before $\mathbf{x}^*$ be the leakage location of the footprint $D_S(\mathbf{x};\mathbf{x}^*;c_t)$ predicted from a simulation. We then have a translation property equivalent to Eq. 4.2 for the footprint of an arbitrary leakage location $\mathbf{x}_0$:

$$D_S^T(\mathbf{x};\mathbf{x}_0;c_t) = D_S(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*;c_t) \tag{6.3}$$

$$D_S^T(\mathbf{x};\mathbf{x}_0;c_t) = \frac{\sum_{l=1}^{L} I_D(C_l(\mathbf{x}^* + (\mathbf{x} - \mathbf{x}_0);\mathbf{x}^*);c_t)}{L}. \tag{6.4}$$

The footprint for $c_t = 2.26 \cdot 10^{-4}$ kg/m^3 used in Hvidevold et al. [7] is shown in Fig. 6.3. We will adopt this footprint as a test case in this chapter.

6.2 Detection function for event

We are now ready to introduce the event method of [7]. Instead of performing a survey with single measurements, we now assume that we have fixed sensors as in the previous chapters. The relative period of time a sensor will experience a concentration above c_t from the leakage is the value of the probability footprint. If this value is above some probability threshold p_t, we assume that the sensor always detects the leakage. If it is below p_t, we assume that the sensor never detects the leakage. This gives the detection function for the event method:

$$D_E(\mathbf{x};\mathbf{x}_0;c_t;p_t) = \begin{cases} 1 & \text{if } D_S(\mathbf{x};\mathbf{x}_0;c_t) \geq p_t \\ 0 & \text{if } D_S(\mathbf{x};\mathbf{x}_0;c_t) < p_t \end{cases}. \tag{6.5}$$

A way to explain this method is that if the concentration is above the threshold concentration c_t sufficiently often, there will within reasonable time be an event where the concentration is large enough that the leakage is detected. Of course, we have no guarantee that these events are spread and not clustered in time. Time series analysis could be used to quantify the risk of this happening, but will not be performed here. Since we use the same predicted probability footprint for all leakage locations, the event detection function will be the same for all leakage locations. This implies that we have the same translation property for an arbitrary leakage location as Eq. 6.3 for the probability footprint:

$$D_E^T(\mathbf{x};\mathbf{x}_0;c_t;p_t) = D_E(\mathbf{x}^* + (\mathbf{x}-\mathbf{x}_0);\mathbf{x}^*;c_t;p_t) \tag{6.6}$$

$$D_E^T(\mathbf{x};\mathbf{x}_0;c_t;p_t) = \begin{cases} 1 & \text{if } D_S(\mathbf{x}^* + (\mathbf{x}-\mathbf{x}_0);\mathbf{x}^*;c_t) \geq p_t \\ 0 & \text{if } D_S(\mathbf{x}^* + (\mathbf{x}-\mathbf{x}_0);\mathbf{x}^*;c_t) < p_t \end{cases}. \tag{6.7}$$

As mentioned before, the event method is similar to the average method of Ch. 4. The probability footprint takes the role of the average concentration footprint and the application of a probability threshold p_t is equivalent to the use of the concentration threshold c_t in Ch. 4. Note that c_t is used also in the event method of this chapter, but in a different way than in Ch. 4. Due to the equivalence of the event and average method, we obtain a detectable region for the event method the same way as Eq. 4.4 for the average method:

$$\boxed{\begin{array}{c}\text{Detectable region} \\ \text{translated event}\end{array}} = \{\mathbf{x}\,|D_S(\mathbf{x}^* + (\mathbf{x}-\mathbf{x}_0);\mathbf{x}^*;c_t) \geq p_t\} \tag{6.8}$$

All sensors within this region is assumed to always detect the leakage, while all sensors outside of this region is assumed to never will detect the leakage. Fig. 6.4, shows some level contours $D_S(\mathbf{x};\mathbf{x}_0;c_t) = p_t$ of the probability footprint in Fig. 6.3 for $c_t = 2.26 \cdot 10^{-4}$ kg/m^3. These level contours give the boundaries of the detectable regions for the different values of p_t. The typical value of p_t used in Hvidevold et al. [7] is $p_t = 0.10$, implying that the area within the blue line in Fig. 6.4 will be the detectable region.

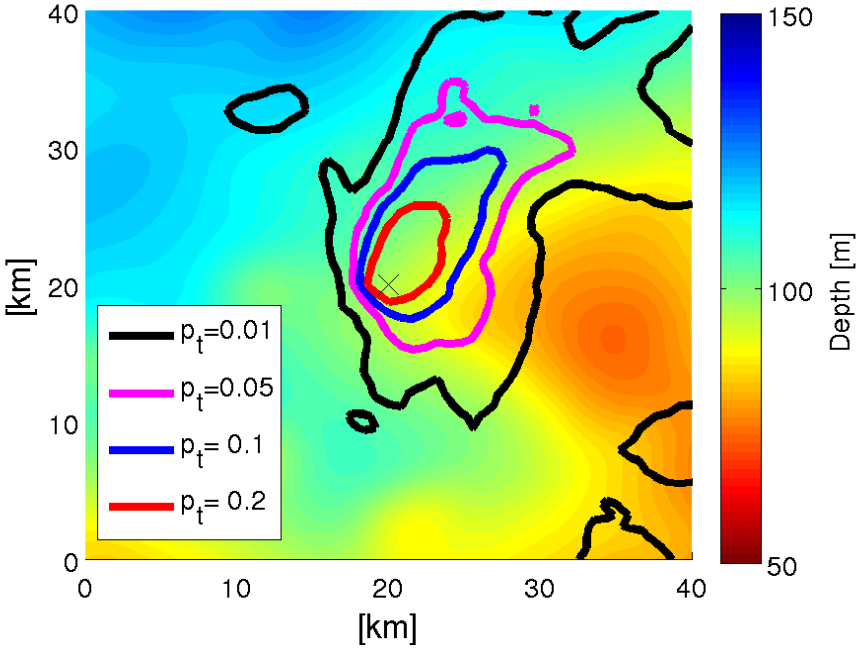


Figure 6.4: Level contours of the probability footprint in Fig. 6.3. The level contours define the boundaries of the detectable regions for the different values of p_t. The contours are plotted together with the depth. The leakage location is marked with a cross. Threshold concentration $c_t = 2.26 \cdot 10^{-4}$ kg/m^3.

6.3 Monitoring function for event

The monitoring function for a sensor at $\mathbf{x}_s$ is as before given implicitly by the detection function and Eq. 2.4 with the extra parameter p_t:

$$M_E(\mathbf{x}; \mathbf{x}_s; c_t; p_t) = D_E(\mathbf{x}_s; \mathbf{x}; c_t; p_t). \tag{6.9}$$

Since we have assumed the probability footprint to be equal everywhere, we have seen that we have the same translation properties for the footprint and the detection function as in Ch. 4. This implies that the inversion to obtain the monitoring function is the same as in Ch. 4. We first express the monitoring function in terms of the detection function for the predicted leakage:

$$M_E^T(\mathbf{x}; \mathbf{x}_s; c_t; p_t) = D_E(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*; c_t; p_t). \tag{6.10}$$

In terms of the probability footprint, the monitoring function becomes

$$M_E^T(\mathbf{x}; \mathbf{x}_s; c_t; p_t) = \begin{cases} 1 & \text{if } D_S(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*; c_t) \geq p_t \\ 0 & \text{if } D_S(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*; c_t) < p_t \end{cases}. \tag{6.11}$$

We see that these equations are equivalent to the ones obtained for the monitoring function of Ch. 4. This enables us to apply Sec. 4.5 to invert the

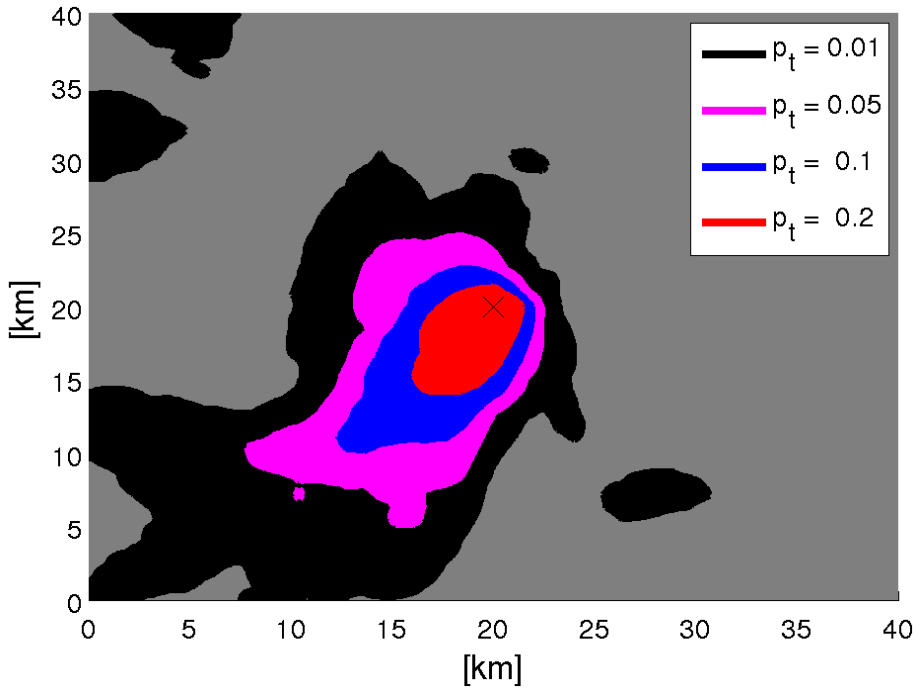


Figure 6.5: Monitored regions for different values of p_t. The sensor location is marked with a cross. Note that the monitored regions are obtained by a 180° rotation of the detectable regions in Fig. 6.4. $c_t = 2.26 \cdot 10^{-4}$ kg/m^3

footprint and the detection function by the 180° rotation described in Sec. 4.4. The monitored region for the event method will be

$$\boxed{\begin{array}{c}\text{Monitored region}\\ \text{translated event}\end{array}} = \{\mathbf{x}\,|D_S(\mathbf{x}^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}^*; c_t) \geq p_t\}\,. \tag{6.12}$$

6.4 Optimal design

We will now apply the method to the same test case (Sec. 2.1.3) as previously. This implies to find the optimal sensor layout for different numbers of sensors for the Sleipner leakage location probability density function of Fig. 2.3. We will use the probability footprint of Fig. 6.3, which is from a simulation at Sleipner. However, this is a different simulation (Tab. C.1) than the one giving the average concentration footprint of Fig. 2.2 used in Ch. 3 and 4. This implies that we can't compare the detection probabilities and sensor locations of Ch. 4 with what we obtain here.

Fig. 6.6 shows the detection probability for the optimal sensor layout as a function of the number of sensors for different probability thresholds p_t. The great difference in the detection probabilities shows that the choice of p_t is of great importance. The optimal placements of four sensors for the different values of p_t are plotted in Fig. 6.7 together with the monitored regions of the

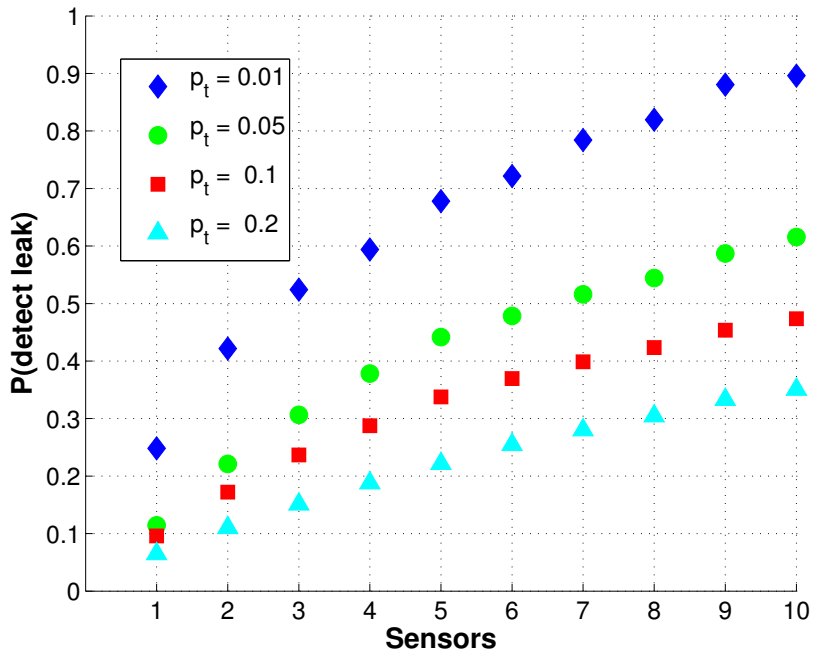


Figure 6.6: Detection probability for the optimal layout as function of the number of sensors for different values of the threshold probability p_t. Threshold concentration $c_t = 2.26 \cdot 10^{-4}$ kg/m^3.

sensors. The placement of the sensors for the different values of p_t are plotted together in Fig. 6.9 for comparison of the layouts. Fig. 6.8 shows the optimal sensor layout for eight sensors having $p_t = 0.1$.

6.5 Discussion of the method

In this chapter we have presented the event method from [7] and applied the inversion of Ch. 4 to it. We will now do a short discussion on some of the aspects of the event method.

Leakage location

As before, we think of the leakage location probability density function as given while developing the method. This implies that the previous discussion on the leakage location in Sec. 3.7 still applies.

Probability footprint and threshold

The first step in the event method is the calculation of the probability footprint. We originally have full time series for the concentration in all grid points surrounding the leakage. These time series are then reduced to a single probability value in each grid point, just as the time series are reduced to

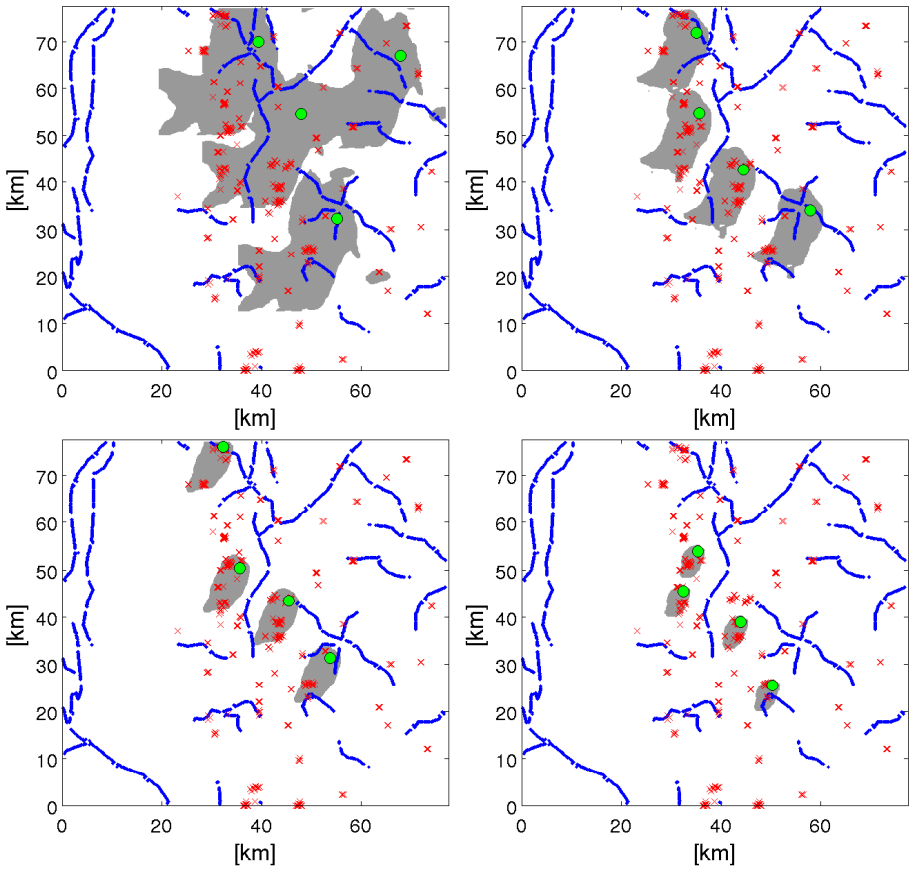


Figure 6.7: Optimal sensor locations (green) and monitored area (gray) using four sensors for different values of the threshold probability p_t: 0.01 (top left), 0.05 (top right), 0.1 (bottom left) and 0.2 (bottom right). Corresponding detection probabilities are $p = 0.59$, $p = 0.38$, $p = 0.29$ and $p = 0.19$. Red crosses are wells and blue lines are faults. $c_t = 2.26 \cdot 10^{-4}$ kg/m^3

the average concentration value in the previous chapters. Since we want to estimate the ability of the monitoring program to detect a leakage at an arbitrary time, it is reasonable to reduce the time series to a typical value. It is also computationally much easier to work with a single value in each grid point than with the large data amount a complete time series in each grid point is.

The probability footprint is a more realistic approach than the average concentration footprint. As discussed in Sec. 3.7, a leakage may be detectable even though the average concentration is below the concentration threshold c_t. However, locations having a concentration above c_t much of the time, and thus a high average concentration, will be a more likely place to detect the leakage. The probability footprint takes this into account by using not only the values 0 and 1, but the full range between 0 and 1.

It can be useful to use the probability footprint as detection function directly, as is done in [7] as the single measurement method. The smooth transition in the detection function between 0 (never detect) and 1 (always detect) is realistic.

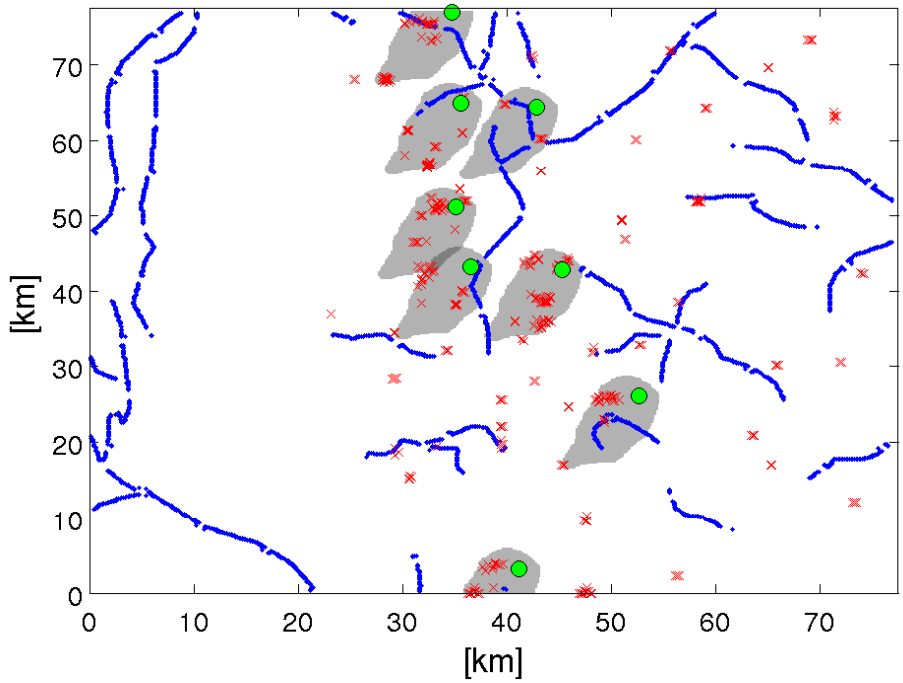


Figure 6.8: Optimal sensor locations (green) and monitored area (gray) using eight sensors for $p_t = 0.1$. The detection probability is $p = 0.47$. Red crosses are wells and blue lines are faults. $c_t = 2.26 \cdot 10^{-4}$ kg/m^3

However, the event detection function of this chapter is also reasonable. A grid point having a certain probability of the concentration exceeding the threshold concentration c_t, will experience a concentration above c_t at a regular basis.

The biggest challenge of the event method is to set the threshold probability p_t correctly. Different values of p_t give completely different detectable regions and detection probabilities as is easily seen by Fig. 6.4 and Fig. 6.6. These different values of p_t then give quite different sensor layouts as seen in Fig. 6.9.

Two of the questions to consider when determining p_t is for how long time it is acceptable that the concentration stays below c_t, and for how long time the concentration has to stay above c_t for it to be significant.

The answer to the first question depends on what is an acceptable detection time. This may depend on several factors determining the severity of a leakage. The answer to the second question depends on how sure we want to be that an alarm really is a leakage. This may also depend on the severity of a potential leakage, together with the cost of investigating an alarm.

After answering the two questions, statistical time series analysis should be applied to find a proper value for p_t. This is not done here, and the values of p_t used are to be considered test values. To do a complete time series analysis would be a thoroughly process, and may also give different values of p_t for different locations. Having a larger p_t gives a smaller detectable region and requires the concentration to be above c_t a larger amount of the time for a leakage to be considered detectable. In the absence of a complete time series

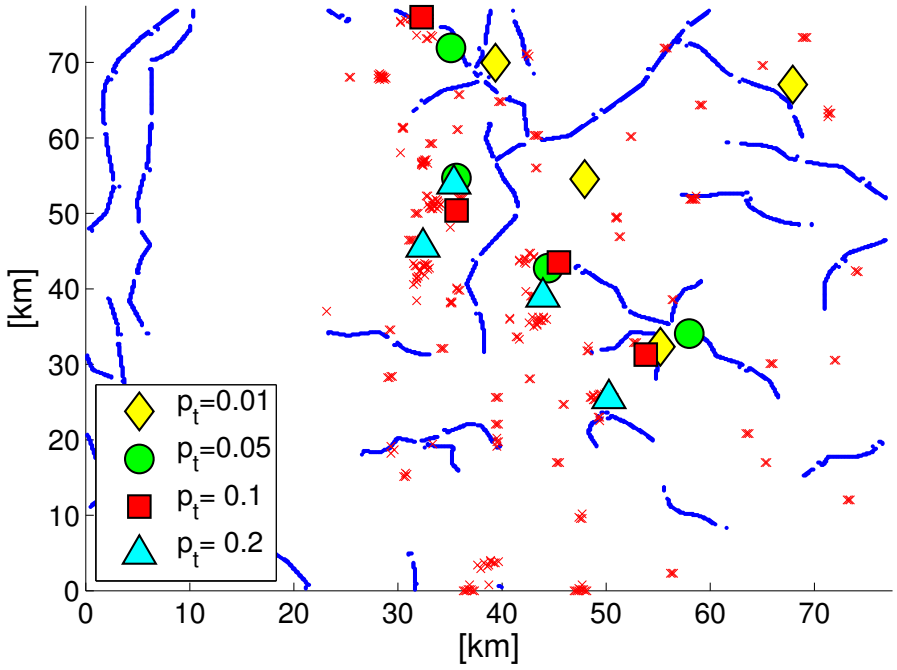


Figure 6.9: Optimal sensor locations using four sensors for different threshold probabilities p_t. Red crosses are wells and blue lines are faults. Threshold concentration $c_t = 2.26 \cdot 10^{-4}$ kg/m^3.

analysis, it is possible to do a conservative approach and choose a larger p_t to be on the safe side.

Spatial independence of footprint

Even though we use a probability footprint instead of an average concentration footprint, the discussion of Sec. 3.7 still applies. The footprint should vary between the different leakage locations due to spatial variations in topography and currents. This calls for a method taking these variations into account. Such a method was developed in Ch. 5 for the average concentration footprint. We will in the next chapter apply this method to the probability footprint.

Chapter 7

Monitoring design: Interpolation of probability footprints

Ch. 6 presented the event based method from [7]. This method uses the probability footprint presented in Sec. 6.1. In Ch. 6, this footprint was assumed to be independent of the leakage location. The further calculations of the event method in the chapter were then similar to the calculations in Ch. 4 for the average concentration footprint with spatial independence.

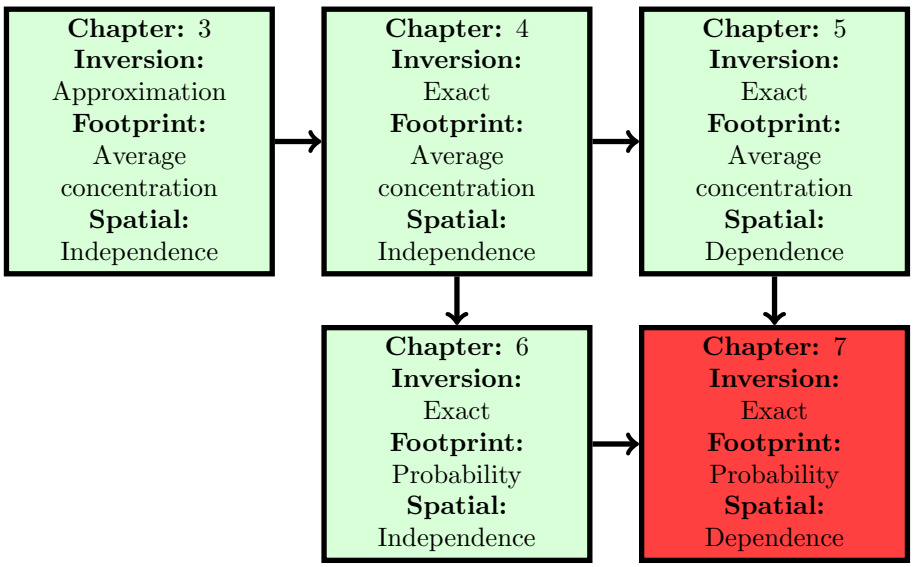


Figure 7.1: Summary of the different methods for monitoring design, the red is the one to be developed in this chapter.

After discussing the need of incorporating spatial dependence in the average concentration footprint, a method including spatial dependence was developed in Ch. 5. The necessity of developing a similar method with spatial dependence

in the probability footprint for the event method is addressed in Sec. 6.5. We will develop such a method in this chapter giving the red method of Fig. 7.1.

This chapter will be an extension of Ch. 6 similar to how Ch. 5 is an extension of Ch. 4. Since much of the calculations of Ch. 4 and Ch. 6 are the same, the method of this chapter is obtained by applying the theory of Ch. 5 to Ch. 6 as is explained in Sec. 5.4 for an arbitrary footprint.

7.1 Estimation of the probability footprint

The first thing we need to do is to estimate the probability footprint for an arbitrary leakage location including spatial dependence. We assume as in Ch. 5 that we have simulations of leakages from K different leakage locations $\mathbf{x}_1^*, \ldots, \mathbf{x}_K^*$. To calculate the predicted probability footprints $D_S(\mathbf{x}; \mathbf{x}_k^*; c_t)$, we need the concentration time series for leakages at each of the K leakage locations. As in Ch. 6, we let $C_l(\mathbf{x}; \mathbf{x}_0)$ be the concentration value at location $\mathbf{x}$ at time step l when having a leakage at $\mathbf{x}_0$. The predicted footprints $D_S(\mathbf{x}; \mathbf{x}_k^*; c_t)$ are then calculated using Eq. 6.1 for each predicted leakage location $\mathbf{x}_0 = \mathbf{x}_k^*$:

$$D_S(\mathbf{x}; \mathbf{x}_0; c_t) = \frac{\sum_{l=1}^{L} I_D(C_l(\mathbf{x}; \mathbf{x}_0); c_t)}{L}. \tag{6.1}$$

As in Ch. 6, L is the total number of time steps in the time series and $I_D(C_l(\mathbf{x}; \mathbf{x}_0); c_t)$ is the detection indicator function

$$I_D(C_l(\mathbf{x}; \mathbf{x}_0); c_t) = \begin{cases} 1 & \text{if } C_l(\mathbf{x}; \mathbf{x}_0) \geq c_t \\ 0 & \text{if } C_l(\mathbf{x}; \mathbf{x}_0) < c_t \end{cases}. \tag{6.2}$$

We now have K predicted probability footprints $D_S(\mathbf{x}; \mathbf{x}_k^*; c_t)$ and want to use them to estimate the footprint of an arbitrary leakage location including spatial dependence. We can apply the theory of Ch. 5 as described in Sec. 5.4 for an arbitrary footprint. The first we do is to translate the footprints to $\mathbf{x}_0$. A leakage located at $\mathbf{x}_0$ having probability footprint k will be given by

$$D_S^T(\mathbf{x}; \mathbf{x}_0; c_t) = D_S(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*; c_t). \tag{7.1}$$

We will interpolate between these footprints, giving the following expression for the footprint of an arbitrary leakage location:

$$D_S^I(\mathbf{x}; \mathbf{x}_0; c_t) = \sum_{k=1}^{K} w_k(\mathbf{x}_0) D_S(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*; c_t). \tag{7.2}$$

The interpolation weights $w_k(\mathbf{x})$ are the same as in Ch. 5, and given by Eq. 5.7 using the leakage locations $\mathbf{x}_1^*, \ldots, \mathbf{x}_K^*$ of the predicted footprints. Fig. 7.2 shows four of the nine probability footprints used in Sec. 7.4. The resulting footprint given by Eq. 7.2 for an arbitrary leakage location is shown in Fig. 7.3.

7.2 Detection function for interpolated event

The only difference between Ch. 6 and this chapter is the estimation of the probability footprint. Once the footprint is estimated, we have the same

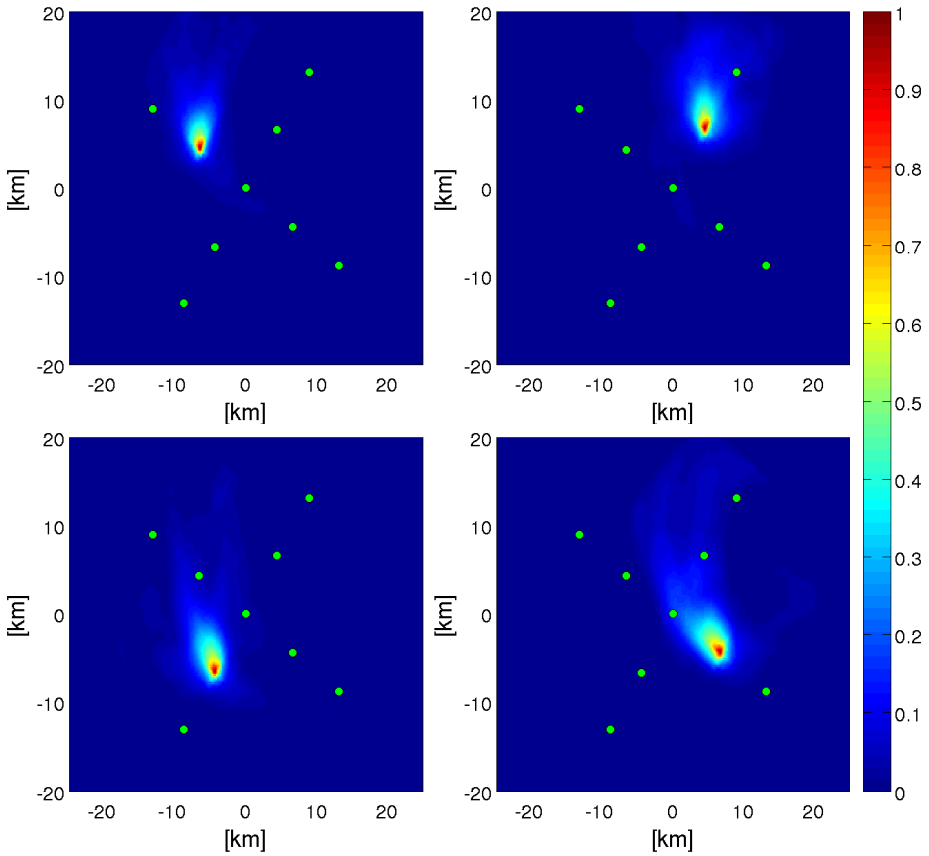


Figure 7.2: Four of the nine probability footprints used as test case in Sec. 7.4. The leakage locations of the other footprints are marked with green circles. We can clearly see the spatial dependence in the footprint. $c_t = 2.26 \cdot 10^{-4}$ *kg/m*^{3}

detection function as in Ch. 6:

$$D_E(\mathbf{x}; \mathbf{x}_0; c_t; p_t) = \begin{cases} 1 & \text{if } D_S(\mathbf{x}; \mathbf{x}_0; c_t) \geq p_t \\ 0 & \text{if } D_S(\mathbf{x}; \mathbf{x}_0; c_t) < p_t \end{cases}. \tag{6.5}$$

Using Eq. 7.2 for the probability footprint, we get the detection function for the event method with spatial dependence:

$$D_E^I(\mathbf{x}; \mathbf{x}_0; c_t; p_t) = \begin{cases} 1 & \text{if } \sum_{k=1}^K w_k(\mathbf{x}_0) D_S(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*; c_t) \geq p_t \\ 0 & \text{if } \sum_{k=1}^K w_k(\mathbf{x}_0) D_S(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*; c_t) < p_t \end{cases}. \tag{7.3}$$

We see that the detection function obtained is equivalent to Eq. 5.8 for the average method with spatial dependence. As before, the detection function gives rise to a detectable region where a sensor inside this region is assumed to detect a potential leakage at $\mathbf{x}_0$:

$$\boxed{\begin{array}{c}\text{Detectable region} \\ \text{interpolated event}\end{array}} = \left\{ \mathbf{x} \,\middle|\, \sum_{k=1}^K w_k(\mathbf{x}_0) D_S(\mathbf{x}_k^* + (\mathbf{x} - \mathbf{x}_0); \mathbf{x}_k^*; c_t) \geq p_t \right\}. \tag{7.4}$$

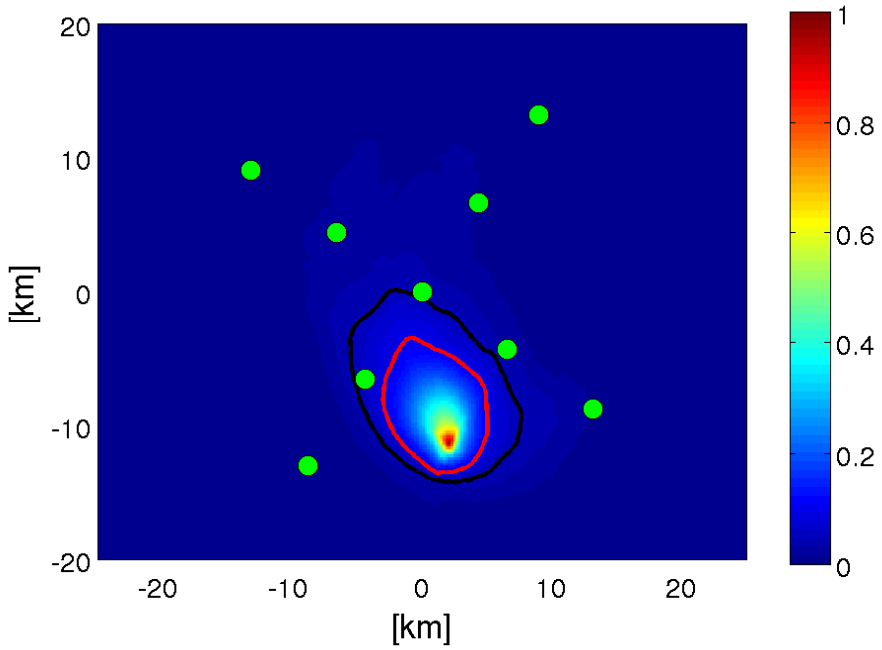


Figure 7.3: Probability footprint for an arbitrary leakage location using nine predicted footprints, whose locations are marked with a green circle. Comparing with four of the predicted footprints shown in Fig. 7.2, we can see the spatial dependence. The black contour line is the level curve for $p_t = 0.05$ and the red contour for $p_t = 0.1$, giving the boundaries of the detectable regions.

The detectable regions for the footprints in Fig. 7.2 are shown in Fig. 7.4.

7.3 Monitoring function for interpolated event

The monitoring function will as before be given implicitly by the detection function by the relation

$$M_E(\mathbf{x}; \mathbf{x}_s; c_t; p_t) = D_E(\mathbf{x}_s; \mathbf{x}; c_t; p_t). \tag{6.9}$$

By using Eq. 7.3 for the detection function, we obtain the monitoring function for the event method with spatial dependence:

$$M_E^I(\mathbf{x}; \mathbf{x}_s; c_t; p_t) = \begin{cases} 1 & \text{if } \sum_{k=1}^K w_k(\mathbf{x}) D_S(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*; c_t) \geq p_t \\ 0 & \text{if } \sum_{k=1}^K w_k(\mathbf{x}) D_S(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*; c_t) < p_t \end{cases}. \tag{7.5}$$

Similarly as for the detection function, the monitoring function is equivalent to Eq. 5.11 for the average method with spatial dependence. We see that the inverted probability footprint is given by

$$D_S^I(\mathbf{x}_s; \mathbf{x}; c_t) = \sum_{k=1}^K w_k(\mathbf{x}) D_S(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*; c_t). \tag{7.6}$$

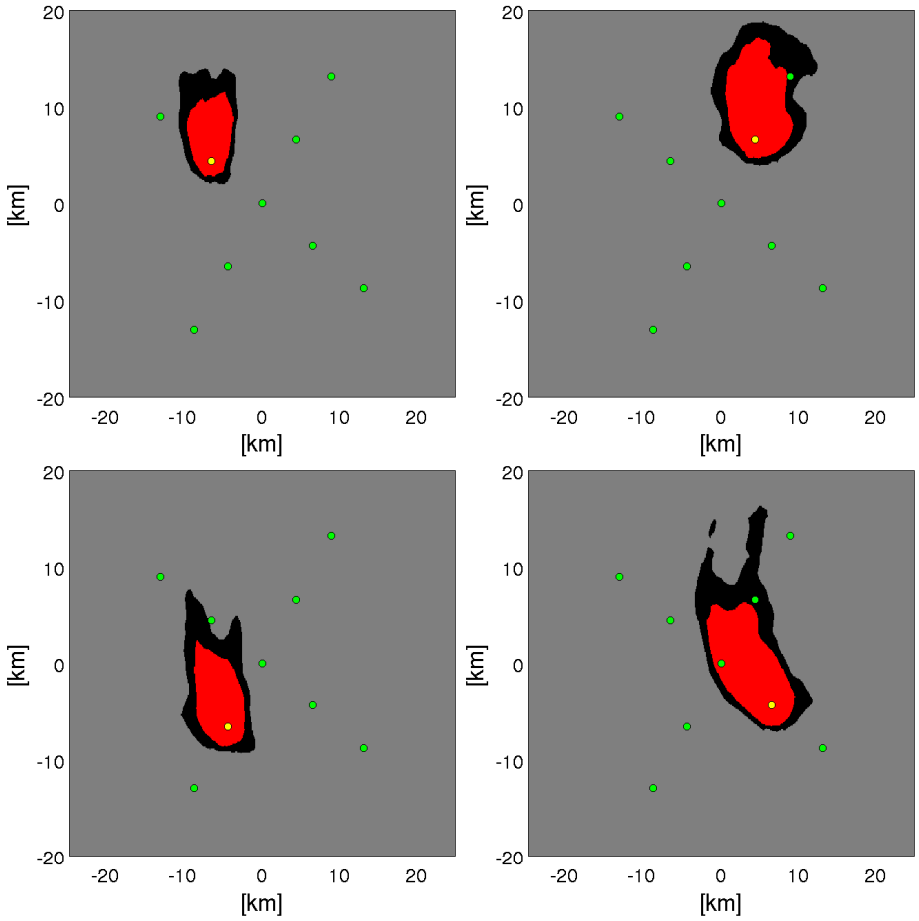


Figure 7.4: The detectable regions for the four footprints in Fig. 7.2. The leakage location of the leakage is marked with a yellow circle, and the leakage locations of the other footprints with green circles. The black region (including the red) is for $p_t = 0.05$ and the red for $p_t = 0.1$.

This inversion is the equivalent of Eq. 5.10 for the average method with spatial dependence. Note again that the argument of the weight functions w_k is the independent variable $\mathbf{x}$, while for the footprint Eq. 7.2 the argument $\mathbf{x}_0$ of the weight functions is fixed. The monitored region will be

$$\boxed{\begin{array}{c}\text{Monitored region}\\ \text{interpolated event}\end{array}} = \left\{ \mathbf{x} \left| \sum_{k=1}^{K} w_k(\mathbf{x}) D_S(\mathbf{x}_k^* - (\mathbf{x} - \mathbf{x}_s); \mathbf{x}_k^*; c_t) \geq p_t \right. \right\}. \quad (7.7)$$

Fig. 7.5 shows the monitored regions for an arbitrary sensor location using the test case footprints of Sec. 7.4.

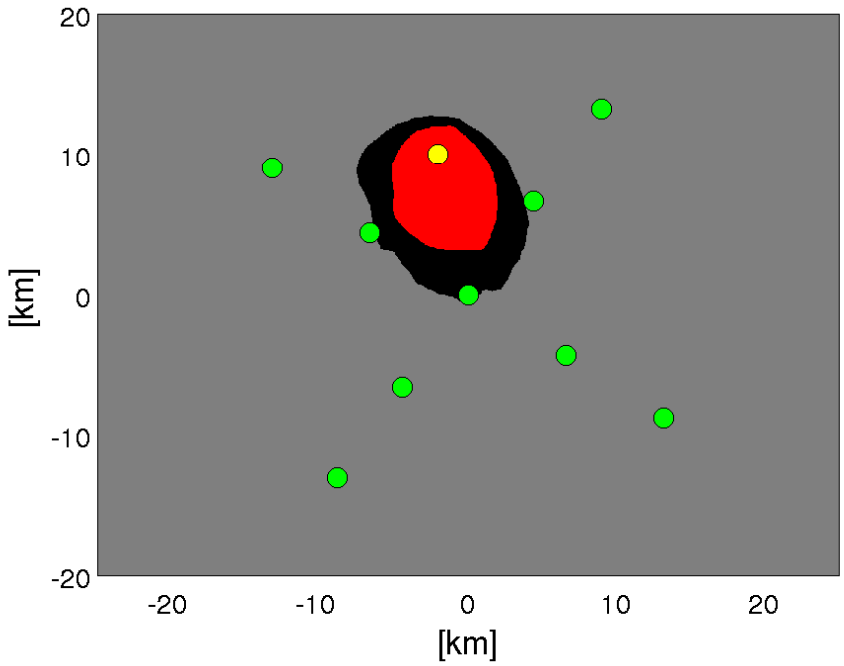


Figure 7.5: Monitored regions for an arbitrary sensor location (yellow circle) using the predicted footprints of Sec. 7.4 The green circles show the location of the nine predicted footprints, where four of them are plotted in Fig. 7.2. The black region (including the red) is for $p_t = 0.05$, and red is for $p_t = 0.1$.

7.4 Optimal design

We now want to apply the event method with spatial dependence to a test case. Simulations for nine different leakage locations at the Sleipner storage site are preformed. The relative location of the leakage locations are shown in Fig. 7.2 together with the probability footprint for four of the locations.

It is possible to use the leakage location probability density function of Sec. 2.1.3 as in previous chapters. However, we will instead use a completely synthetic leakage location probability density function shown in Fig. 7.6.

The main reason for this is that parts of the work in this chapter was used in Ali et al. [4], where the synthetic leakage location probability density function was desirable. As mentioned before we are not interested in site studies, implying that the test case of Sec. 2.1.3 is not of specific interest. Since the footprints used to test the different methods are from different simulations (Tab. C.1), we can not compare the optimal layouts and detection probabilities obtained for the different methods independently of the leakage location probability density function used. For this reason, changing test case for is no problem.

The leakage location probability density function $f(\mathbf{x})$ is created similarly to in Hvidevold et al. [7]. Every location $\mathbf{x}_0$ that is assumed to be a potential

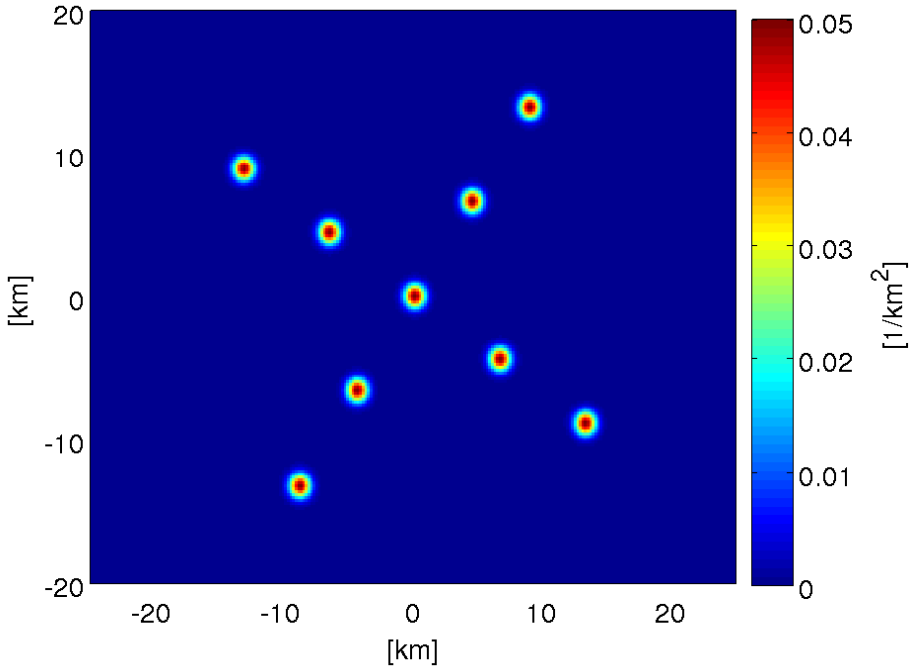


Figure 7.6: Leakage location probability density function used as test case in Sec. 7.4. The leakage locations of the predicted footprints are the mode of a Gaussian distribution. In addition, a small background probability is added everywhere to account for unknown leakage pathways.

leakage location is the mode of an Gaussian distribution on the form

$$\exp\left(-\frac{(\mathbf{x}-\mathbf{x}_0)^T(\mathbf{x}-\mathbf{x}_0)}{\sigma^2}\right) \qquad \mathbf{x}\in A. \tag{7.8}$$

To create the function in Fig. 7.6, a shape parameter of $\sigma = 0.8$ km was used. We assume that the locations $\mathbf{x}_1^*,\ldots,\mathbf{x}_K^*$ of the predicted footprints are the only possible leakage locations that are given a Gaussian distribution. All of these locations are assumed to have the same probability of being the leakage location. In addition, a constant background probability is added everywhere to account for potential unknown leakage pathways. The leakage location probability density function $f(\mathbf{x})$ is then given by:

$$f(\mathbf{x}) = f_{Norm}\left(\sum_{k=1}^{K}\exp\left(-\frac{(\mathbf{x}-\mathbf{x}_k^*)^T(\mathbf{x}-\mathbf{x}_k^*)}{\sigma^2}\right)+f_{Back}\right) \tag{7.9}$$

The constant f_{Back} gives the ratio between the total background probability and the probability of the leakage locations $\mathbf{x}_1^*,\ldots,\mathbf{x}_K^*$. We have chosen f_{Back} such that the total background probability is approximately 1/10 of the total probability of the leakage locations, i.e.

$$\int_A\sum_{k=1}^{K}\exp\left(-\frac{(\mathbf{x}-\mathbf{x}_k^*)^T(\mathbf{x}-\mathbf{x}_k^*)}{\sigma^2}\right)dxdy \approx 10\int_A f_{Back}dxdy. \tag{7.10}$$

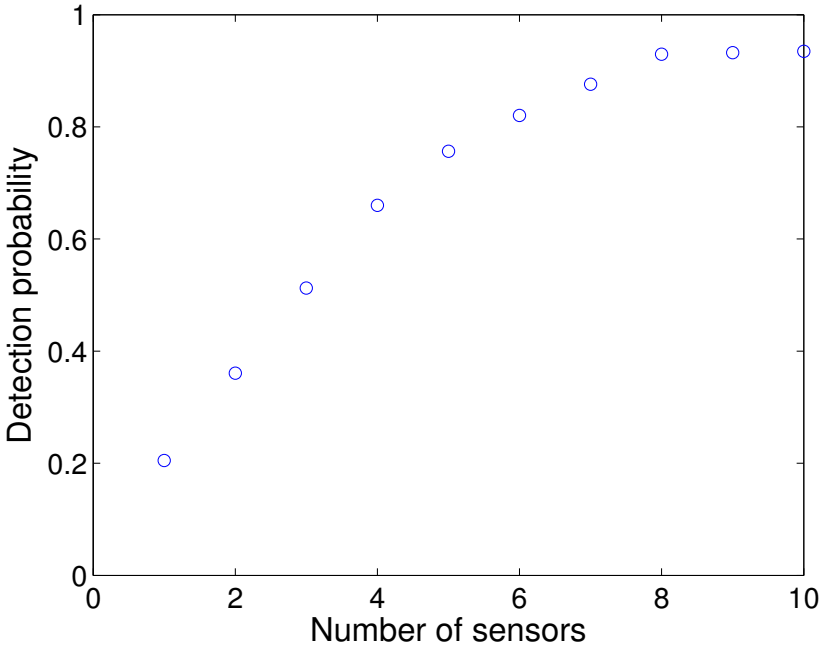


Figure 7.7: Detection probability for the optimal sensor layout as a function of the number of sensors. The leakage location probability density function is shown in Fig. 7.6. A probability threshold of $p_t = 0.1$ is used.

The constant f_{Norm} is used for normalization such that f integrates to 1 over the area A to be monitored, i.e. satisfying Eq. 2.2:

$$f_{Norm} = \left(\int_A \left(\sum_{k=1}^{K} \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_k^*)^T (\mathbf{x} - \mathbf{x}_k^*)}{\sigma^2} \right) + f_{Back} \right) dxdy \right)^{-1}. \quad (7.11)$$

Using the nine footprints and the leakage location probability density function in Fig. 7.6, optimal sensor layouts for different number of sensors are obtained. Fig. 7.7 shows the detection probability for the optimal sensor layouts as a function of the number of sensors. The optimal sensor layouts for two and four sensors are shown in Fig. 7.8 and Fig. 7.9 together with the monitored regions.

7.5 Discussion of the method

In this chapter we have introduced the event method with spatial dependence, combining Ch. 5 and 6. We will now do a short discussion on the method.

Leakage location

As in all of the previous chapters, we are not interested in specific site studies. We are interested in developing methods for finding the optimal sensor layout and estimating the detection probability. For this reason, we take the

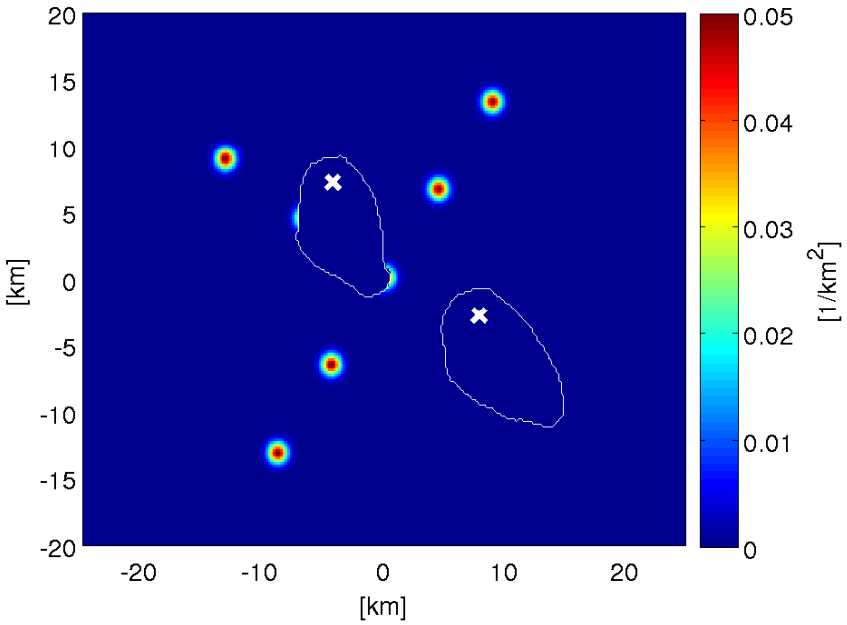


Figure 7.8: Residual leakage location probability density function for the optimal sensor layout of two sensors (white crosses). The monitored regions of the sensors are zeroed out and the boundaries of the monitored regions are marked with white lines. Compare with the original leakage location probability density function in Fig. 7.6 to see the monitored (zeroed) probability. A probability threshold of $p_t = 0.1$ is used. The detection probability is $p = 0.36$.

leakage location probability density function as given and use it as a test case. In this chapter, we have used the synthetic leakage location probability density function of Fig. 7.6 instead of the test case of Sec. 2.1.3 used in the previous chapters. A complete study necessary to create a realistic leakage location probability density function was not performed in Sec. 2.1.3. In addition, we can not compare the method output of the different chapters since the footprints used are from different simulations. For this reason, the synthetic function can equally well be used as described in Sec. 7.4.

Probability footprint and threshold

The only difference of the method of this chapter and the on of Ch. 6, is the estimation of the footprint for the leakage locations we don't have simulations for. The probability footprints of the predicted leakage locations $\mathbf{x}_1^*, \ldots, \mathbf{x}_K^*$ are calculated exactly as in Ch. 6. This implies that the discussion of Sec. 6.5 on the probability footprint and the choice of probability threshold p_t is the same.

Spatial dependence of footprint

Ch. 6 presented the event method originally introduced in Hvidevold et al. [7]. This method used the same footprint for all leakage locations, similarly to

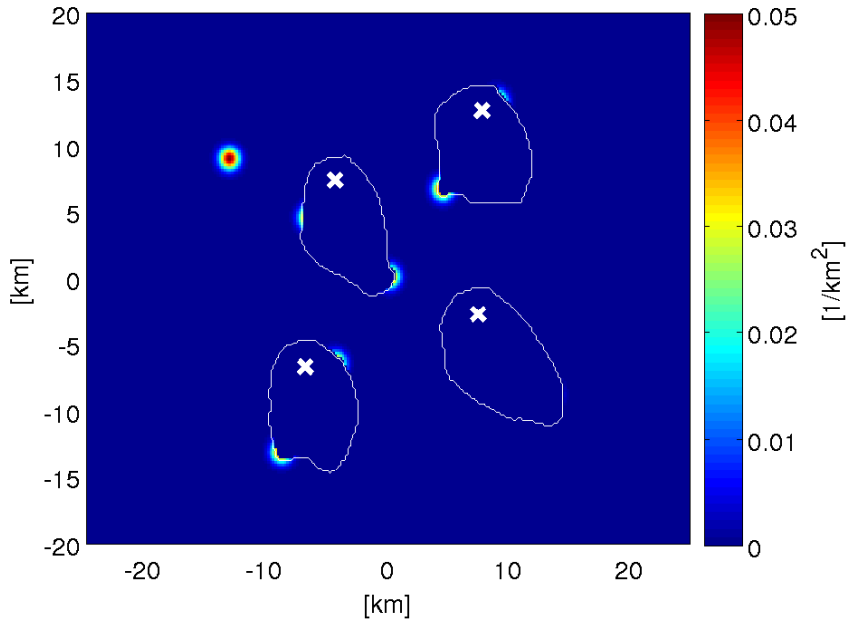


Figure 7.9: Residual leakage location probability density function for the optimal sensor layout of four sensors (white crosses). The monitored regions of the sensors are zeroed out and the boundaries of the monitored regions are marked with white lines. Compare with the original leakage location probability density function in Fig. 7.6 to see the monitored (zeroed) probability. A probability threshold of $p_t = 0.1$ is used. The detection probability is $p = 0.66$.

how the same average concentration footprint was used for all leakage locations in Ch. 4. Just as the need to introduce spatial dependence in the average concentration footprint was addressed in Sec. 4.7, Sec. 6.5 suggests that spatial dependence is incorporated in the probability footprint for the event method. This has successfully been done in this chapter using the theory of Ch. 5. Since this chapter is to Ch. 6 as Ch. 5 is to Ch. 4, the discussion on spatial dependence in Sec. 5.6 also applies to this chapter.

Chapter 8

Discussion and further work

8.1 Developed methods for monitoring design

Fig. 8.1 shows a summary of the methods for monitoring design presented in Ch. 3 to Ch. 7. The methods are constructed using different combinations of footprints, inversion and spatial dependency on the leakage location in the footprint estimation. Each of these features are discussed later in this chapter.

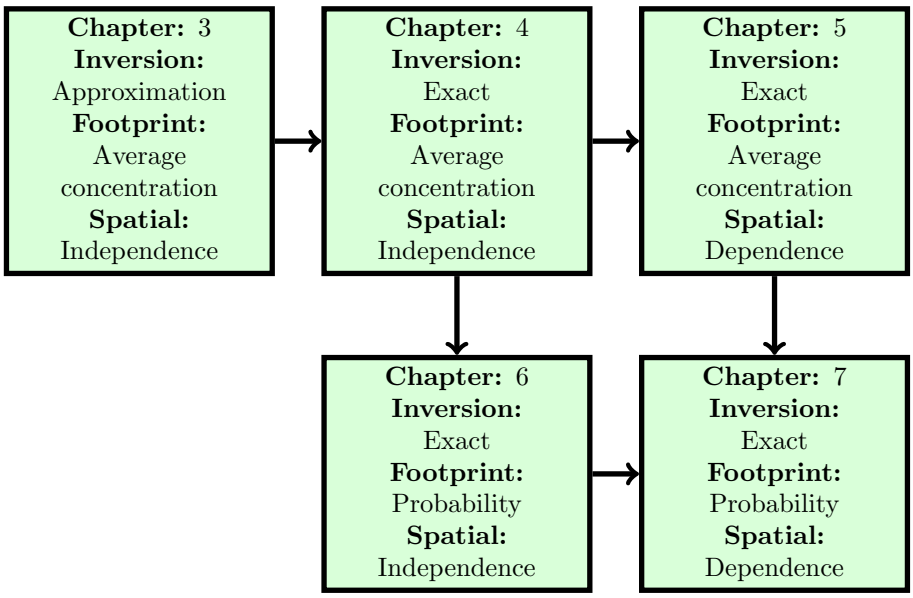


Figure 8.1: Methods for monitoring design.

First of all, we have shown that the footprint should be inverted exactly as discussed in Sec. 8.4. This was the motivation for Ch. 4 where the first improvement of the original method from [8] is developed. After having developed the exact inversion of Ch. 4, there is no reason to pursue other approximations as the one of Ch. 3.

The next thing developed in Ch. 5 was spatial dependence on the leakage location in the average concentration footprint using interpolation. As discussed

in Sec. 8.3, this is a simple way to introduce spatial dependence in the footprint, but will be computationally more expensive than the method of Ch. 4.

We note that the method of Ch. 5 reduces to the method of Ch. 4 if only one footprint is used in the interpolation. For this reason, Ch. 4 can be viewed as a special case of Ch. 5.

As discussed in Sec. 8.2, the use of the probability footprint and the event method is a more realistic approach than the use of the average concentration footprint. For this reason, the method of Ch. 6 where the exact inversion of Ch. 4 is applied to the probability footprint, was a natural next step.

The final method of Ch. 7 extends Ch. 6 similarly as Ch. 5 extends Ch. 4. The direct connection between Ch. 5 and 7 in Fig. 8.1 also emphasises that those two methods are the same except for the type of footprint used.

Due to the use of different test cases, we can not compare the output of all the methods directly. The use of different test cases has been necessary due to limited available test data. Some of the test data was made available to us after we had completed the testing of some methods. Ideally, we would of course have liked to perform optimal design with all the methods using the same test data. However, we have obtained good insight in how the different methods work without doing so.

If desirable, one or several of the suggestions for further work in the following sections can be combined to create new methods for monitoring design.

8.2 Footprint prediction and detection function

To estimate the leakage footprint for an arbitrary leakage location, we need one or several predicted footprints to base the estimation on. It is expensive to obtain such predictions experimentally, limiting the number of leak scenarios that can be predicted experimentally. For this reason, we must rely on simulations from numerical models to predict footprints. In our case the test data is from simulations with Bergen Ocean Model (BOM). The leakage locations of the predicted footprints should ideally be chosen based on the leakage location probability density function as is further discussed in Sec. 8.3 and 8.5.

The prediction of the footprint can be divided into two parts. The first is the simulation where a tool like BOM is used to obtain time series of the concentration originating from the leakage, e.g. Fig. 6.2. For some methods, it may also be interesting with information such as the time series for currents.

The second part of the prediction is to calculate the footprint from the data obtained from the simulation. In our work, we have a statistical and stationary footprint independent of time. This implies that the time series are reduced to only one value for each grid point. We have used two different methods for this, the average concentration footprint of Ch. 3 to 5 and the probability footprint of Ch. 6 and 7. The data sets used are listed in Tab. C.1.

Average concentration footprint

A leakage is modelled in Hvidevold et al. [8] by the time average concentration footprint resulting from the leakage. This approach is adopted in Ch. 3 to 5. Fig. 2.2 shows the average concentration footprint used in [8] and also in Ch. 3 and 4. For our test data, the average is taken over approximately two months.

The detection function for the average concentration is given by Eq. 3.1. It has a sharp boundary between where the leakage is detectable and not. The detectable region of the footprint in Fig. 2.2 can be seen to the left in Fig. 3.2.

As discussed in Sec. 3.7, the approach of using average concentration is too conservative. Even though the long term average concentration at a location is below c_t, the concentration may be above c_t for statistically significant periods of time. This implies that the leakage will in real life be detectable in a larger region than the average concentration approach indicates.

An other problem with the average concentration approach is the sharp boundary of the detectable region. It is not realistic that all sensors within a region always will detect the leakage while all sensors outside this region never will detect the leakage. However, this is a necessary consequence of applying the concentration threshold c_t directly to the average concentration footprint. In addition, it gives easy calculations for the average concentration method.

Probability footprint

Having discussed the insufficiency of the average concentration footprint, Ch. 6 and 7 use a different footprint presented in Hvidevold et al. [7]. The probability footprint given by Eq. 6.1 is the fraction of time where the concentration is above the detection threshold c_t. The footprint used in [7] and Ch. 6 can be seen in Fig. 6.3.

The event method assumes that a leakage is detected if the concentration is above c_t a sufficient amount of the time. Or, in other words, if the probability footprint is above a threshold probability p_t. This gives the detection function of Eq. 6.5 for the event method. As discussed in Ch. 6, this detection function is similar to Eq. 3.1 for the average concentration, only with a different type of footprint as input.

The event method is considered more realistic than the method using the average concentration. For the test data used, the event method estimates a larger detectable region than the average concentration method. However, we still have a sharp boundary between where the leakage is detectable and not.

It should also be noted that for the calculation of the probability footprint, the complete time series of the concentration in the grid points surrounding the leakage are needed. However, only one value at each grid point is extracted from the time series. Having the complete time series available, it could be beneficial to use this opportunity to do more calculations or construct more complicated methods. Especially the quantification of the correct threshold probability p_t could be improved by applying time series analysis.

Further work on footprint prediction

How to model a leakage and what footprint predictions to use is of great importance for monitoring design. This is what determines where sensors are assumed to be able to detect leakages and not. Based on this, which footprint and detection function to use is the most important part to focus further research on. If the footprint and detection function model a leakage poorly, the importance of the estimation of spatial dependency and inversion of the footprint become less important.

One of the problems addressed for both the average concentration and event method, is the sharp boundary of the detectable region. We have argued that this should in real life be a smooth transition where the detection function varies smoothly from 1 at the leakage to 0 far from the leakage.

There are several possibilities for constructing a more smooth detection function. One could for instance apply two threshold values. The leakages could be assumed to always be detected at all locations having a value above the higher threshold value and never detected at all location having a value below the lower threshold value. Between the two threshold values, the detection function could vary smoothly between 0 and 1. For instance, assuming the detection function to be linear in this region gives using average concentration

$$D(\mathbf{x};\mathbf{x}_0;c_t) = \begin{cases} 1 & \text{if } \bar{C}(\mathbf{x};\mathbf{x}_0) \geq c_t^h \\ \frac{\bar{C}(\mathbf{x};\mathbf{x}_0)-c_t^l}{c_t^h-c_t^l} & \text{if } c_t^l \leq \bar{C}(\mathbf{x};\mathbf{x}_0) < c_t^h \\ 0 & \text{if } \bar{C}(\mathbf{x};\mathbf{x}_0) < c_t^l \end{cases} \tag{8.1}$$

where c_t^l is the lower and c_t^h the higher concentration threshold value.

This is only one of endless possibilities, and even more complicated footprints and detection functions should be investigated to model leakages better.

Another large drawback of the predicted footprints used, is that they are without dependence of time. Throughout the year, there could be seasonal variability in the footprint. In addition, the time series used for calculation of the footprints are about two months long in this work. Reducing them to only one footprint value in each grid point is a major simplification. One could also use measurements to do live updates of the predicted footprints.

Even though seasonal variability is introduced in the footprint prediction, it may be hard to design a monitoring program accounting for it. This would require the sensors to be moved several times a year to account for the variability. However, this is expensive, and calls for a fixed sensor layout.

8.3 Spatial dependency on leakage location in footprint estimation

Independently of the type of footprint used, we are not able to predict it for all leakage locations by doing simulations. As discussed in Ch. 5, this would be computationally too demanding as we have a large number of possible leakage locations. Instead, we need to estimate the footprint of an arbitrary leakage location from the predicted footprints of fewer leakage locations. How to do this is an important question when developing a monitoring program.

Spatial independence by approximation

The original method from [8] of Ch. 3 assumes that all leakage locations have the same footprint predicted by a simulation. To enable easy inversion, this footprint was approximated by a Gaussian function given by Eq. 3.5.

It is suggested in [8] to try other approximation functions. The main reason of the approximation in [8] is to enable easy inversion. As discussed in Sec. 8.4, we have developed an exact inversion that can be applied to any footprint.

Knowing this, there is no longer a reason for approximating the footprint when we can use the simulated one directly. This is also the reason why the testing of other approximation functions than Eq. 3.5 is not included in this work.

Spatial dependence by interpolation

A method based on interpolation is developed in Ch. 5. Interpolation is the natural choice for introducing spatial dependence using a limited number of predicted footprints. Using predicted footprints for K different leakage locations, the interpolation is given by Eq. 5.13.

The weight functions given by Eq. 5.7 are only dependent on the distances from the leakage location to the locations of the predicted footprints. However, the weights are calculated before performing the optimization giving a possibility to account for local effects that should influence the weights. A rough topography could for instance cause two close locations to have completely different footprints.

Having a large area to be monitored, the need for proper spatial dependence in the footprint increases. The test footprints of Ch. 5 shown in Fig. 5.2 and 5.5 illustrates this having large variation between them.

An important question when using Eq. 5.13 to estimate the footprint, is for how many and what locations footprints should be predicted by simulations. Using many footprints to interpolate increases the accuracy of the estimation, but also increases the numerical cost when performing optimization. The extra numerical cost of spatial dependence in the footprint compared with spatial independence is discussed in Sec. 8.6.

The most important is that the footprints of the locations having a high probability of leakage, are estimated well. For this reason, choosing the locations of the predicted footprints should ideally be done after the calculation of the leakage location probability density function such that the footprints in areas of high leakage probability are estimated well.

Spatial independence by translation

A method assuming spatial independence of the footprint is developed in Ch. 4. Assuming the footprint to be equal for leakage locations, the estimation of the footprint from a leakage is only a translation of a predicted footprint known from simulations. This translation is given by Eq. 4.22 and illustrated in Fig. 4.2. We note that Eq. 5.13 for the interpolation reduces to Eq. 4.22 when using only $K = 1$ footprint. The method of Ch. 4 can thus be looked on as a special case of the method in Ch. 5. However, computations are a lot faster assuming spatial independence as is shown in App. B.

Further work on spatial variability

The spatial dependence of the leakage footprint is an important part of monitoring design that should be studied further. Sufficient spatial dependence should be included without making it computationally too demanding.

It is natural to base the estimation on interpolation between predicted footprints as in Eq. 5.13, but there are many things that can be experimented with. One of the most obvious things is where the predicted footprints we

interpolate between are placed. Another is the calculation of the weight functions, possibly including other information than the distances to the predicted footprints.

One could also make more advanced methods not only using interpolation. A method tested, but not included in this work, rotates the footprint according to the topography in addition to interpolation. It could also be a natural extension to include methods of geostatistics, e.g. from [6].

A completely different approach, would be to divide the area to be monitored into smaller regions and apply simpler estimation in each of these regions. For instance, one could apply spatial independence within each of the regions. The problem with this is how to couple the regions together, both when evaluating the sensor layout optimization cost function and to avoid unwanted boundary effects in the estimation. If this can be solved in a good way, it would probably be computationally beneficial over the methods based on interpolation.

8.4 Inversion of the footprint

One of the most important calculations in the work done is the inversion of the footprints obtained for the leakages. Let as in Sec. 4.5 $g(\mathbf{x};\mathbf{x}_0)$ be the value at $\mathbf{x}$ of any kind of footprint resulting from a leakage at $\mathbf{x}_0$. Our main goal in this work is to find optimal sensor layouts and estimate the probability of the layout to detect a leakage at an unknown location. To do this, we need to know what potential leakage locations are monitored by sensors at given locations. This leads to the inversion of the footprint. We now need to know the footprint value $g(\mathbf{x}_s;\mathbf{x})$ at the sensor location $\mathbf{x}_s$ for all possible leakage locations $\mathbf{x}$.

Approximated inversion

In Ch. 3, the inversion was done using an approximation of the footprint from [8]. This inversion was based on the two assumptions that the footprint is the same for all leakage locations and that the footprint can be approximated by a Gaussian function. The approximation was discussed in Sec. 3.7 using the average concentration footprint from Hvidevold et. al. [8] shown in Fig. 2.2 as test case. In accordance with the discussion in [8], the approximation is not found to be satisfactory. The anisotropy and area of the detectable region is badly approximated, as can be seen in Fig. 3.2 for $c_t = 2.26 \cdot 10^{-4}$ kg/m^3.

It is suggested in Hvidevold et al. [8] that other approximation functions with better abilities to approximate the anisotropy of the footprint are tested. As noted in Sec. 3.7, but not included in the thesis, such other approximation functions are tested without giving significantly better approximations. In addition, the choice of the proper approximation function depends on the footprint at hand. Since we want to develop a general method for any site and footprint, it is impossible to find a suitable approximation function. Based on these considerations, it is suggested in Sec. 3.7 that the footprint is inverted exactly.

Exact inversion

A method for exact inversion of the footprint is developed in Ch. 4 giving the inverted footprint of Eq. 4.23. Just as the method of Ch. 3, this method is based on the assumption that all leakage locations have the same footprint.

The inverted footprint of Eq. 4.23 has a geometric interpretation of rotating the original footprint 180° as explained in Sec. 4.4 and illustrated in Fig. 4.5. This interpretation is useful for the calculation of the inverted footprint, as the footprint can be rotated and saved beforehand leaving only the translation to the sensor location to be done when the inverted footprint for a sensor location is to be calculated. By doing this, it is actually just as easy to calculate the exact inverted footprint as the approximate inverted footprint used in Ch. 3 and [8]. Being able to calculate the exact inverted footprint without any extra cost, the inversion procedure of Ch. 4 is a major result of this work. After the exact inversion is developed in Ch. 4, we no longer use the approximation of Ch. 3 in the further methods developed.

Exact inversion for spatial dependence

The inversion of Eq. 4.23 is based on the assumption that the footprint is independent of the leakage location. This implies that the inversion can only be applied directly to the method of Ch. 6 since this is the only other method with spatial independence of the footprint. However, the footprints of Ch. 5 and 7 with spatial dependence can be inverted using Eq. 4.23 as basis.

The interpolated footprint given by Eq. 5.13 can be inverted by inverting each of the footprints used in the interpolation to obtain the inverted footprint given by Eq. 5.14. Just as before, the known footprints can be inverted (rotated) and saved beforehand. When we want to calculate the inverted footprint of a sensor location $\mathbf{x}_s$, all the known inverted footprints can be translated to the sensor location and then interpolated between. Note, however, as mentioned before, that the argument $\mathbf{x}$ of the weight functions is the independent spatial variable. This makes the calculations slightly more costly, as the interpolation can't be done beforehand. However, to include spatial independence in the footprint we must expect the calculations to be more costly. Without exact inversion, one could also argue that the introduction of spatial independence in the footprint is useless. For this reason, the exact inversion of footprints including spatial independence is of great importance.

Further work on footprint inversion

Having obtained an exact inversion for the methods in Ch. 4 to 7, there is not more mathematical work to be done on the inversion used in these methods. Once the footprint is predicted, an exact inversion is the best we can ask for. However, fast implementation of the inversion has not been discussed in this work. An optimal algorithm for inversion can be extremely useful, allowing for faster optimization of the sensor layout, especially when interpolating between many footprint. If other methods than the ones of this work are developed, one should use an exact inversion.

8.5 The leakage scenarios

We have stressed that we are not doing site studies in this work, only developing methods for calculating the optimal sensor layout and detection probability for a given leakage location probability density function. However,

a proper calculation of the leakage location probability density field is of great importance when performing monitoring design.

Test cases used

Ch. 3 to 6 use the test case introduced in Sec. 2.1.3. This test case is based on a map of faults and wells from a public web page. The leakage location probability density function is created by assuming that all wells and all faults have the same probability of being the leakage location, with wells having a higher probability than faults. It is further assumed that no other locations can be the location of the leakage. This is a very simple approach only suitable for test cases.

There is no distinction between wells and faults reaching and not reaching the storage formation. In addition, the area close to the injection well would most likely have a higher probability of being the leakage location than further away from the injection well. It is also a simplification that the wells in the test case of Sec. 2.1.3 are modelled as points, whereas they in real life give an increased probability of leakage in a region around the well.

Since the test case of Sec. 2.1.3 is of no particular interest, we have used a different leakage location probability density function in Ch. 7. This function is completely synthetic and not based on a map like the one previously used. The nine locations having predicted footprints in Ch. 7 are each assumed to be the mode of an Gaussian distribution. In addition, a background probability is added everywhere to obtain the leakage location probability density function as described in Sec. 7.4.

The approach of a background probability and a distribution about each of the wells being a possible leakage location is probably a much better approach than the simplified use of a map in Sec. 2.1.3. However, the cumbersome part of the process is still to identify the wells being a possible leakage location and quantify the relative probability between them.

Further work on leakage scenarios

When doing monitoring design for a storage site, the identification of possible leakage location is really important. If the leakage location probability density function is inaccurate, sensors could be placed to monitor regions that in real life have a low probability of a leakage. More critical than that, regions of high probability of a leakage may end up not being monitored. In addition, the leakage location probability density function should be a deciding factor when choosing which locations to predict the footprint for as discussed in Sec. 8.3. For these reasons, how to calculate the leakage location probability density function should absolutely be studied further. This includes studies of the overburden to identify leakage pathways and quantify their leakage probabilities.

8.6 Computational considerations

The work has been done with a mathematical approach, not focusing on implementation of the methods developed. However, the implementation will

of course be important when having a large optimization problem as Eq. 2.29. Some computational considerations can be found in App. B.

Further work

As more complicated methods are developed and optimization should be performed for a large number of sensors, the numerical implementation becomes important. The simplest way to speed up the optimization is to minimize the evaluation time of the cost function by smart implementation, includes developing good data structures.

The Genetic Algorithm is probably a good choice of optimization algorithm, but other algorithms should absolutely be assessed. It would be especially useful if the chosen algorithm has a numerical toolbox allowing for parallel computing. Finally, investigating when the recursive optimization procedure of Eq. 2.30 is a good approximation could allow for improved optimization when it applies.

8.7 Closing remarks

Several possibilities for further work have been suggested in the previous sections. Of the different features of monitoring design, the footprint prediction and detection function should be the focus of further research. As discussed in Sec. 8.2, having a wrong type of footprint or detection function disables the whole method for monitoring design of performing well independently of the rest of the method. Investigating new types of footprint, possibly with dependency on time, should be a natural extension of this work.

Another subject, is the spatial dependence on the leakage location in the estimation of the footprint as discussed in Sec. 8.3. More advanced methods for this estimation of the footprint could give better results than the ones used in the methods of this work. However, one should beware the extra computational cost calling for longer optimization time and compare this with the gain of a more advanced estimation.

The numerical improvements discussed briefly in Sec. 8.6 could speed-up calculations and allow for more advanced methods or faster optimization. This has not been studied in this work, allowing for a large possibility of topics to be studied. Finally, the prediction of the footprint by simulation is the most important numerical calculation, and should of course also be studied further.

Bibliography

[1] Recent greenhouse gas concentrations. `http://cdiac.ornl.gov/pns/current_ghg.html` Accessed: 2015-09-14.

[2] Ocean acidification. `https://en.wikipedia.org/wiki/Ocean_acidification` Accessed: 2015-09-14.

[3] Guttorm Alendal, Marius Dewar, Alfatih Ali, Yakushev Evgeniy, Lisa Vielstädte, Helge Avlesen, and Baixin Chen. Technical report on environmental conditions and possible leak scenarios in the north sea. Tech. Rep. D3.4, ECO2 deliverables, 2013. `http://www.eco2-project.eu`.

[4] Alfatih Ali, Håvard G. Frøysa, Helge Avlesen, and Guttorm Alendal. Simulating spatial and temporal varying CO_2 signals from sources at the seafloor to help designing risk based monitoring programs. *Journal of Geophysical Research - Oceans*, 2015. Submitted.

[5] J. Berntsen. Users guide for a modesplit sigma-coordinate numerical ocean model. Tech. rep., Department of Mathematics, University of Bergen, 2004.

[6] Noel A. C. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. Wiley, 1993. ISBN 9780471002550.

[7] Hilde Kristine Hvidevold, Guttorm Alendal, Truls Johannessen, and Alfatih Ali. Survey strategies to quantify and optimize detecting probability of a CO_2 seep in a varying marine environment. *Env. Modelling & Software*, 2015. Submitted.

[8] Hilde Kristine Hvidevold, Guttorm Alendal, Truls Johannessen, Alfatih Ali, Trond Mannseth, and Helge Avlesen. Layout of CCS monitoring infrastructure with highest probability of detecting a footprint of a CO_2 leak in a varying marine environment. *International Journal of Greenhouse Gas Control*, 37:274–279, jun 2015.

[9] B. Metz and Intergovernmental Panel on Climate Change. Working Group III. *Carbon Dioxide Capture and Storage: Special Report of the Intergovernmental Panel on Climate Change*. 25 CM. Cambridge University Press, 2005. ISBN 9780521866439.

[10] Bert Metz, Ogunlade Davidson, Heleen de Coninck, Manuela Loos, and Leo Meyer. *Carbon Dioxide Capture and Storage*. 2005.

[11] Zbigniew Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1996. ISBN 9783540606765. Chapter 1-3.

[12] ECO_2 project. Best Practice Guidance for Environmental Risk Assessment for offshore CO_2 geological storage. 2015.

Appendix A

Symbols used

Symbol	Explanation	Reference
A	Area to be monitored.	Sec. 2.1
$A(\mathbf{z})$	Parameter matrix for approximation of average concentration.	Sec. 3.3
A_H	Horizontal eddy diffusivity.	App. C
$B(\mathbf{x}_0)$	Normalization value at $\mathbf{x}_0$ for interpolation.	Sec. 5.1
BOM	Bergen Ocean Model.	App. C
CCS	Carbon Capture and Storage.	Sec. 1.1
c_t	Concentration threshold of sensors.	Sec. 2.2.1
c_t^h	Higher concentration threshold.	Sec. 8.2
c_t^l	Lower concentration threshold.	Sec. 8.2
C	CO_2 Concentration.	App. C
$\bar{C}(\mathbf{x};\mathbf{x}_0)$	Average concentration at $\mathbf{x}$ resulting from a leakage at $\mathbf{x}_0$.	Ch. 3-5
$\bar{C}_i$	Simulated average concentration in grid point i.	Sec. 3.3
$\bar{C}_0$	Simulated average concentration at leakage point.	Sec. 3.3
$C_l(\mathbf{x};\mathbf{x}_0)$	Concentration time series value at $\mathbf{x}$ at time step l for a leakage at $\mathbf{x}_0$.	Sec. 6.1
$d_k(\mathbf{x}_0)$	Distance from $\mathbf{x}_0$ to the leakage location $\mathbf{x}_k^*$ of predicted footprint k.	Sec. 5.1
$D(\mathbf{x};\mathbf{x}_0;c_t)$	Detection function, the probability that a leakage at $\mathbf{x}_0$ is detected by a sensor at $\mathbf{x}$ with a concentration threshold c_t.	Sec. 2.2.2
$D_A(\mathbf{x};\mathbf{x}_0;c_t)$	Detection function using average concentration.	Sec. 3.1
$D_A^A(\mathbf{x};\mathbf{x}_0;c_t)$	Detection function using approximation of an average concentration footprint.	Sec. 3.4
$D_A^T(\mathbf{x};\mathbf{x}_0;c_t)$	Detection function using translation of an average concentration footprint.	Sec. 4.2

$D_A^I(\mathbf{x};\mathbf{x}_0;c_t)$	Detection function using interpolation of average concentration footprints.	Sec. 5.2
$D_S(\mathbf{x};\mathbf{x}_0;c_t)$	Probability footprint, detection function for a single measurement.	Sec. 6.1
$D_E(\mathbf{x};\mathbf{x}_0;c_t;p_t)$	Detection function for the event method using the probability footprint $D_S(\mathbf{x};\mathbf{x}_0;c_t)$ and probability threshold p_t.	Sec. 6.2
$D_S^T(\mathbf{x};\mathbf{x}_0;c_t)$	Probability footprint estimated using translation of a predicted footprint.	Sec. 6.1
$D_E^T(\mathbf{x};\mathbf{x}_0;c_t;p_t)$	Detection function for the event method using the translated probability footprint $D_S^T(\mathbf{x};\mathbf{x}_0;c_t)$.	Sec. 6.2
$D_S^I(\mathbf{x};\mathbf{x}_0;c_t)$	Probability footprint estimated using interpolation of predicted footprints.	Sec. 7.1
$D_E^I(\mathbf{x};\mathbf{x}_0;c_t;p_t)$	Detection function for the event method using the interpolated probability footprint $D_S^I(\mathbf{x};\mathbf{x}_0;c_t)$.	Sec. 7.2
$f = f(\mathbf{x})$	Leakage location probability density function.	Sec. 2.1.1
$f_n = f_n(\mathbf{x})$	Residual leakage location probability density function of $f(\mathbf{x})$ after placing n sensors.	Sec. 2.2.4
f_{Back}	Leakage background probability.	Sec. 7.4
f_{Norm}	Probability normalization value.	Sec. 7.4
GA	Genetic Algorithm.	Sec. 2.2.5
$g(\mathbf{x};\mathbf{x}_0)$	Value at $\mathbf{x}$ of an arbitrary footprint from a leakage having leakage location $\mathbf{x}_0$.	Sec. 4.5 Sec. 5.4
$G(\mathbf{x};\mathbf{x}_0;\mathbf{z})$	Approximated average concentration at $\mathbf{x}$ from a leakage at $\mathbf{x}_0$ using parameter vector $\mathbf{z}$.	Sec. 3.3
G_i	Approximated concentration in grid point i.	Sec. 3.3
i	Grid point index.	Sec. 3.3
I	Number of grid points.	Sec. 3.3
$I_D(C_l(\mathbf{x};\mathbf{x}_0);c_t)$	Detection indicator function.	Sec. 6.1
k	Index of predicted footprint.	Ch. 5 Ch. 7
K	Number of predicted footprints.	Ch. 5 Ch. 7
K_H	Vertical eddy diffusivity.	App. C
l	Time step index.	Sec. 6.1
L	Number of time steps.	Sec. 6.1
$M = M(\mathbf{x};\mathbf{x}_s;c_t)$	Monitoring function, the probability that a sensor at $\mathbf{x}_s$ with a threshold concentration c_t detects a leakage at $\mathbf{x}$.	Sec. 2.2.3
$M_n = M(\mathbf{x};\mathbf{x}_s^n;c_t)$	Monitoring function for sensor n.	Sec. 2.2.4

$M_A(\mathbf{x};\mathbf{x}_s;c_t)$	Monitoring function using average concentration.	Sec. 3.2
$M_A^A(\mathbf{x};\mathbf{x}_s;c_t)$	Monitoring function using approximation of an average concentration footprint.	Sec. 3.5
$M_A^T(\mathbf{x};\mathbf{x}_s;c_t)$	Monitoring function using translation of an average concentration footprint.	Sec. 4.3
$M_A^I(\mathbf{x};\mathbf{x}_s;c_t)$	Monitoring function using interpolation of average concentration footprints.	Sec. 5.3
$M_E(\mathbf{x};\mathbf{x}_s;c_t;p_t)$	Monitoring function for the event method using the probability footprint $D_S(\mathbf{x};\mathbf{x}_0;c_t)$ and probability threshold p_t.	Sec. 6.3
$M_E^T(\mathbf{x};\mathbf{x}_s;c_t;p_t)$	Monitoring function for the event method using the translated probability footprint $D_S^T(\mathbf{x};\mathbf{x}_s;c_t)$.	Sec. 6.3
$M_E^I(\mathbf{x};\mathbf{x}_s;c_t;p_t)$	Monitoring function for the event method using the interpolated probability footprint $D_S^I(\mathbf{x};\mathbf{x}_0;c_t)$.	Sec. 7.3
n	Sensor index	
N	Number of sensors in layout.	
$p = p(\mathbf{x}_s^1;\ldots;\mathbf{x}_s^N;c_t)$	Detection probability, the probability that a leakage in an arbitrary location is detected by the N sensors.	Sec. 2.2.4
p_t	Probability threshold used in the event method.	Sec. 6.2
Q	Source flux rate.	App. C
$\mathbf{r}$	Displacement vector.	Sec. 4.4
R	Rotation matrix.	Sec. 4.4
$\vec{U}$	Velocity field.	App. C
$w_k(\mathbf{x_0})$	Weight function for predicted footprint k when estimating a leakage in $\mathbf{x}_0$.	Sec. 5.1
$\mathbf{x}$	Coordinate vector for independent variable.	
$\mathbf{x}_0$	Coordinate vector for leakage location.	
$\mathbf{x}^*$	Coordinate vector for leakage location of predicted footprint.	Ch. 4 Ch. 6
$\mathbf{x}_k^*$	Coordinate vector for leakage location of predicted footprint k.	Ch. 5 Ch. 7
$\mathbf{x}_s$	Coordinate vector for location of sensor.	
$\mathbf{x}_s^n$	Coordinate vector for location of sensor n.	
x	x-coordinate for independent variable.	
x_0	x-coordinate for leakage location.	
x_k^*	x-coordinate for leakage location of predicted footprint k.	Ch. 5 Ch. 7
y	y-coordinate for independent variable.	

y_0	y-coordinate for leakage location.	
y_k^*	y-coordinate for leakage location of predicted footprint k.	Ch. 5 Ch. 7
$\mathbf{z}$	Parameter vector for approximation.	Sec. 3.3
$\mathbf{z}_n$	Optimization parameter vector for sensor n.	Sec. 2.2.5
α	Angle for rotation matrix R.	Sec. 4.4
σ	Shape parameter for Gaussian distribution.	Sec. 7.4
σ_x	Length parameter for approximated average concentration.	Sec. 3.3
σ_y	Length parameter for approximated average concentration.	Sec. 3.3
$\hat{\sigma}_x$	Semi axis of detectable region using approximated average concentration.	Sec. 3.4
$\hat{\sigma}_y$	Semi axis of detectable region using approximated average concentration.	Sec. 3.4
Σ_i	Standard deviation of simulated concentration in grid point i.	Sec. 3.3
θ	Angle parameter for approximated average concentration.	Sec. 3.3

Appendix B

Computational considerations

The work has been done with a mathematical approach, not focusing on implementation of the methods developed. However, the implementation will of course be important when having a large optimization problem as Eq. 2.29. The simulations of the leakage scenarios used are assumed given, and are for that reason not part of the discussion on the numerical implementation.

Evaluation of cost function

A numerical optimization routine such as the Genetic Algorithm used in this work optimizes a function be evaluating the function a large number of times for different parameters. The ability of the optimization to find the layout having the highest detection probability, will thus highly depend on the time it takes to evaluate the cost function. A faster evaluation of the cost function would enable more function evaluations in the same time span, alternatively shorter run time for the same number of evaluations.

The methods of Ch. 4 and 6 assumes spatial independence of the footprint, leaving only one footprint to be used. This gives the same cost function for the optimization for these methods, only with different input files. The methods of Ch. 5 and 7 introduce spatial dependence, implying that several footprints are used and interpolated between. This gives a more complicated cost function for the two methods with spatial dependence. For this reason, the optimization of the sensor layout will be computationally more demanding when applying spatial dependence.

To compare the cost function with spatial independence and dependence, we estimate the evaluation time of the cost functions. The two different cost functions are evaluated 240000 times for each number of sensors from one to twelve. We use the test case of Sec. 2.1.3 as leakage location density field. For the spatial independence, we use the footprint of Fig. 2.2 as in Ch. 4. For the spatial dependence, we use the four footprints of Ch. 5. The evaluation time of the cost function for spatial dependence will of course increase if a large number of footprints are used for interpolation.

The results are shown in Tab. B.1 and plotted in Fig. B.1. We see that the

	Time per evaluation [s]		
N	Spatial independence	Spatial dependence	Factor
1	$5.45 \cdot 10^{-3}$	$5.23 \cdot 10^{-2}$	9.60
2	$8.99 \cdot 10^{-3}$	$1.02 \cdot 10^{-1}$	11.37
3	$1.22 \cdot 10^{-2}$	$1.53 \cdot 10^{-1}$	12.51
4	$1.63 \cdot 10^{-2}$	$2.04 \cdot 10^{-1}$	12.52
5	$1.92 \cdot 10^{-2}$	$2.57 \cdot 10^{-1}$	13.35
6	$2.34 \cdot 10^{-2}$	$3.09 \cdot 10^{-1}$	13.18
7	$2.65 \cdot 10^{-2}$	$3.57 \cdot 10^{-1}$	13.45
8	$3.05 \cdot 10^{-2}$	$4.08 \cdot 10^{-1}$	13.41
9	$3.37 \cdot 10^{-2}$	$4.60 \cdot 10^{-1}$	13.66
10	$3.75 \cdot 10^{-2}$	$5.14 \cdot 10^{-1}$	13.70
11	$4.05 \cdot 10^{-2}$	$5.58 \cdot 10^{-1}$	13.78
12	$4.50 \cdot 10^{-2}$	$6.21 \cdot 10^{-1}$	13.82

Table B.1: Time per evaluation of the cost function used for optimization for different number N of sensors. Each function is evaluated 240000 times. For the spatial dependence, four footprints are used for interpolation.

current implementation is at least ten times faster for the spatial independence than for the spatial dependence using four footprints. We also see that the advantage is larger for a higher number of sensors. The extra computation time needed for a more complicated method should be considered when choosing what method to use.

Genetic Algorithm (GA)

We have used the Genetic Algorithm implementation *ga* in MATLAB as numerical optimization toolbox. As described in Sec. 2.2.5, a Genetic Algorithm is a stochastic optimization algorithm that uses properties of evolution. The algorithm has previously been used by Hvidevold et al. in [8] and [7] for the same optimization problem as in this work. This use was successful, and is the main reason why we have used this algorithm. However, the stochastic optimization algorithm including mutation seems to fit the optimization problem well as good solution have "large distance" between them in the solution space. As described in Sec. 2.2.5, it is found beneficial to apply a large mutation rate to compensate for this "large distance" between the feasible solutions.

Even though applying a large mutation rate makes *ga* perform well, it sometimes fail to find a good solution. For this reason, we have run the same optimization problem several times to verify the obtained solution. This can be neatly implemented using parallelization. There may be better optimization algorithms for this problem. However, the Genetic Algorithm is found to be sufficient for the test cases run and we have not studied other algorithms.

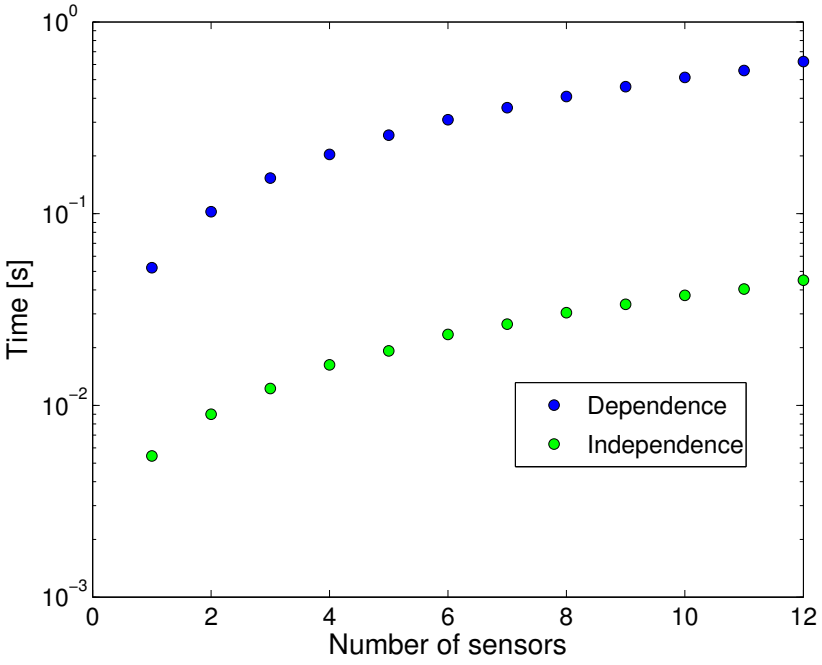


Figure B.1: Time per evaluation of the cost function used for optimization. Each function is evaluated 240000 times. For the spatial dependence, four footprints are used for interpolation.

Parallel computing

When having a large optimization problem like the one of this work, parallel computing can be an invaluable tool. Applying parallelization, the cost function can be evaluated for several different arguments simultaneously. However, this requires quite complicated communication between the different workers in order to be beneficial. Many numerical optimization toolboxes use have such built-in support for parallelization. This is not the case of the current implementation *ga* of the Genetic Algorithm in MATLAB used in this work. For this reason, we have not been able to apply parallelization directly.

As discussed in Sec. 2.2.5, we have been able to apply parallel computing even though the optimization routine we are using does not support it. Since we have found it necessary to run every optimization several times for verification, this allows for a simple parallelization. Each optimization problem is run independently on each of the workers. Having e.g. 12 workers, the optimization problem will be solved 12 times simultaneously. We then choose the best of the obtained solutions, and trust this to be close to the optimal solution.

Parallelization of the optimization is more important for complicated cost functions and high numbers of sensors to be placed. This implies that as more complicated methods are developed, parallelization will be of even greater value. When choosing what numerical optimization toolbox to use, support for parallel

computing may be a deciding factor.

Recursive placement of sensors

The optimization of the sensor layout is a large optimization problem having $2N$ degrees of freedom where N is the number of sensors:

$$\left[\mathbf{x}_s^1; \ldots; \mathbf{x}_s^N\right] = \text{argmax}\left(p(\mathbf{z}_1; \ldots; \mathbf{z}_N; f; c_t)\right) \qquad f, c_t \text{ fixed.} \tag{2.29}$$

This finds the optimal layout for all sensors at the same time. However, the intuitive approach will be to place one sensor at the time, always choosing the location having the highest detection probability. This breaks the problem into N recursive optimization problems of two degrees of freedom:

$$\mathbf{x}_s^n = \text{argmax}\left(p(\mathbf{z}_n; f_{n-1}; c_t)\right) \qquad n = 1, \ldots, N \qquad f, c_t \text{ fixed.} \tag{2.30}$$

Not only is the recursive approach more intuitive, but it could also be preferable numerically. However, it is found in Hvidevold et al. [8] that this may not give the optimal sensor layout. This is also discussed in Sec. 2.2.5, where it is found that the recursive method of Eq. 2.30 gives the same result as Eq. 2.29 if the monitoring functions of the sensor in the optimal layout have disjoint support.

Most of optimal sensor layouts obtained in Ch. 5 to 7 have sensors with small monitored regions. This gives disjoint, or almost disjoint, support of the monitoring functions of the sensors. Knowing this, we could have used the recursive optimization of Eq. 2.30. To be sure that the supports are disjoint for the optimal layout allowing for the recursive algorithm, we must however know the optimal layout beforehand. As discussed in Sec. 2.2.5, this is the main motivation why we always have used the full optimization problem of Eq. 2.29.

The time used per optimization of the test case of Ch. 5 for different number N of sensors is shown in Tab. B.2. Each optimization is performed 60 times using a population size of 400, a generation limit of 200, a relative tolerance of 10^{-3} and a mutation rate of 0.4. We see that the time used is almost the same, leaving it to the fitness of the obtained solutions to decide algorithm.

The detection probabilities obtained by the best optimizations from Tab. B.2 are plotted in Fig. B.2. We see that the detection probabilities are almost the same. Knowing from Tab. B.2 that the regular optimization of Eq. 2.29 actually is a little bit faster, it is natural to use this since it is the mathematically correct.

N	**Time per optimization** [s]		Factor
	Regular (Eq. 2.29)	Recursive (Eq. 2.30)	
1	$4.09 \cdot 10^1$	$4.08 \cdot 10^1$	1.00
2	$7.89 \cdot 10^1$	$8.19 \cdot 10^1$	1.04
3	$1.16 \cdot 10^2$	$1.23 \cdot 10^2$	1.06
4	$1.53 \cdot 10^2$	$1.64 \cdot 10^2$	1.07
5	$1.90 \cdot 10^2$	$2.05 \cdot 10^2$	1.08
6	$2.27 \cdot 10^2$	$2.45 \cdot 10^2$	1.08
7	$2.64 \cdot 10^2$	$2.87 \cdot 10^2$	1.09
8	$3.02 \cdot 10^2$	$3.28 \cdot 10^2$	1.09
9	$3.40 \cdot 10^2$	$3.69 \cdot 10^2$	1.08
10	$3.75 \cdot 10^2$	$4.11 \cdot 10^2$	1.09
11	$4.15 \cdot 10^2$	$4.52 \cdot 10^2$	1.09
12	$4.52 \cdot 10^2$	$4.91 \cdot 10^2$	1.09

Table B.2: Time per optimization for different number N of sensors for the test case of Ch. 5. Each optimization is performed 60 times using a population size of 400, a generation limit of 200, a relative tolerance of 10^{-3} and a mutation rate of 0.4.

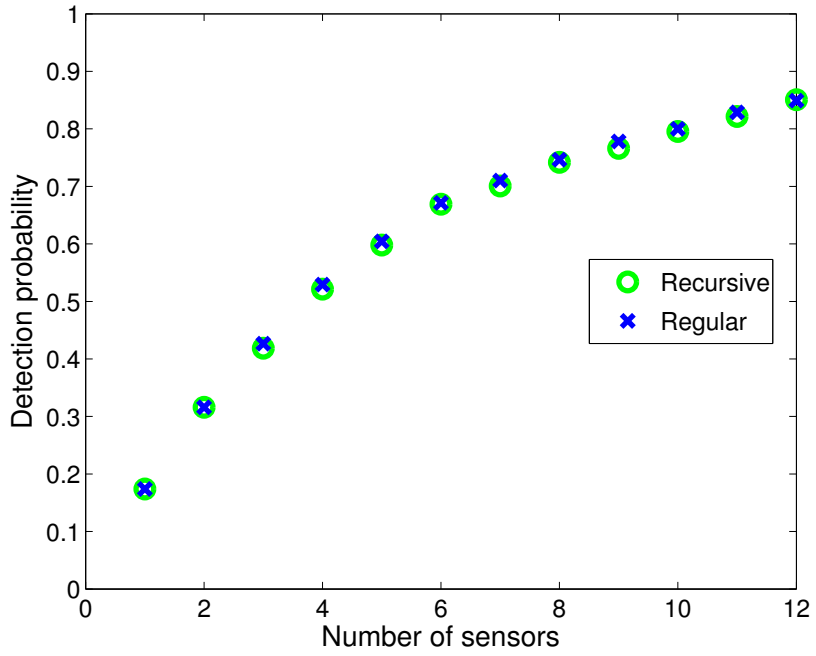


Figure B.2: Detection probabilities for the best sensor layouts obtained by the optimizations in Tab. B.2 for the test case of Ch. 5. The regular optimization (2.29) and the recursive optimization (2.30) give almost the same result.

Appendix C

Prediction of leakage scenarios

To design a monitoring program, we need to predict one or several leakage scenarios. Such predictions are expensive and hard to perform experimentally. For this reason, the predictions must be done by numerical simulations.

We have used data from simulations performed by Alfatih Ali [4] using the General Circulation Model Bergen Ocean Model (BOM). The data sets used are listed in Tab. C.1.

#	**Number of leakages**	**Available data**	**Used in chapter**
1	1	Average concentration	3, 4
2	4	Average concentration	5
3	1	Concentration time series	6
4	9	Probability footprint (Sec. 6.1)	7

Table C.1: Data sets used.

Bergen Ocean Model (BOM)

BOM is a three-dimensional terrain-following non-hydrostatic ocean model. It has capabilities of resolving processes from mesoscale to large scale. BOM is implemented in modern Fortran, and the source code is freely available from *http://www.mi.uib.no/BOM/*. For further information, see the user's guide [5].

For the data used in this work, the hydrostatic parallel version of BOM has been used. The spatial resolution is 800 meters and the sampling rate of the time series 1 min. CO_2 is released from the injection well and is assumed to be diluted enough to not influence the density of the sea water, implying that the concentration of CO_2 can be simulated as a passive tracer. This is modeled by the advection-diffusion equation

$$\frac{\partial C}{\partial t} + \vec{U} \cdot \nabla C = \frac{\partial}{\partial z}\left(K_H \frac{\partial C}{\partial z}\right) + \frac{\partial}{\partial x}\left(A_H \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(A_H \frac{\partial C}{\partial y}\right) + Q \qquad \text{(C.1)}$$

where C is the CO_2 concentration, Q the source flux rate, $\vec{U}$ the velocity field, A_H the horizontal eddy diffusivity and K_H the vertical eddy diffusivity.

For further information on the BOM setup and forcing, see [4].

Leakage scenarios

The leakage scenarios used are from the ECO_2 project [3] and the metocean data used for forcing are from spring 2012. Each scenario is simulated for two months using BOM as described in the previous section.

Most of the scenarios simulated have the chimney flux rate of $Q \sim 150$ T/day. Some of the methods only need simulations for one leakage location, whereas other need for several. For this reason, different data sets are used in the different chapters. Tab. C.1 shows an overview of the different available data sets used.