Quantifying uncertainties when monitoring marine environments in connection with geological storage of CO_2

Hilde Kristine Hvidevold

Dissertation for the degree of philosophiae doctor (PhD)

at the University of Bergen

2015

Dissertation date: 26.06.15

Preface

This dissertation is submitted as a partial fulfillment of the requirements for the degree Doctor of Philosophy (PhD) at the Department of Mathematics, University of Bergen. The subject of the thesis is on quantification of uncertainties when monitoring marine environments in connection with geological storage of CO_2. The thesis has been done in collaboration with the Geophysical Institute, University of Bergen and Uni Research CIPR, however the emphasis has been on mathematical aspects. The four included papers are based on the same mathematical theory; assessment of parameter uncertainties through linearized covariance analysis and optimization using genetic algorithms.

This work has been funded by SUCCESS centre for CO_2 storage under grant 193825/S60 from the Research Council of Norway.

The advisory committee has consisted of Guttorm Alendal (University of Bergen, Department of Mathematics), Truls Johannessen (University of Bergen, Geophysical Institute and Uni Climate) and Trond Mannseth (Uni Research CIPR).

Outline

The thesis is organized in two parts. Part I gives a motivation and a short introduction to the theory used in the papers in Part II.

The background theory in Part I is structured in the following way. The first chapter gives a short introduction to global warming and ocean acidification, ending with mitigation options. Chapter 2 introduces the basic concepts behind carbon capture, transportation and storage (CCS), focusing on the storage part. Thereafter monitoring of the marine environment is addressed. The chapter ends with a comment on the thesis contribution to CCS and monitoring of the marine environment. In Chapter 3 the equations describing rise and dissolution of a single CO_2 droplet/bubble in the water column are explained. The mathematical theory the four papers are founded on is introduced in Chapter 4. First linear covariance analysis and curvature measures of nonlinearity is explained, thereafter the optimization method, the genetic algorithm is briefly introduced. Part I concludes with a summary of the papers with an suggestion for further work, aiming to put the work into a broader perspective.

List of included papers in Part II of the thesis

Paper A:

Hvidevold, H.K., Alendal, G., Johannessen, T. and Mannseth, T, Assessing model parameter uncertainties for rising velocity of CO_2 droplets through experimental design, *International Journal of Greenhouse Gas Control* **11**, 283–289, (2012), doi:10.1016/j.ijggc.2012.09.008.

Paper B:

Hvidevold, H.K., Alendal, G., Johannessen, T. and Mannseth, T, Assessing model uncertainties through proper experimental design, *Energy Procedia*, **37**, 3439–3446, (2013), doi:10.1016/j.egypro.2013.06.233.

Paper C:

Hvidevold, H.K., Alendal, G., Johannessen, T., Ali, A., Mannseth, T. and Avlesen, H., 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, (2015), doi:10.1016/j.ijggc.2015.03.013.

Paper D:

Hvidevold, H.K., Alendal, G., Johannessen, T. and Ali, A., Survey strategies to quantify and optimize detection probability of a CO_2 seep in a varying marine environment, *Submitted to: Environmental Modelling & Software*, april 2015.

Acknowledgements

Firstly, I would like to thank my supervisors Guttorm Alendal, Truls Johannessen and Trond Mannseth, for sharing their knowledge and guiding me during my time as a PhD student. I would especially thank Guttorm Alendal, my main supervisor, for his enthusiasm, constant support and believe in me. The thesis would not have been possible without you.

To my colleagues and friend at the department, thanks for many nice moments and discussions. I especially would like to thank "3 kaffi gjengen", for making one of the daily coffee breaks much more pleasant. And of course, the friday lunch people, the highlight of the week. I am also grateful and happy to have shared office with Huiyai, who have always been helpful and supporting, I have appreciated our many talks.

Finally, Fredrik, for always being there, you make every day a little bit better.

Contents

Part I

Background

Chapter 1

Introduction

Quote IPCC: [46]

> Warming of the climate system is unequivocal. ... Human influence has been detected in warming of the atmosphere and the ocean, in changes in the global water cycle, in reductions in snow and ice, in global mean sea level rise, and in changes in some climate extremes. ... It is extremely likely ($95-100\%$ probability) that human influence has been the dominant cause of the observed warming since the mid-20th century. Continued emissions of greenhouse gases will cause further warming and changes in all components of the climate system. Limiting climate change will require substantial and sustained reductions of greenhouse gas emissions.

1.1 Climate change and ocean acidification

In the last 50 years (from February 1965 to February 2015), the monthly mean atmospheric carbon dioxide (CO_2) levels have increased from $\sim$320 ppm to $\sim$400 ppm, Figure 1.1 [1]. The anthropogenic emission of greenhouse gases contributes to a warming of the earth, and the last three decades have been successively warmer than any of the preceding decades since 1850. The estimated globally averaged combined land and ocean surface temperature show a warming of $0.85 \pm 0.21^{\mathrm{o}}\mathrm{C}$ over the period from 1880 to 2012 [75].

The Fifth Assessment Report from IPPC, released between September 2013 and November 2014, strengthens the concluded that it is extremely likely that human influence is the main cause of global warming [75], with anthropogenic CO_2 release being the most dominant. On the timescale of centuries to millennia, the magnitude of global warming will be determined primarily by anthropogenic CO_2 emissions. This is due to CO_2 having a very long lifetime in the atmosphere. If CO_2 emission is stopped today; a scenario that is not probable, it would essentially never return to its pre-industrial level (around 280 ppm) on timescales relevant for our society.

The effect the climate changes and global warming have on the earth will influence the life on earth. Some of the effects are rise in global mean sea water level and more extreme weather events. Common to them all is that it will have an influence on human lifestyle. Although some changes may be positive it will mainly be negative effects. Rise in sea level may in some areas cause inundation that is expected to threaten vital infrastructure and human settlements. More extremes and regional changes in temperature will influence crop production and most likely increase the risk of food insecurity, especially some vulnerable groups [47].

To limit the effects of global warming a climate target is suggested, aiming to keep the global mean temperature rise below 2^{o}C above pre-industrial level. To reach this goal a quota on the total amount of CO_2 that can be emitted is given, and two thirds of this quota has already been used. The quota will most likely be exhausted within the next 30 years if emissions of CO_2 continue at the 2014 rate [35]. Studies show that to limit global warming to below 2^{o}C, more than 80% of current coal reserves, half of all gas reserves and a third of oil reserves must remain unburned through at least 2050 [59].

Recently questions about the possibility to stay below the 2^{o}C target has been raised. Even though nearly every country in the world has endorsed it, to few nations are doing enough to meet the stated target. To meet this, annual emissions of carbon dioxide and other greenhouse gasses must drop 40-70 percent by 2050. The last years the opposite has happened and on average the global emission has risen. Current pledges by governments to reduce emissions by 2020 have set the world on a path between 3 and 5 degree Celsius of warming by 2100, according to the IPCC - a potentially catastrophic level. However, if a sustainable effort is done, the IPCC report states that the goal is not out of reach.

To reduce the global warming the CO_2 emission must be substantially decreased the next decades, and be global net negative before 2100 [48], meaning that atmospheric CO_2 must be removed, hence captured and stored. If the CO_2 level is to be stabilized between 450 - 500 ppm within 2100 many model scenarios states that negative-emission technologies must be deployed within in the second half of this century [48]. A stabilization without negative-emission technologies require reduction in CO_2 emission, which essentially should start today.

Ocean acidification is also a consequence of the increased CO_2 emissions. About 27% of the anthropogenic CO_2 released into the atmosphere is slowly absorbed by the oceans [54]. Absorption of CO_2 in the ocean is causing a shift in the ocean chemistry and a lowering of the pH, essentially decreasing carbonate ions and increasing hydrogen ions. This ocean acidification and has a huge effect on biological processes. A direct consequence is that it will inhibit calcification, accumulation of calcium in body tissue, which is important for corals and shell growth. If nothing is done to significantly reduce CO_2 emissions, it may mean that there will be no place in the future oceans for many of the species and ecosystems known today [67]. The impact of ocean acidification will be globally, some areas will of course be more sensitive, and the effect will persist for thousands of years before natural processes return the pH to normal values.

The ocean and marine-system are very resistant to changes in pH, i.e. highly

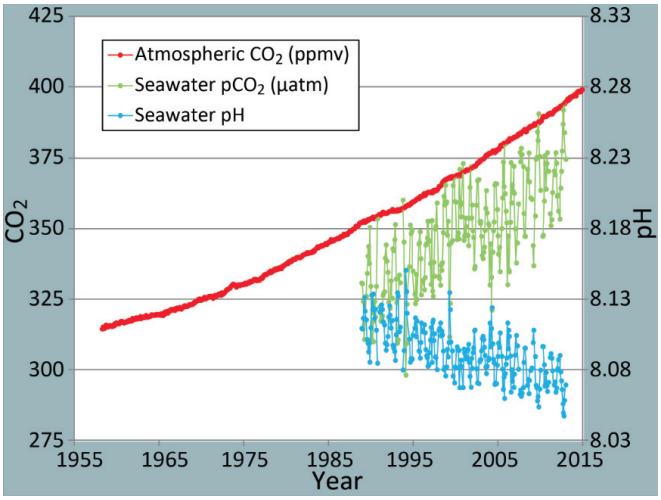


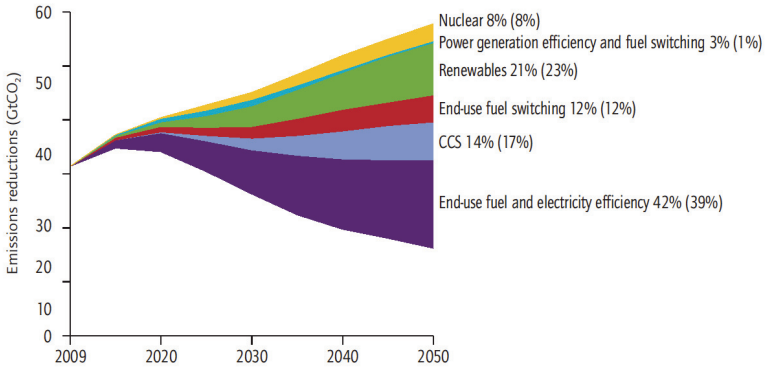
Figure 1.1: *The correlation between rising levels of CO_2 in the atmosphere at Mauna Loa with rising CO_2 levels in the nearby ocean at Station Aloha [1]. As more CO_2 accumulates in the ocean, the pH of the ocean decreases. The figure is from [1].*

buffered with respect to changes in CO_2, but only via slow millennia scale processes. The marine system has evolved with a very stable carbonate system with global mean pH between 8.0 - 8.2 over 20 million years [66]. However, the comparatively short-term perturbations as ocean acidification alter pH far beyond the ranges experienced over the evolutionary timescales of contemporary species.

Since the beginning of the Industrial Revolution, around 1750, the acidity of surface ocean waters has increased by about 30 percent, resulting in a lowering in pH of 0.1, Figure 1.1 [1]. The amount of CO_2 absorbed by the upper layer of the oceans is increasing by about 2 billion tons per year. If known fossil fuel reserves are combusted it is predicted that the global mean ocean pH will fall significantly to 7.5 or less within the timeframe of 100-300 years, depending on emission scenarios and mitigation [23]. A change that is greater than experienced in the past 300 million years, with the possible exception of those resulting from rare, catastrophic events through the history of earth [22].

Climate change and global warming is one of the greatest challenges of our time. The main contributor to global warming is anthropogenic CO_2 emissions, mainly from burning of fossil fuels. It it is inevitable that the CO_2 emissions have to be dramatically reduced to stop global warming. However, a growing population and increasing welfare has and will lead to a tremendous growth in the energy demand of the world. According to the International Energy Agency (IEA) fossil fuel will continue to be the most important source of energy for a long time. The challenge is therefore to substantially decrease the anthropogenic CO_2 emission and at same time provide the world with energy.

1.2 Mitigation options



Note: numbers in brackets are shares in 2050. For example, 14% is the share of CCS in cumulative emission reductions through 2050, and 17% is the share of CCS in emission reductions in 2050, compared with the 6DS.

Source: IEA, 2012c.

Figure 1.2: *Show how different technologies have to contribute to meet the energy sector target of cutting CO_2 emissions by more than half by 2050. CCS contributes to 14% of total emission reductions through 2050. The figure is from [44].*

There are multiple options for overcoming this challenge, such as; an increased use of renewable energy, concentrate on energy conservation and efficiency, and Carbon Capture Transportation and Storage (CCS), see Figure 1.2. It is clear that one of these actions can not do the job alone, and all of them have to contribute if we shall be able to decrease greenhouse gases and limit the global temperature.

There exist some studies explicitly limiting or eliminating CCS technologies in mitigation scenarios without ruling out the feasibility of a 2^{o}C target [9]. Then substantial reductions in CO_2 emission is required the next decades. However the ones following the 2020 emission reduction pledge by countries were found to be unfeasible or extremely costly [10, 35, 53]. Hence, CCS seems inevitable and should be explored as a potential toll to reduce further CO_2 emissions.

Chapter 2

Carbon Capture Transportation and Storage (CCS)

Carbon Capture Transportation and Storage (CCS) is a promising technology to overcome the problems with the increased anthropogenic CO_2 in the atmosphere. To be able to store the CO_2 under ground it must firstly be captured, preferably from large power plants producing/wasting huge amounts of CO_2, as cement manufactures, iron and steel making and natural gas treatment. After the CO_2 is captured it must be transported to the storage sites. The CO_2 is then injected into stable geological formation that can trap the CO_2 for hundreds and thousands of years preventing it from seeping into the atmosphere or ocean. For CCS to be a feasible and reliable option storage sites must be designed not to leak [40]. To secure this general international regulations for CO_2 storage projects exist, among them the EU directive 2009/31/EC [34]. However, more research is needed and in the case of subsea CO_2 storage the potential effect of leakage into the marine environment is addressed.

As long as fossil fuels and carbon-intensive industries play dominant roles in our economies, CCS will remain a critical greenhouse gas reduction solution. If the long-term global average temperature is to be limited, International Energy Agency IEA analysis shows that CCS is an integral part of any lowest-cost mitigation scenario [44]. To address the emissions reduction challenge the total CO_2 capture and storage rate should be around 10 gigatonnes of CO_2 per year ($GtCO_2$/yr) in 2050. As a comparison in 2013 more than 36 $GtCO_2$ were emitted.

The components of the CCS chain have been demonstrated at or close to full industrial scale [69, 80]. Most of the cost lies within the with capture part, while the storage part poses most uncertainty. There are challenges in quantification of storage potential and in the tracking verification and monitoring of injected CO_2. There is also uncertainty in the fail-safe retention of CO_2 so that a storage site can be transferred to government as a low-risk proposition for long-term care and maintenance. Hence, significant challenges exist within engineering and researching to ensure that the CO_2 is trapped.

2.1 CCS and the marine environment

If CCS is to be considered a safe and reliable option for mitigating the increasing levels of CO_2 in the atmosphere, it is important to assure that storage sites does not leak and keep CO_2 stored for decades [40]. If not the decreasing effectiveness of the power plants with CCS might lead to increased CO_2 releases, the opposite of what is the scope.

In Norway all possible storage sites except from Svalbard are offshore. The London Convention and the Oslo-Paris Convention (OSPAR for the North Sea) will apply for these sites. These guidelines include six stages of risk and impact assessment to execute before, during and after injection of CO_2. This includes, among others, an impact assessment in the marine environment, i.e. monitoring to detect a potential leak into shallow sediments, seafloor and the water column. In this thesis a leakage is used about CO_2 escaping from a storage site and into the marine waters. The EU Emission Trading Scheme (EU ETS) treats leakage from geological storage as emissions, and it needs to be quantified [45].

Government and international regulations will for any CO_2 storage project require risk and environmental status assessment to sustain future monitoring programs. These assessments will have to include potential impact on marine environment, and will have to outline a subsequent monitoring program for detection, localization and quantification of a leak. This is a challenging task to perform in marine waters since it will be close to impossible to prove that an identified increase in CO_2 concentration stems from a particular storage site, unless the injected CO_2 is marked with an unique tracer.

Understanding the potential for CO_2 leaked from CCS and the impact it may have on the marine environment is a complex problem. It requires an understanding of physical dispersion, the behavior of plumes, marine chemistry and ecological relationships. One of the key issues is understanding the spread, persistence and impact of a CCS derived leak. All storage sites will be designed not to leak, however a probability of leakage will always be present, and if leakage should occur it is important to be convinced that its impact is less than the long term consequences of climate change and the uptake of anthropogenic created atmospheric CO_2 [56].

Simulation of different leakage scenarios imply that an alternation in pH due to leakage from geological stored CO_2 will only have local effects [3, 16]. Nevertheless the local acidification may have impact on marine biota in the area [23]. In the vicinity of the leak the pH may decrease lower than ~ 7.5. However, the pH will return to normal within a relatively short timeframe, provided that the leak is temporary. Therefore, a potential leakage must be detected as soon as possible, and to avoid fatal consequences it should be stopped. To be able to detect leakages, knowledge of how CO_2 is distributed and absorbed in the ocean and how this affects the marine environment is vital.

Broadly speaking there are two potential leak sources connection with CCS; 1) leakage during transportation from pipelines, and 2) leakage directly from the geological storage site, see Figure 2.1. A leakage from the storage site will either

be a plume or individual rising droplets or bubbles, depending on the flux. In this process models describing rise and dissolution of CO_2 droplets/bubbles play an important role [4, 28].

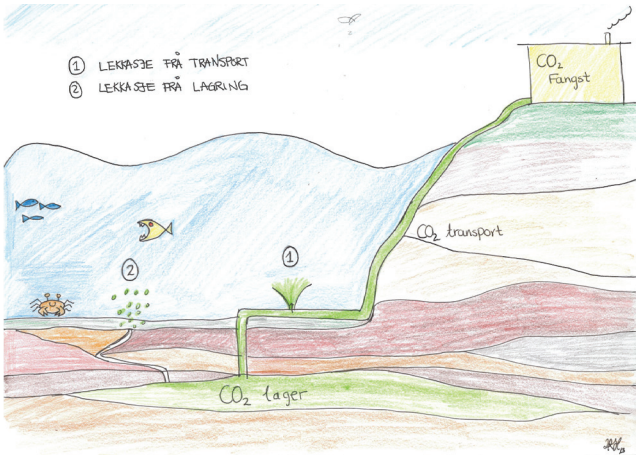


Figure 2.1: *Possible leakage scenarios in connection with CCS; 1) leakage from transportation, or 2) leakage from the storage site.*

An important part of the imposed requirements for a storage project is an adequate design of monitoring program. The program will have three main objectives; i) assure that a potential leak will be early detected, ii) if no leak is detected the program should contribute to build an accurate base line to capture changes in trends caused by natural or human activities, and iii) to assure that other changes in CO_2 concentration than those caused by CCS is added on the CCS account to prevent unjustifiable accusations about leaks and environmental damages [18].

2.2 Monitoring of CO_2 storage

CO_2 injected into aquifers, or other geological formations, will be buoyant and hence try to rise. A number of different trapping mechanism will keep the injected CO_2 contained in the formation selected. Although several trapping mechanisms exist and will hinder unexpected transportations of CO_2 upwards in the formation with the potential of eventually leaking into the water column, leakage cannot be ruled out as a possibility. The transport of the CO_2 within the formation might cause pathways to the surface to become probable, or the CO_2 might create new pathways [64]. Such pathways may be wells, faults or fractures that perforate the storage formation, or is reached if the CO_2 rises through to other shallower formations. This might lead to very localized leaks, with possible high flux rates, far away from the injection site. More diffusive leaks may be caused by CO_2 migrating through the overburden reaching the surface over a relatively large area.

Even if the formation is being monitored through seismic and other remote methods, and the bulk part of the CO_2 plume within the formation is followed from year to year, there will still be uncertainties regarding how much of the CO_2 is being contained in the plume. In other words, what is the detection limit of CO_2 migrating toward the surface, and are we looking at the right place for such pathways? Numerical simulation tools will be good tools in this respect, but even these will be inaccurate and have uncertainties [62].

In connection with monitoring and quantification in marine waters there are many challenges to overcome. Knowledge on how CO_2 is distributed and absorbed in the ocean and how this affects the marine environment is vital. Understanding near-field diffusion and dispersion processes are important to assess effects on the marine biota in the vicinity of a seep. In addition, the transport on larger spatial and temporal scales, including the rate of outgassing to the atmosphere and the buffering capacity offered by the ocean, are essential to be able to design a monitoring program that covers large areas.

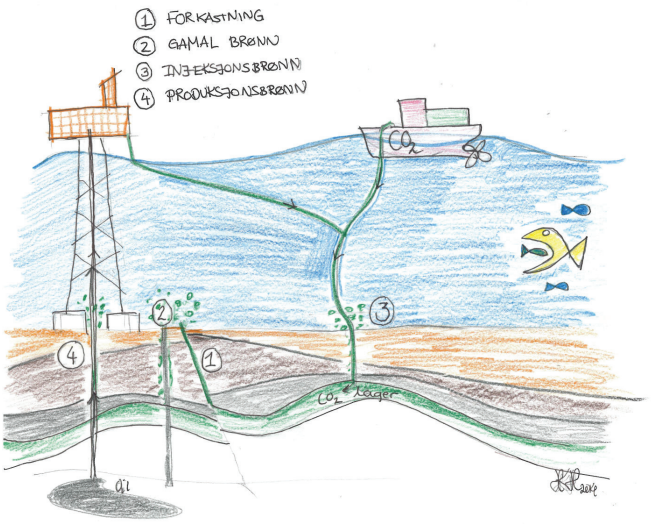


Figure 2.2: *Possible leakage pathways from the storage site to seabed; (1) through fractions or faults, old wells (2), injections wells (3), along production wells (4) or a more diluted leakage through the sediments.*

There exist several possible leakage pathways from storage site to the seafloor; through wells, fractures and faults in addition to maybe more diffuse leakage through the sediments, see Figure 2.2. In general the leakage will either be diluted with a limited mass flux, for instance through sediments, or more dense with a high mass flux. Failure in injection wells or transportation pipes will most likely result in the latter case. Local current conditions and the small scale mixing will

determine where a signal of a CO_2 leak is transported, and how quickly it disperses and become indistinct from the natural variability. The mixing rate changes with atmospheric forcing, tidal currents and local topography.

Studies indicates that CO_2 bubbles dissolves in the bottom layer and in a small area around the leak source, and strong currents and tidal cycles will dilute the solution [3]. This indicated that a leakage from a point source will have limited localized environmental impact. More studies are needed in the case of a more diffuse leakage scenario, when number of spatial distributed leaks interacts.

Acidification of marine waters caused by leakage of CO_2 is one trend that is important to capture when monitoring the marine environment. In all cases an increase in CO_2 concentration is expected in the vicinity of a leak, with subsequent possible environmental impacts [16]. Environmental changes might serve as indicators that a leak is occurring, either through changes in bottom fauna, such as new occurrences of bacterial mats [81], or changes in the pelagic ecosystem [61].

Several monitoring equipment and methods already exists and are being developed [5, 52, 73]. Sidescan Sonar Bathymery and Seabed Multibeam Bathymery are effective methods to cover a large area in relatively short time. Bubble stream detection using remotely operated vehicles (ROV) or autonomous underwater vehicles (AUV) are also possible. If there is a need to continuously monitor the marine environment, fixed installation can be placed on the seafloor or in the water column. Acoustic methods can be used to detect bubble trains, plumes and quantify fluxes.

To design a monitoring program is highly dependent on having a proper understanding of the environment prior to the injection start. A baseline database of important environmental parameters is an absolute requirement. Which data and parameters to collect will be highly site specific, and the data should be gathered over years. Keep in mind that the area to be studied will be much larger than other offshore activities and the time frame of study is much longer. It is therefore important that long time series are gathered and that the monitoring is designed to keep adding to the environmental baseline in order to incorporate trends caused by other activities or by natural variations.

2.3 Contribution

The work done in this thesis contributes in the marine monitoring program in the CCS research field. It is motivated from the desire to contribute in designing reliable and safe storage sites, and to assure that a potential leak from geological CO_2 will be detected. The papers use the same mathematical tools, uncertainty assessment using linearized covariance analysis and optimization using the genetic algorithm. The overall subject is quantification of uncertainties in models that are all linked to CCS and monitoring, either through the mathematical equations describing a sub model in simulation of leakage or in quantification of the probability to detect a CO_2 leakage on the seafloor.

Only with a proper understanding of the processes involved during a seep will

an estimate of the spatial and temporal footprint of a potential leak be reliable. Such a footprint is essential for subsequent impact assessments and for designing a monitoring program for the marine environment. In a hypothetical leak scenario, CO_2 bubbles/droplets will ascent through the water column and gradually dissolve, acidifying the water. The rise and dissolution of CO_2 in the water column is important sub models in the complex system simulating advection, diffusion and mixing of CO_2 in the water column.

In Paper A and B experimental design is used to assess the parameter uncertainties in the rise velocity for a CO_2 droplet. Since the parameter uncertainty can be estimated prior to the experiment a suggestion on how to design an experiment aiming to minimize the uncertainties is also given. How these uncertainties effects in the rise and dissolution of a CO_2 droplet in the water column is also studied. The coupled differential equation describing rise and dissolution is a sub-model in a plume model. The simulations from a bubble plume model [3] is used as a basis for describing the leakage distribution around a leakage point in the Papers C and D.

Seafloor depth above possible leakage sites in the North Sea is shallower than 500 m, hence a CO_2 leakage will in general be in the bubble regime. Hence, the droplet equations studied in the first two papers are not directly useable for this region. However the methods used in the papers can also be used for CO_2 bubble equations, and experimental design should be included in the process of parameter estimation.

Paper C and D demonstrate how environmental statistics may assist in designing a monitoring program. They give a quantification of the probability to detect a leakage in a variable marine environment. Optimal strategies for placement of sensors maximizing the probability of detecting a leak are provided, both at fixed installations and during surveys. As a basis a synthetic map is used to predict a leakage scenario, where faults and wells have different probability of leakage. The leakage footprint is simulated under varying current conditions, where a bubble plume model is used as the source [29] in a general circulation model.

Chapter 3

Modeling CO_2 particles in marine waters

The marine waters are in constant motion and are characterized with high spatial and temporal variability. The tides change current directions, wind alters the amount of mixing, and local topography might change the local current conditions [2]. Seasonal variations might also be significant. Hence, leakage footprints are varying signals depending on the local oceanic and atmospheric conditions.

Dispersion of CO_2 in seawater is a complex process. Observations form experiments and natural leakage show that CO_2 will most likely enter the water as either bubble (gaseous phase) or droplets (liquid phase) of varying size [20], dependent on the leakage pathway. CO_2 is buoyant and soluble in seawater above 3000 m, if between $\sim$ 3000-500 m the CO_2 will be in liquid phase if shallower than $\sim$ 500m the CO_2 will in gaseous phase [4]. The collective term for a bubble or a droplet is a particle.

Depending on the release depth, temperature of surroundings and flux the CO_2 will either completely solve in seawater or rise all the way up to the surface and get released to the atmosphere. If the temperature of the surrounding seawater is below 9^oC, hydrates may form on the particle surface, changing the CO_2 seawater dynamic, essentially slow down the dissolution of CO_2 [8, 60], see Figure 3.1 [14]. A plume is emerged if the number of CO_2 particles are high. Through interfacial drag and mass transfer there will be a two way coupling between the particles and the surrounding water [4, 28].

One of the key features in all of this is the speed of which CO_2 particles ascend through the water column, and how fast mass transfer dissolves the CO_2 content into the water masses. These two processes are interlinked; a fast moving particles will dissolve faster. An interface separates the particle, dispersed phase, from the surrounding medium, the continuous phase. Coupling through mass, momentum and energy transfer occurs in the interaction between the particle and the surrounding medium. In the following section the concept of momentum and mass will be addressed to define the numerical model describing the rise and dissolution of a single CO_2 particle in seawater. Most of the basic theory presented in the next sections is taken from the textbooks [25, 27].

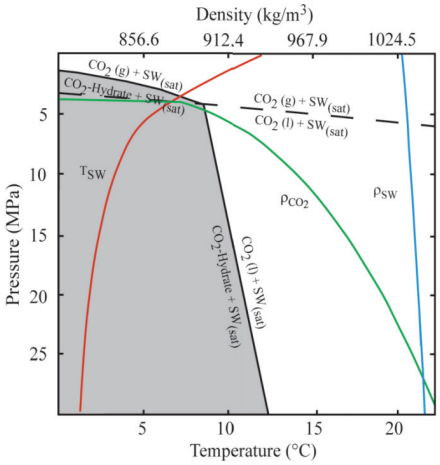


Figure 3.1: *Phase Diagram for CO_2 in the ocean, Figure is from [14]. Gray area: region of stability for CO_2 hydrates; dashed line - gas/liquid phase transition for pure CO_2; red line - depth -temperature profile of the Pacific Ocean at 152^{o}W, 40^{o}N; green line - CO_2 density; blue line - seawater density (35.0 PSU).*

3.1 Rise velocity

Newton's second law, stating that the net force acting on a system is equal to the rate of chance of momentum in the system, governs the motion of particles in fluids. The momentum equation for an incompressible fluid particle is

$$m_d \frac{d\boldsymbol{u}}{dt} = (m_s - m_d)\boldsymbol{g} - \boldsymbol{F}_d. \tag{3.1}$$

Here m_d and $\boldsymbol{u}$ is the mass and the velocity of the particle respectively, t is the time. The first term on the right hand side is the buoyancy forces, where m_s is the mass of the displaced surrounding medium and $\boldsymbol{g}$ the gravitational acceleration. $\boldsymbol{F}_d$ is the drag force of the particle and represent the surface force, acting on the surface of the particle.

Fluid particles can have a range of shapes under influence of external fluid fields, and the shape effects both mass transfer and motion. Moving in free rise or fall under influence of gravity, fluid particles are generally grouped under three shapes: the spherical, the ellipsoidal and the spherical-cap or ellipsoidal-cap regime [25].

In this thesis CO_2 in seawater is studied, hence the surrounding medium is seawater and particle medium is CO_2. The rise velocity of the CO_2 particle is derived from the spherical equivalent to equation (3.1), given by

$$\frac{4}{3}\pi r^3 \rho_d \frac{d\boldsymbol{u}}{dt} = \frac{4}{3}\pi \boldsymbol{g} r^3 (\rho_s - \rho_d) - \boldsymbol{F}_d.$$

Here ρ_d and ρ_s is the density of the particle and the seawater respectively, and r is the radius. The drag force acts opposite to the direction of the movement of the particle and represents a resistance that the particle experience due to the presence of the continuous phase. A quadratic law is assumed for the drag force,

$$\boldsymbol{F}_d = C_d \frac{1}{2} \rho_s |\boldsymbol{u} - \boldsymbol{v}| (\boldsymbol{u} - \boldsymbol{v}) \pi r^2,$$

where C_d is a dimension less drag coefficient, and will be discussed later, and $\boldsymbol{v}$ the velocity of the medium surrounding the particle. If there is no acceleration of the relative velocity between the particle and the surrounding medium a balance of buoyant and friction forces is obtained

$$\frac{4}{3} \pi g r^3 (\rho_s - \rho_d) = C_d \frac{1}{2} \rho_s u_T^2 \pi r^2.$$

Then the vertical terminal velocity u_T of a particle rising in sea water is given by ($u_T = \|\boldsymbol{u} - \boldsymbol{v}\|$)

$$u_T = \sqrt{\frac{8gr(\rho_s - \rho_d)}{3C_d \rho_s}}. \tag{3.2}$$

The particles studied here are in general not spherical, and the radius, r, is replaced by the equivalent radius, r_e,

$$r_e = (a^2 b)^{1/3}.$$

The equivalent radius is the radius of a sphere having the equivalent volume as the particles studied [25]. In the equation above a and b are major and minor semi axes of the imagined particle respectively. Spheroids can be approximated by combining two semi-spheroids of different height and same major axis, such that $b = (b_1 + b_2)/2$, see Figure 3.2.

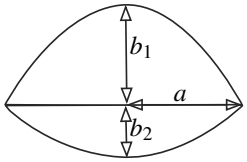


Figure 3.2: *Idealized shape of a rising particle, two semi-spheroids of different height and same major axis.*

The drag coefficient

The drag depends on hydrodynamic characteristics of the flow, on particle shape and orientation with respect to flow, on properties of the particles, the shape of the surface past which the fluid moves and on the properties of the surrounding

medium. Numerous expressions for the drag coefficient exist, they are in general complex and expressed though dimensionless numbers such as the Reynolds number, Re, the Weber number, We, and the Eötvös number, Eo, given by

$$\mathrm{Eo} = 4gr_e^2(\rho_s - \rho_d)\sigma^{-1} \quad \mathrm{Re} = 2r_e\rho_s u_T \mu^{-1} \quad \mathrm{We} = 2r_e\rho_s u_T \sigma^{-1}$$

where μ is the dynamic viscosity and σ is the interfacial tension.

There are very limited experimental and in-situ measurements of CO_2 seeping from the seafloor and behavior of CO_2 in seawater for verification of numerical models. However a few in-situ experiments by MBARI [21, 36] and laboratory experiments [14, 15] have led to various expressions for the drag coefficient. Standard Rayleigh curves for a single sphere moving at constant velocity are derived from experimental studies. For deformable particles, changing in size and shape, theoretically determining the drag coefficient involves serious difficulties, and is in most cases estimated using experimental data and empirical correlations [50]. In [50] a review over the drag coefficient of various shaped particles is given, stating that the choice of drag coefficient is, among other, dependent on flow pattern of the dispersion, values of the Re, Mo and We and particle size and shape.

The drag coefficient can be split into the friction factor, $f_{c,h}$, and the deformation factor, D:

$$C_d = f_{c,h}D, \tag{3.3}$$

where the deformation factor, $D = (a/r_e)^2$, account for non sphericity of the particle. The friction factor is dependent on whether the particle is covered with a hydrate skin or not. To distinguish between the two situations the subscript c will be used for hydrate free, and h for hydrate coated particles.

In Paper A and B we study the rise velocity for liquid CO_2 droplets with and without hydrate skin. The studies use expressions from [15], where the authors performed parameter estimation on experimental data from [14] to derive the expression. Under the assumption that bubbles attain shapes that minimize their energies, the deformation factor can be expressed as

$$\begin{aligned} D =& [2/(3.974\times10^{-3}(\mathrm{We}-12.62)^2 - 7.186\times10^{-4}(\mathrm{Eo}-17.87)^2 \\ &+3.280\times10^{-5}\mathrm{EoWe}((\mathrm{Eo}-27.77)(\mathrm{We}-8.405)+67.08)+1.130)]^2. \end{aligned} \tag{3.4}$$

To estimate the friction factor, known asymptotical behaviors are used as first approximation, and subsequent adjustments are added using curve fitting. The experiment was conducted in a stainless steel tank where the temperature and pressure were held constant during an experiment. Droplets of different shapes and sizes were released and cameras were recording shape, size and velocity for each droplet, for more details about the experiment see [14] and reference therein.

The friction factor, f_c, will in general vary with the physical properties of the droplet studied [14, 19, 36]. For particles without hydrate skin [14] calculated the friction factor to be

$$f_h = \frac{9}{\mathrm{Re}} + 0.9\frac{0.75\mathrm{Eo}^2}{0.75\mathrm{Eo}^2+4.5} = \frac{9}{\mathrm{Re}} + \frac{\alpha_h\mathrm{Eo}^2}{\gamma_h\mathrm{Eo}^2+\omega_h}, \tag{3.5}$$

for hydrate coated droplets. While the friction factor for hydrate free droplets is

$$f_c = \frac{48}{\mathrm{Re}} + 0.9\frac{0.75\mathrm{Eo}^2 + 0.6}{0.75\mathrm{Eo}^2 + 14.5} = \frac{48}{\mathrm{Re}} + \frac{\alpha_c \mathrm{Eo}^2 + \beta_c}{\gamma_c \mathrm{Eo}^2 + \omega_c}. \tag{3.6}$$

Here $\alpha_{h,c}, \beta_c, \gamma_{h,c}$ and $\omega_{h,c}$ are parameters decided using curve fitting, which is a well known method for finding unknown parameters in an expression.

In Paper A the uncertainty in the parameters in equation (3.6) are investigated and a suggestion on how to optimize the experiment aiming to reduce the uncertainties are presented. To make the study of the terminal velocity possible, the parameters in the friction factor are reduced to three linear independent parameters, $\mathbf{\Theta} = [\Theta_1\ \Theta_2\ \Theta_3]$, giving

$$f_c = \frac{48}{\mathrm{Re}} + \Theta_1 \frac{(\mathrm{Eo}^2 + \Theta_2)}{(\mathrm{Eo}^2 + \Theta_3)}.$$

The rise velocity is seen on as a function of the equivalent radius and the parameters, while the other variables are assumed to be constant. Both the friction factor $f_c{=}f_c(u_T, \Theta_i)$, equation (3.6), and the deformation factor $D{=}D(u_T)$, equation (3.4), are dependent on the velocity, through the Reynolds number and the Weber number respectively. Solving equation (3.2) gives a polynomial $P(u_T){=}0$ of order eight and the solution is

$$u_T = u_T(r_e, \mathbf{\Theta}),$$

where u_T must be solved numerically.

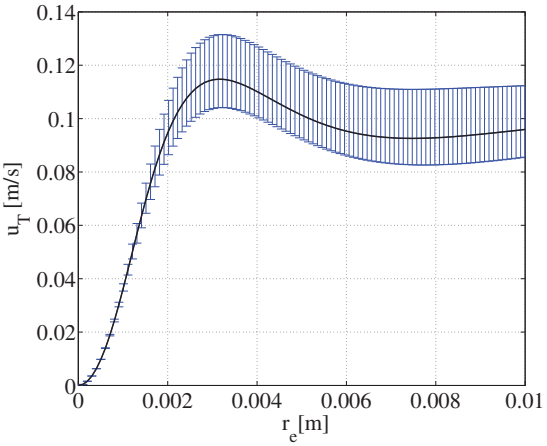


Figure 3.3: *Show the rise velocity, black line, for a hydrate free CO_2 droplet as function of the equivalent radius using the parameter values found in [14],* P=22 MPa, T=13.9$^{\mathrm{o}}$C. *The blue line shows how the parameter uncertainties effect the rise velocity.*

In Bigalke et. al [14] they performed three experiments for hydrate free CO_2 droplets for three different pressure and temperature (P/T) conditions. Figure 3.3 shows the rise velocity, black line, for one of these P/T conditions as function of the equivalent radius. Note that, in the beginning the velocity increases together with increasing radius, when the droplet gets bigger the velocity decreases due to friction before it gradually increases again. The uncertainty in the rise velocity due to parameter uncertainties is also shown in the Figure 3.3, blue lines, the true parameter values lines with 95% probability within this area.

3.2 Dissolution of CO_2 in the water column

The continuity equation for a particle states that the rate of change of particle mass is the negative value of the mass flux through the particle surface,

$$\frac{dm_d}{dt} = -\rho_s w S, \tag{3.7}$$

where m_d is the mass of the particle, ρ_s is the density of the continuos phase (seawater) at the particle surface S and w is the velocity through the particle surface (uniform and positive if mass is transferred out of the particle). The mass flux, $\rho_s w$, through the particle surface is due to evaporation or condensation, hence the transport of particle vapor to or from the particle surface and the change of phase on the surface. This process is driven by the difference in fluid mass concentration between surface and ambient fluid.

Fick's law gives the mass flux at the surface,

$$\rho_s w = -D_v \frac{\partial \rho_d}{\partial n_n},$$

where n_n is the coordinate normal to the surface in the direction out from the particle, D_v is the diffusion coefficient and ρ_d the particle density. Based on the assumption that the mass fraction of the particle is much less than unity the equation can be expressed as

$$\rho_s w = -\rho_s D_v \frac{\partial \rho_d/\rho_s}{\partial n_n} = \rho_s D_v \frac{\partial \omega_d}{\partial n_n},$$

where ω_d is the mass fraction of the dispersed phase. The diffusion coefficient is scalar in water

$$D_v = \frac{7.1141 \cdot 10^{-15}(273.15 + \mathrm{T})}{\mu_c},$$

where T is the temperature of the fluid in the continuous phase, and the dynamic viscosity for seawater at 20°C and salinity S=36 psu is $\mu_c = \mu_{SW} = 1.075 \cdot 10^{-3}$ kgm/s.

For a particle of radius r the gradient in mass fraction must be proportional to the mass fraction between the surface and the free stream and inverse proportional to the particle diameter,

$$\rho_s w \sim \rho_c D_v \frac{\omega_{d,s} - \omega_{d,\infty}}{2r},$$

where the subscripts s and ∞ denote at the surface and in the free stream respectively. The density ρ_c is the representative density and can be the average density between the surface and the free stream, but in most cases it should be considered as the density in the free stream (continuous phase).

From the continuity equation (3.7), the rate of change of a spherical particle is proportional to,

$$\frac{dm}{dt} \sim \pi 4 r^2 D_v \frac{\omega_{d,s} - \omega_{d,\infty}}{2r} = 2\pi r D_v (\omega_{d,s} - \omega_{d,\infty}).$$

The Sheerwood number is the proportionality constant and

$$\frac{dm}{dt} = 2\mathrm{Sh}\pi r D_v (\omega_{d,s} - \omega_{d,\infty}). \tag{3.8}$$

As for the friction factor and deformation factor in the rise velocity, the Sherwood number in the mass transfer equation should in general be estimated from experiments for each specific case.

The Sherwood number

The Sherwood number is found by a correlation that represents the effect of relative velocity between the particle and the surrounding medium, this effect is to increase the rate of mass transfer. The correlation is often referred to as the Ranz-Marshall correlation and is formulated as [27],

$$\mathrm{Sh} = 2 + \alpha \mathrm{Re}^{1/2} \mathrm{Sc}^{1/3}.$$

Sc is the Schmidth number, $\mathrm{Sc} = \nu_c / D_v$, where $\nu_c = \mu_c / \rho_c$ is the kinematic viscosity and Re is the Reynolds number. The value of the factor α in the Ranz-Marshall equation is dependent on which regime the Reynolds number is in

- $\alpha = 0.60$ if $2 < \mathrm{Re} < 200$, this is the standard factor for most textbooks.
- If $30 < \mathrm{Re} < 2000$, then a better value is $\alpha = 0.69$ [27].
- If $20 < \mathrm{Re} < 2000$, then $\alpha = 0.79$ for mass transfer to water [68] .

In paper B,

$$\mathrm{Sh}_c = 2 + 0.96 \mathrm{Re}_r^{1/2} \mathrm{Sc}^{1/3}, \tag{3.9}$$

is used as the Sherwood number for a hydrate free droplet, see black dashed line in Figure 3.4.

For hydrate covered particles, the Sherwood number can be found through various correlations or can be derived from experimental data. There exist a limited amount of experimental data for mass transfer of CO_2 particle in seawater. From those experiments, e.g. [21, 63], different models have been derived and suggested, e.g. [36, 42, 60, 65]. Since a direct comparison is difficult, especially in the presence of hydrate, the model suggested by Chen et al. [24] is used in Paper B, together with their expression for the solubility of CO_2. Hence, the Sherwood number for hydrate coated droplets is

$$\mathrm{Sh}_h = 2 + 0.96\mathrm{Re}_r^{1/2}\mathrm{Sc}^{1/3} \\ (1 + 4.6707 \times 10^{-4}\mathrm{Re} - 1.1871 \times 10^{-6}\mathrm{Re}^2 + 1.4766 \times 10^{9}\mathrm{Re}^4), \quad (3.10)$$

where the last term is the ration of total effective droplet area to total area of an equivalent sphere, see blue line in Figure 3.4. Solubility of CO_2 droplets decreases with decreasing pressure and due to hydrate coating it increases with increasing temperature, in contrast to hydrate free CO_2 droplets [8].

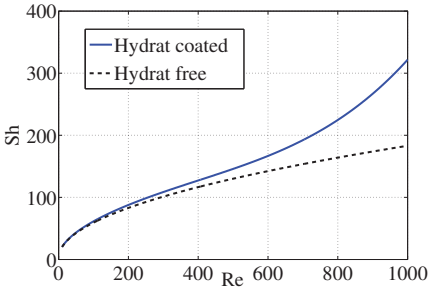


Figure 3.4: *Show the Sherwood number as a function of the Reynolds number for the hydrate free droplet, black dashed lie (equation (3.9)), and hydrate coated droplet, blue solid line (equation(3.10)).*

3.3 Rise and dissolution of a single CO_2 droplet in seawater

Given the equation for the rise velocity, equation (3.2), and the mass transfer, equation (3.8), the numerical model for the rise and dissolution of a CO_2 single droplet in seawater can be described by the coupled differential equation

$$\frac{dz}{dt} = u_T = \sqrt{\frac{8gr_e(\rho_s - \rho_d)}{3D(u_T)f_{c,h}(u_T)\rho_s}}, \quad (3.11)$$

$$\frac{dm}{dt} = \dot{M} = 2\mathrm{Sh}_{c,h}(u_T)\pi r D_v(\omega_{d,s} - \omega_{d,\infty}), \quad (3.12)$$

where it is assumed that it is no horizontal current, and that the concentration in the ambient water $\omega_{d,\infty} = 0$. The drag coefficient is defined in equation (3.4), the friction factor and Sherwood number for hydrate coated and hydrate free droplets is given in equations (3.5), (3.10), (3.6) and (3.9) respectively.

A theoretical example of how a CO_2 droplet dissolves in the water column is shown in Figure 3.5. The ODE's are integrated with a 4-5 order Runge-Kutta method. Seawater and droplet densities are calculated from Gibbs seawater (GSW) Oceanic toolbox [58] and a 32 term Modified Benedict-Webb-Rubin equation of state [77] respectively. The salinity S=34.8 psu, and temperature T=2^{o}C. The droplets are released at 800 meter and the initial droplet diameter is 1 cm, when the droplet diameter is 1 mm the simulation stops and it is assumed that the remaining mass is dissolved at the last depth. The droplets surface concentration, $C_s = 1363.33 \times 44.01 \text{g/m}^3$, is constant [78]. Note that hydrate free CO_2 droplets do not exist in seawater for temperatures lower than 10^{o}C, it is included in Figure 3.5 as a graphical illustration. Hence the simulations are strictly theoretical.

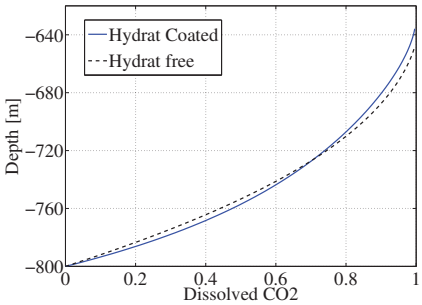


Figure 3.5: *Dissolved CO_2 for a single CO_2 droplet with, blue solid line, and without, black dashed line, hydrate coating. The release depth is 800 m and the seawater temperature is set to 2^{o}C, the initial droplet diameter is 1 cm.*

The simple model above follow the motion of individual particle in the ocean, described in Lagrangian coordinates. When the number of particles is high the modeling becomes challenging, and the approach is only practical for a small amount of particles. Therefore, one must turn to field description, Eulerian coordinates, to model the volume fraction occupied by the particle within a control volume. Then there is a need to account for the two way coupling and the dual interaction between particles and seawater. A momentum source term to account for interfacial drag must be included in the standard hydro mechanical equations . In addition, the increased level of CO_2 must be incorporated through an equation of state, $\rho_{SW} = \rho(S,T,C_{d,\infty})$ [4, 24, 31, 32].

Chapter 4

Tools and Methodology

In this chapter, the mathematical theory used in the papers in the thesis are explained. Firstly the theory behind experimental design through linearized covariance analysis is introduced. Thereafter the optimization method, the genetic algorithm is briefly described.

4.1 Assessing parameter uncertainties by linearized covariance analysis

An important part of a parameter estimation problem is to determine the uncertainty in the parameters, by doing that one can state a region for which the parameters are likely to be. This region is dependent on uncertainties in the measurements and the experimental setup. Preferably the uncertainty region should be as small as possible, this is achieved by optimizing the experiment, hence designing a best way to conduct the experiment. Having knowledge about the parameters and the profile of the model in question this optimization can be done prior to the experiment using theory from nonlinear regression, the linearized covariance analysis (LCA) method. Most of the theory presented in the next sections is taken from the textbooks [12, 72].

4.1.1 Parameter estimation

The basic principle behind parameter estimation is as follow: given a set of data, often obtained from measurements, and a function dependent on a set of unknown parameters, then the aim is to estimate the parameter values such that the data matches the function. It is grounded on the assumption that there exist a true parameter set $\hat{\boldsymbol{\Theta}}$ and a physical correct model f such that the data y_j, from for instance an experiment, may be expressed as

$$y_j = f(\boldsymbol{x}_j, \hat{\boldsymbol{\Theta}}) + \varepsilon_j \qquad j = 1, \ldots, n. \tag{4.1}$$

Here ε_j are the measurement errors of case j which are assumed to be independent and normally distributed with zero mean, variance ϑ^2 and covariance matrix $\boldsymbol{\Sigma} =$

$\boldsymbol{\vartheta}^2\boldsymbol{I}$. The vector $\boldsymbol{x}_j$ consist of (measurable) variables for the jth case, and n is the number of measurements. When analyzing a particular set of data the vector $\boldsymbol{x}_j$ is considered as fixed, and the focus is on the dependence of the model f on $\boldsymbol{\Theta}$. By creating the n-vectors $\boldsymbol{f}(\boldsymbol{\Theta}) = [f(\boldsymbol{x}_1,\boldsymbol{\Theta}), f(\boldsymbol{x}_2,\boldsymbol{\Theta}),\ldots,f(\boldsymbol{x}_n,\boldsymbol{\Theta})]^T$ and $\boldsymbol{Y} = [y_1,,y_2,\ldots,y_n]$ the measurements of the nonlinear model studied can be write as

$$\boldsymbol{Y} = \boldsymbol{f}(\hat{\boldsymbol{\Theta}}) + \boldsymbol{\varepsilon},$$

where $\hat{\boldsymbol{\Theta}}$ is the true parameter vector. Because of the measurement errors the matching of $\boldsymbol{Y}$ to $\boldsymbol{f}$ can never be exact, the aim is to find the set of parameters $\boldsymbol{\Theta}^*$ so that the matching is satisfying. The most commonly used method to find this set of parameters is the least square method which provides the estimated parameters, $\boldsymbol{\Theta}^*$, that minimizes the objective function [72]

$$S(\boldsymbol{\Theta}) = (\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\Theta}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\Theta})), \tag{4.2}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix for the measurement errors. Solving scalar nonlinear optimization problems is a widely discussed and researched topic, several methods with citations are discussed in detail in for instance Seber and Wild [72] and Aster et al. [6]. Least square optimization methods exploit the structure of sums of squares and Newton's method, due to its speed, serve as basis in many methods in nonlinear optimization. Both Gauss-Newton method and the Levenberg-Marquradt method [55, 57] are example of such methods.

It should be mentioned that nonlinear least square problem can have difficulties with ill-posed problems, and the problem typically grows with increasing number of model parameters [6]. Essentially ill-posed means that a small change in the input value leads to an enormous change in the estimated model, this problem can be overcome by including a regularization term, $R(\boldsymbol{\Theta})$, in the least square minimization. Hence solving

$$S(\boldsymbol{\Theta}) = (\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\Theta}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\Theta})) + R(\boldsymbol{\Theta}),$$

instead of equation (4.2). One of the most used regularization techniques is the Tikhonove regularization [79]. Due to the low number of parameters in the problems discussed in the papers it is assume that $R(\boldsymbol{\Theta}) = 0$.

4.1.2 Confidence region

Due to (measurement) errors $\boldsymbol{\Theta}^* \neq \hat{\boldsymbol{\Theta}}$, therefor it is important to specify a region in parameter space, called a $100(1-\alpha)\%$ confidence region, which with a probability of $100(1-\alpha)\%$ contains the true parameters. The confidence region is found by examine the shape of the objective function $S(\boldsymbol{\Theta})$, equation (4.2). The estimated parameters that minimize equation (4.2) is $\boldsymbol{\Theta}^*$ such that $\partial S/\partial\boldsymbol{\Theta} = 0$. If the model is linear, an exact confidence region is given by

$$(\boldsymbol{\Theta}^* - \hat{\boldsymbol{\Theta}})\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{\Theta}^* - \hat{\boldsymbol{\Theta}}) \leq \chi^2_{p,\alpha},$$

where $\boldsymbol{X}$ is the coefficient matrix of the linear model, i.e. $\boldsymbol{f}(\boldsymbol{\Theta}) = \boldsymbol{X}^T\boldsymbol{\Theta}$ and $\chi^2_{p,\alpha}$ is the upper α quantile for the χ^2 distribution with p degrees of freedom. When the model is nonlinear the problem is more complex. However if the nonlinear model can be well approximated by a linear function for the set of parameters contained in the confidence region a similar inequality as above can be found where $\boldsymbol{X}^T\boldsymbol{X}$ is replaced by an approximated covariance matrix for the estimated parameters, $\boldsymbol{P} = E[(\boldsymbol{\Theta}^* - \hat{\boldsymbol{\Theta}})(\boldsymbol{\Theta}^* - \hat{\boldsymbol{\Theta}})^T]$. It is a basis for linear covariance analysis; a derivation of $\boldsymbol{P}$ and a detailed description on LCA can be found in [7, 39], and is also included the following section.

Derivation of the linear covariance matrix

Assuming that the true parameters exist $\hat{\boldsymbol{\Theta}}$ such that the data $\boldsymbol{Y}$ is given by the nonlinear model $\boldsymbol{f}(\hat{\boldsymbol{\Theta}})$ in the following way

$$\boldsymbol{Y} = \boldsymbol{f}(\hat{\boldsymbol{\Theta}}) + \boldsymbol{\varepsilon}.$$

Since the errors, $\boldsymbol{\varepsilon}$, are assumed to be normally distributed with covariance matrix $\Sigma = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]$, the data vector has the probability distribution

$$p(\boldsymbol{\varepsilon}) = \frac{1}{(2\pi)^{n/2}\|\boldsymbol{\Sigma}\|^{1/2}} e^{-\frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{\Sigma}^T\boldsymbol{\varepsilon}}.$$

By minimizing the objective function

$$S(\boldsymbol{\Theta}) = (\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\Theta}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{f}(\boldsymbol{\Theta})),$$

with respect to $\boldsymbol{\Theta}$, the estimated parameters $\boldsymbol{\Theta}^*$ is found. Since the true parameter vector is fixed the minimization problem can be minimized with respect to the estimation error $\boldsymbol{\kappa} = \boldsymbol{\Theta}^* - \hat{\boldsymbol{\Theta}}$.

Linear covariance analysis is based on the assumption that the model function $\boldsymbol{f}(\boldsymbol{\Theta})$ can be approximated by a linear function in the vicinity of $\hat{\boldsymbol{\Theta}}$. Hence

$$\boldsymbol{f}(\hat{\boldsymbol{\Theta}} + \boldsymbol{\kappa}) \approx \boldsymbol{f}(\hat{\boldsymbol{\Theta}}) + \sum_i \frac{\partial \boldsymbol{f}(\hat{\boldsymbol{\Theta}})}{\partial \Theta_i}\kappa_i = \boldsymbol{f}(\hat{\boldsymbol{\Theta}}) + \boldsymbol{A}\boldsymbol{\kappa}. \tag{4.3}$$

At the minimum of $S(\boldsymbol{\Theta})$

$$\begin{aligned}
\frac{\partial}{\partial \kappa_i}\left[(\boldsymbol{Y} - \boldsymbol{f}(\hat{\boldsymbol{\Theta}} + \boldsymbol{\kappa}))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{f}(\hat{\boldsymbol{\Theta}} + \boldsymbol{\kappa}))\right] &= 0 \qquad i = 1, 2, \ldots, p \\
&\Downarrow \\
\frac{\partial}{\partial \kappa_i}\left[(\boldsymbol{\varepsilon} - \boldsymbol{A}\boldsymbol{\kappa})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varepsilon} - \boldsymbol{A}\boldsymbol{\kappa})\right] &= 0 \\
&\Downarrow \\
\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\varepsilon} - \boldsymbol{A}\boldsymbol{\kappa}) &= 0 \\
&\Downarrow \\
\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{\kappa} &= \boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}.
\end{aligned} \tag{4.4}$$

Hence $\boldsymbol{\kappa}$ is normal distributed since $\boldsymbol{\varepsilon}$ is, and the covariance matrix $\boldsymbol{P} = E[\boldsymbol{\kappa}^T\boldsymbol{\kappa}]$ can be calculated by right multiplying equation (4.4) by its transposed and taking the expectance,

$$E\left[\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{\kappa}\boldsymbol{\kappa}^T(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^T\right] = E[\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1})^T]$$
$$\Downarrow$$
$$\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}E[\boldsymbol{\kappa}\boldsymbol{\kappa}^T]\boldsymbol{A}^T\boldsymbol{\Sigma}^{-T}\boldsymbol{A} = \boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]\boldsymbol{\Sigma}^{-T}\boldsymbol{A}$$

since $\boldsymbol{A}$ and $\boldsymbol{\Sigma}$ are constant in this expression. Substituting $\boldsymbol{\Sigma} = E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T]$ and $\boldsymbol{P} = E[\boldsymbol{\kappa}\boldsymbol{\kappa}^T]$, if follows by canceling $\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}$ on both sides that

$$\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{I}.$$

If the sensitivity matrix $\boldsymbol{A}$ has full rank the expression for the covariance matrix of the estimated parameters is $\boldsymbol{P} = \left(\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\right)^{-1}$.

4.1.3 Linear covariance analysis

For a generalized least square objective function of the kind shown in equation (4.2), $\boldsymbol{P}$ is given by

$$\boldsymbol{P} = (\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A})^{-1}. \tag{4.5}$$

The sensitivity matrix of the simulated values is the Jacobian of $\boldsymbol{f}(\boldsymbol{\Theta})$ and have columns $\boldsymbol{a}_i = \partial\boldsymbol{f}/\partial\Theta_i$ $(i = 1,2,\ldots,p)$ where p is the number of parameters. It is easily seen that the matrix $\boldsymbol{P}$ becomes a square matrix with dimension $p \times p$. Note that equation (4.5) is only valid if the sensitivity matrix $\boldsymbol{A}$ has full rank, that is, the parameters must be linear independent and the number of measurements must be greater or equal to the number of parameters ($n \geq p$).

An approximate confidence region is given by

$$(\boldsymbol{\Theta}^* - \hat{\boldsymbol{\Theta}})\boldsymbol{P}^{-1}(\boldsymbol{\Theta}^* - \hat{\boldsymbol{\Theta}}) \leq \chi^2_{p,\alpha}, \tag{4.6}$$

where $\chi^2_{p,\alpha}$ is the upper α quantity for the χ^2 distribution with p degrees of freedom. If the model is linear the region will be exact. If the covariance matrix exists it is symmetric and positive definite, and $\boldsymbol{P}^{-1}$ can be diagonalized as

$$\boldsymbol{P}^{-1} = \boldsymbol{W}^T\boldsymbol{\Lambda}\boldsymbol{W}, \tag{4.7}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of positive eigenvalues and the columns of $\boldsymbol{W}$ are orthonormal eigenvectors. The confidence region can be visualized as a p-dimensional ellipsoid where the i-th semi major ellipsoid axis direction is defined by $\boldsymbol{W}_{\cdot,i}$ and has length $\sqrt{(\chi^2_{p,\alpha}/\boldsymbol{\Lambda}_{i,i})}$, visualized in Figure 4.1.

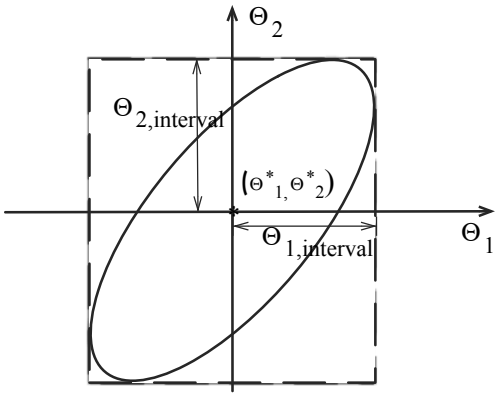


Figure 4.1: *A 2-dimensional illustration of a projection of the confidence region to obtain a* $100(1-\alpha)\%$ *joint confidence interval for the parameters. Note that the box include many points that lie outside of the* $100(1-\alpha)\%$ *confidence region.*

Donaldson and Schnabel [30] compared different types of linearization methods for calculating confidence regions and concluded that the method above (using the Jacobian matrix of $\boldsymbol{f}(\boldsymbol{\theta})$ in $\boldsymbol{P}$), the simples among them, was as good as any of the more complex methods suggested. However, it is also pointed out that the validity of the linear approximation should be investigated before trusting the resulting confidence values. The curvature measure of nonlinearity [11] is such an investigation and will be explained later in the chapter.

4.1.4 Experimental design based on LCA

The covariance matrix of the estimated parameters, $\boldsymbol{P}$, depends on the experiment setup, hence the accuracy of the estimated model function depends on the measurement errors and the variables $\boldsymbol{x}_j$. This means that the set of variables $\{\boldsymbol{x}_j\}$, which is used in an experiment will change the value of $\boldsymbol{f}$ and thus the covariance matrix $\boldsymbol{P}$. An increase in the measurement errors will generally result in an increase in size of the confidence region, therefor it will be crucial to do as exact measurement as possible. It is important to note that $\boldsymbol{P}$ is independent of measurements and can be calculated prior to the experiment if the estimated profiles is given. The experiment can be designed to minimize the uncertainty in the estimated parameters. Hence, a LCA-based experimental design seek to find the set of measurement points that minimize some scalar measure of $\boldsymbol{P}$.

There exists several different optimization criteria based on different properties of the matrix $\boldsymbol{P}$. Examples are D-Optimality, A-optimality and E-optimality which minimizes the determinant, the trace and the largest eigenvalue of the matrix $\boldsymbol{P}$ respectively [74]. The determinant of the covariance matrix for the measurement

errors, $\boldsymbol{P}$, is proportional to the volume of the error ellipsoid [17]. So the smaller the determinant of the matrix $\boldsymbol{P}$ is, the smaller the confidence region for the estimated parameters will be. The optimization problem is then to find the set of measurement points which minimize the determinant of the matrix $\boldsymbol{P}$,

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}}(\det \boldsymbol{P}(\boldsymbol{x})). \tag{4.8}$$

This is called the D-criteria in optimization theory and is one of the most common criteria in experimental design [13]. The D-criteria and is in general a good criteria, but if one parameter has very large variance compared to the others it may be better to use E-optimality [33].

4.1.5 Curvature measurements of nonlinearity

Equation (4.8), used to find a optimal setup for an experiment, is based upon the theory of LCA which is founded on the assumption that the model equation can be approximated with a linear equation with respect to the parameters $\boldsymbol{\Theta}$, from now called the linear assumption. If this assumption fails, that means if the nonlinearity is too severe, the results cannot be trusted without further investigation. The equation $\boldsymbol{f}(\boldsymbol{\Theta})$ is nonlinear in the parameters, so to find out if the linear assumption is valid a measure on how large this nonlinearity is must be found. This can be done by curvature measurements of nonlinearity (CMN) suggested by [12]. CMN gives a measure on the curvature and indicates whether the linearity assumption is valid. The curvature measures of nonlinearity are based upon the fact that a linear function has zero second and higher order derivatives, therefore an investigation of the first and second derivative of the function is done.

The curvature c_h in the direction $\boldsymbol{h}$ at a point is defined as the ratio of the length of the acceleration vector to the squared length of the tangent vector

$$c_h^{\Theta} = \frac{\|\ddot{\boldsymbol{f}}_h^{\Theta}\|}{\|\dot{\boldsymbol{f}}_h\|^2}, \tag{4.9}$$

$$c_h^{l} = \frac{\|\ddot{\boldsymbol{f}}_h^{l}\|}{\|\dot{\boldsymbol{f}}_h\|^2}. \tag{4.10}$$

Here c_h^{Θ} is the parameter effect curvature which measures the apparent nonlinearity caused by the parameterization of the model plane. The intrinsic curvature, c_h^{l}, measures the intrinsic nonlinearity of the solution locus in the direction of $\dot{\boldsymbol{f}}_h$, and is not changed by re-parameterization of the model.

The curvature measures above are defined by studying how the model equation $\boldsymbol{f}(\boldsymbol{\Theta})$ changes with respect to the parameter values. Now, let $\boldsymbol{\Theta}$ vary through all allowed values, then the function $\boldsymbol{f}(\boldsymbol{\Theta})$ defines a p-dimensional surface S, called *solution locus*, [12, 17], in the n-dimensional "sample" space. The linear assumption assumes two kinds of linearization of the solution locus near the estimated parameters, called the planar- and the uniform- assumption. The planar assumption is that the solution locus S may be replaced by the p-dimensional tangent plane

T. The uniform assumption is that the curved parameter lines on the tangent plane may be replaced with a grid of straight, parallel, equidistant lines. A measure of the linear- and uniform assumption is called the intrinsic and parameter effect curvature respectively.

The shape of the solution locus is investigated through the nonlinearity. An arbitrary straight line through $\boldsymbol{\Theta}_0$ in the parameter space, $\boldsymbol{\Theta}(b) = \boldsymbol{\Theta}_0 + b\boldsymbol{h}$, where $\boldsymbol{h}$ is a nonzero vector and b is a real constant, defines a curve $\boldsymbol{f}_{\boldsymbol{h}}(b)$ on the solution locus

$$\boldsymbol{f}_{\boldsymbol{h}}(b) = \boldsymbol{f}(\boldsymbol{\Theta}(b)) = \boldsymbol{f}(\boldsymbol{\Theta}_0 + b\boldsymbol{h}).$$

If $\boldsymbol{f}(\boldsymbol{\Theta})$ is linear then S will be a plane and $\boldsymbol{f}_{\boldsymbol{h}}(b)$ a straight line on S. In general, $\boldsymbol{f}(\boldsymbol{\Theta})$ is not linear and the curve $\boldsymbol{f}_{\boldsymbol{h}}(b)$ is not a straight line, see Figure 4.2. The tangent to $\boldsymbol{f}_{\boldsymbol{h}}$ at $b = 0$ is

$$\dot{\boldsymbol{f}}_h = \frac{\partial \boldsymbol{f}}{\partial b} \,|_0 = \sum_i \frac{\partial \boldsymbol{f}}{\partial \Theta_i} \,|_{\Theta_0} \, \frac{\partial \Theta_i}{\partial b} \,|_0 = \sum_i a_i h_i = \boldsymbol{A}\boldsymbol{h}.$$

Here $\boldsymbol{A}$ is the sensitivity matrix with dimension $n \times p$ introduced in equation (4.3). The tangent vector to $\boldsymbol{f}_h$ is a linear combination of the vectors $\{a_i\}$ and the set of all such linear combinations is the tangent plane, T, at $\boldsymbol{f}(\boldsymbol{\Theta}_0)$. The acceleration of the curve is similarly calculated

$$\ddot{\boldsymbol{f}}_h = \frac{\partial^2 \boldsymbol{f}}{\partial^2 b} \,|_0 = \sum_j \frac{\partial \sum_i a_i h_i}{\partial \Theta_j} \,|_{\Theta_0} \, \frac{\partial \Theta_j}{\partial b} \,|_0 = \sum_i \sum_j h_i h_j a_{ij} = \boldsymbol{h}^T \ddot{\boldsymbol{A}} \boldsymbol{h},$$

where $\ddot{\boldsymbol{A}}$ for one single measurement is a $p \times p$ matrix with elements $a_{ij} = \partial^2 \boldsymbol{f}(\boldsymbol{\Theta}_0) / \partial \Theta_i^T \partial \Theta_j$. For n measurements $\ddot{\boldsymbol{A}}$ will be an $n \times p \times p$ array.

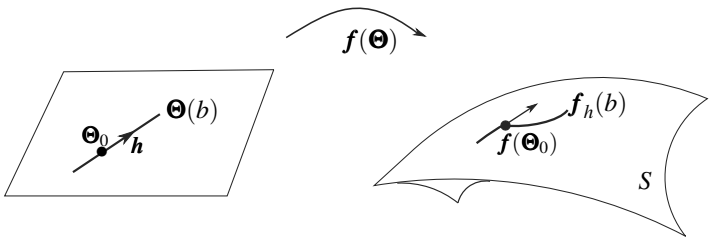


Figure 4.2: *Mapping of the line $\boldsymbol{\Theta}(b)$ from parameter to sample space in 2 dimensions.*

The acceleration can be decomposed into a tangential, $\ddot{\boldsymbol{f}}_h^{\Theta}$, and normal, $\ddot{\boldsymbol{f}}_h^{l}$, component, $\ddot{\boldsymbol{f}}_h = \ddot{\boldsymbol{f}}_h^{\Theta} + \ddot{\boldsymbol{f}}_h^{l}$ [11]. Then the parameter effect and intrinsic curvature in the direction $\boldsymbol{h}$ at a point is defined by equation (4.9) and (4.10). For an algorithm on how to calculate the curvatures, see [12, 72].

The extent to which the acceleration vectors lie outside the tangent plane measures how much the expectation surface deviates from a plane, and hence measures

the nonplanarity of the expectation surface. This nonplanarity is called intrinsic nonlinearity because it does not depend on the parameterization chosen for the expectation function, only on the design and the expression for the expectation function. The projections of the acceleration vectors in the tangent plane depend on the parameterization, and measure the non uniformity of the parameter lines on the tangent plane and is called parameter effects (nonlinearity).

To make the curvatures invariant under scaling Bates and Watts [11] points out that c_h^{Θ} and c_h^{l} must be converted to standardized *relative curvatures* denoted γ^{Θ} and γ^{l} respectively. This should be done by normalized the curvature with respect to the number of parameters used and the magnitude of the model function values. If the standard deviation of the measurements is assumed to be known, a convenient scaling is to divide the output of the model function, and thus all derivatives thereof, by the appropriate standard deviation ϑ. Assuming this is done, the *maximum* relative parameter effects curvature is defined as

$$\gamma^{\Theta} = \max_{\boldsymbol{h}} c_{\boldsymbol{h}}^{\Theta}, \tag{4.11}$$

and the *maximum* relative intrinsic curvatures as

$$\gamma^{l} = \max_{\boldsymbol{h}} c_{\boldsymbol{h}}^{l}. \tag{4.12}$$

To assess the impact of the nonlinearity on a particular confidence region, Bates and Watts [11] suggested to compare the relative curvature measures with cut off values. When the error distribution is assumed to be known, the cut off value $\gamma_c = 1/(\chi^2_{p,\alpha})^{1/2}$. If the parameter effect curvature is small compared to this critical value for all directions, then the uniform coordinate assumption is acceptable over the region of interest. Similarly the planar assumption is acceptable if the intrinsic curvature is small compare to the cut off value for all directions $\boldsymbol{h}$. If the value

$$\frac{\gamma_h^{\Theta,l}}{\gamma_c} > 1,$$

then linear covariance analysis should not be trusted without further investigation. However this cut-off value does not represent a sharp division between acceptable and unacceptable nonlinearity.

A wide study by different authors [26, 30, 76], is done on the validation and accuracy on the curvature measures and the cut of value suggested by Bates and Watts [11]. These studies investigate how well warnings about use of LCA obtained from CMN correspond with comparison of LCA results to Monte-Carlo results. Sylte et al. [76] indicate that the cut off value may be too strict for some types of problems and show, through Monte Carlo study, that the linear assumption may be valid even if the parameter curvature is above the critical value, at least if one is content with LCA providing conservative results. Grimstad et al. [38] elaborate further on this observation, and point out a connection to a link between low sensitivity and high nonlinearity for certain directions in parameter space for a large class of parameter estimation problems.

4.2 The genetic algorithm

The main optimization method used in the included papers is the genetic algorithm (GA), due to its flexibility, effectivity and curiosity in the method. The method is directly inspired by biology and was first introduced by John Henry Holland in the book "Adaptation in natural and artificial systems" in 1975 [43]. The genetic algorithm has become very popular recent years because that it has been shown to be efficient on problems that are not analytically solvable, or are too time consuming to solve with dynamic programing due to explosion of possible combinations. It is favorable because it can handle many variables and it does not need information on the derivative of the function, which is often the case for classical optimization methods. Below is a short introduction to genetic algorithm. Much of the presented text is taken from [37, 70], which are good introductions to this subject.

The genetic algorithm is a method that imitates population genetics based upon Darwin's theory about survival of the fittest. Having an objective function, GA starts out with a random population of possible solutions, called individuals, thereafter selection, crossover and mutation are applied to produce a new population of individuals, see Figure 4.4. The idea is that these steps will eventually make the solution converge towards the optimal solution. The essence is that the good individuals have high probability to reproduce, resulting in even better offspring, while bad individuals are rejected. The idea and concept behind GA is very simple, however the algorithm can be complex and hard to implement. The work presented in the thesis has used Matlab's inbuilt genetic algorithm called *ga*.

First of all, a fitness function is needed to evaluate the success of each solution, a quantitative score, usually low score is connected to high success, hence the GA seeks after the solution that minimizes the fitness function. Note that throughout this chapter we will use the term high fitness and high score for the best individuals. After the fitness function is given, a set of many possible solutions, called a population, is generated. The solutions, referred to as individuals, which are different combinations of variables in the problem can be represented as chromosomes. The two most common representation methods are binary strings and floating points. Usually a floating point representation is preferred, as it generally is faster [41]. However, for simplicity, the algorithm will be explained by using a binary representation.

4.2.1 Selection

An initial population is created by randomly drawing possible solutions (individuals). Therafter, each individual is equipped with a fitness given by the original fitness function. A population consist of N_{pop} individuals and the user chooses how large that number should be. The larger the population is, the slower the algorithm will be. However, if N_{pop} is to low it may not cover a big enough area and the method can fail to find the optimum.

After the initial population is drawn and evaluated a selection is performed. In the selection process the best individuals, denoted N_{keep}, are selected and survives

to the next generation. The individuals with lowest fitness are thrown away, and how many to reject must be decided in advanced. For instance it is possible to keep the same amount of individuals each time or to set a threshold on the fitness function and keep those above. If a fixed number of individuals is kept each time they must be sorted.

When the N_{keep} individuals that survive is decided the selection starts, to favor the best fitted individuals a selection is done from a probability distribution corresponding to the rank or the fitness value [41]. One of the methods used to select individuals is roulette wheel selection; all the individuals are given a spot in a roulette wheel corresponding to their fitness, the higher fitness the larger space they occupies. Then the wheel is spun the individual at the stop is selected, this continues until enough parents are selected. Alternatively tournaments can be used to select individuals, then a predefined number of individuals are randomly chosen and the best one is picked out to reproduce. This prevents the least fitted individual to be selected.

4.2.2 Crossover

The process when one or more offspring is created from the selected parents is termed crossover. Most commonly two parents are selected to produce two new offsprings. One crossover point is chosen to separate the parents into two strings, and two offsprings are created by combining the strings. The front part of parent 1 is crossed with the back part of parent 2 to create offspring 1 and to create offspring 2 the back part of parent 1 is crossed with the front part of parent 2, see illustration in Figure 4.3. It is of course possible to choose several crossover points. The size of the new population should be the same as the initial population, hence it must be created N_{new}=N_{pop}−N_{keep} new individuals.

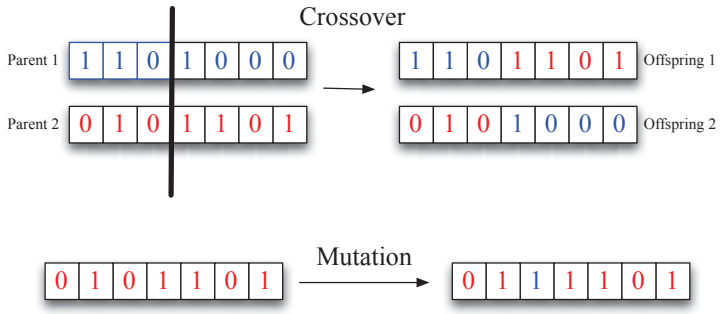


Figure 4.3: *Illustration of single point crossover between to parents creating two new offsprings, top figure. The bottom figure show an example of a mutation in one individual.*

4.2.3 Mutation

Once all new individuals are created, one more step is needed before the new population is created. The step, called mutation, is an alternation in the "characteristic" of the individual and in binary string this mean to alter a 0 to a 1 or a 1 to a 0 in the individual, see Figure 4.3. To avoid loosing the best individual, it is given immunity against mutation. The best individual is termed elite individual, and can be one or several individuals. Even though the mutation rate in general is low, often less than 1% of the individuals are mutated, it is a very important step in the algorithm. A mutation will force the algorithm to seek after solutions in a broader area of the solution domain and hopefully it prevents the algorithm converging towards a local minimum.

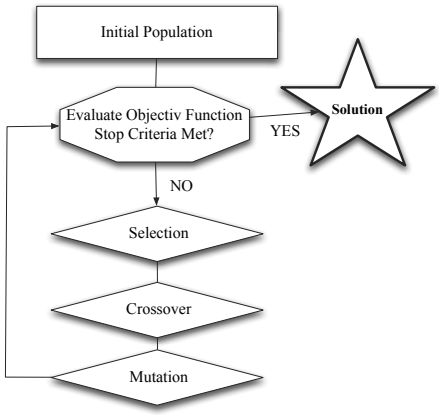


Figure 4.4: *Steps in the Genetic Algorithm*

4.2.4 Convergence

When the selection, crossover and mutation is finished the new population is evaluated by the fitness function, if the stop criteria is satisfied, the algorithm stop and the solution is given. If not a new population is created using selection, crossover and mutation over again, as illustrated in Figure 4.4. There are several ways to set a stop criteria; it can be a given fitness value; that the elite individual has been the same for several generations; that the max generation or time limit is reached; or that the difference between the best individual and the mean of the n next best individuals are closer than a given limit.

Note that there are no rigorous proofs showing that the GA will converge towards the global optimum. Nevertheless, several theorems and hypotheses have been developed to give theoretical explanations to the effectiveness of GA [49, 71].

Chapter 5

Summary of papers and outlook

The main contribution from this thesis is found in the included papers. In this final chapter the papers are summarized and put into a broader context. The papers can be grouped into two categories based on their objective. Paper A and B assessing parameter uncertainty in the rise velocity of a single droplet using experimental design and investigate its effect on rise and dissolution of CO_2 in the water column. In Paper C and D the leakage detection probability in a variable marine environment is studied.

5.1 Rise and dissolution of a single CO_2: Assessing parameter uncertainties using experimental design

The risk and environmental impact assessments required for geological CO_2 storage projects will have to rely on different types of numerical models, which will have to be calibrated and validated against measurements. Available measurements from ongoing demonstration projects are limited; hence it is necessary to turn to analog processes or laboratory experiments to estimate model parameters. In any case, parameter estimates will have uncertainties that will be important to assess when predicting future scenarios. Two important sub-models in this respect is the rise velocity and mass transfer of CO_2 in the water column, both of them have parameters to estimate from experiments, as discussed in Chapter 3.

Summary

In Paper A we study a model for the rise velocity of single hydrate free CO_2 droplets in the ocean, as the origin we use the parameter estimation study by [14] based on a tank experiment [15]. Under three different pressure and temperature (P/T) conditions droplets of varying sizes were released and the rise velocity was measured. The unknown parameters in the rise velocity was estimated using least square approximation.

We illustrate how linearized covariance analysis (LCA) can be used to assess the parameter uncertainties. The underlying equation for calculating the parame-

ter uncertainly, the covariance matrix for the estimated parameters, equation (4.5), depends on the measurement errors and the experimental setup. Hence it is independent of the measurements, therefore the experimental setup can be designed to minimize the parameter uncertainty. In Paper A we suggest how to design an experiment for the three different conditions to reduces these uncertainties. It is shown that up to $\sim 63\%$ reduction in uncertainties is achieved by choosing the right droplet size distribution and by extending the range of droplet sizes to include larger droplets the uncertainties are reduced by another $\sim 88\%$.

In paper B we include both hydrate coated and hydrate free CO_2 droplets, and parameter uncertainties in the rise velocity for both droplet types are estimated from the experiment data [15]. As a progress from the paper A we now include all the experimental data for each droplet type, hence merge the three different P/T conditions to one, and then calculate parameter uncertainties. The assessment shows that parameter uncertainties effects large droplets more.

Finally the rise velocity is coupled with a mass transfer model to simulate the fate of dissolution of a single droplet to assess the effect the parameter uncertainties have on where and how fast the CO_2 dissolves in the water column. Since the parameter uncertainties are highest for large droplets the uncertainty effects simulation of larger droplets more. Keeping in mind that the simulations performed in the work is strictly theoretical, it is concluded that the uncertainties in the rise velocity model have limited effect on the simulations performed in the study.

It is important to note that the underlying linear assumption that the LCA method is based upon must be carefully investigated before the results are trusted. For this curvature measures of nonlinearity is used, and our investigation gave curvature values both above and below the critical cut of value suggested by [12]. Since the associated curvature values above the critical value were very similar to those with curvature values below the critical cut of value it strengthened our confidence in the accuracy of the LCA results. However, if the uncertainties should be studied and included in a more complex model a proper investigation, through Monte Carlo simulations, should be performed.

The rise velocity equation used is based on experimental data at pressures higher than 800 meter. This is not necessarily representative for shallower marginal seas, like the North Sea, where the natural and seasonal temperature variations are higher than in the deep ocean. The challenge is to predict the impact on marine waters from CO_2 release and dissolution at such depths. For this experimental and in-situ data at appropriate depths will be vital for model calibration and validation. A possible next step would be to include the method described and used in the papers to assist in designing similar experiments for CO_2 bubbles aiming to reduce parameter uncertainty in the equations.

5.2 Optimizing leak detection probability

Monitoring of the marine environment for indications of a leak, or precursors of a leak, will be an intrinsic part of any subsea CO_2 storage projects. A real challenge will be quantification of the probability of a given monitoring program to detect a leak and to design the program accordingly. The task complicates by the number of pathways to the surface, difficulties to estimate probabilities of leaks and fluxes, and predicting the fluctuating footprint of a leak.

Summary

The objectives of paper C and D is to present a procedure for optimizing the layout of chemical sensors on the seafloor, using the probability of detecting a leak as metric. Both the use of fixed installations and single measurement is included in the study. Synthetic maps are used as a basis for probable leakage points, and it is assumed that the probability for leakage inside the map is one. The spatial footprint is based on results from a General Circulation Model (GCM). An 800 meter resolution regional ocean model, BOM, set up for the North Sea is advecting and dispersing CO_2 as a passive tracer [3]. The model is terrain following and the grid cells closest to the seafloor is used in both studies.

In Paper C the average CO_2 concentration from BOM is used as a basis for CO_2 leakage on the seafloor. The simulated footprint is approximated by an analytical function, which is used as the leakage footprint. Given a threshold concentration for when a sensor can detect a leakage, the leakage footprint is used to define a area around a leakage for when a sensor can detect a leakage. If a senor is placed inside this area the leakage will be detected. Assuming that the leakage is independent of leakage site, the problem can be inverted from the leakage location to the sensor location. Now it is possible to quantify the probability for a sensor to detect a leakage, and the optimization problem is to find the set of sensor location such that the probability to detect leakage is optimized. It is shown that with an optimization the detection probability can almost double compared to placing sensors in an equidistant grid. It is not necessarily best to place the first in the location of the highest probable leakage point, one sensor can monitor several potential leakage points.

Admitting that the leakage footprint used in Paper C suffer from major drawbacks, limiting the detection area compared with the simulated average and introducing unnecessarily uncertainties into the problem, Paper D introduce a more realistic detection approach. Now the time series in a array of 51×51 grid points around the leakage location, is used to define three detection methods; i) the average method for comparison with Paper C, ii) the event based method accounting for occasionally elevated concentrations and iii) the single measurement method. The event based method represent fixed installations and the last one represent measurements taken during survey. The optimization problem is the same as in Paper C, however we use the predicted footprint directly as the leakage probability around the seep location, hence eliminating an unnecessarily uncertainty and simplification imposed by the fitting to an analytical function. The main conclusion

is that fixed installations offers much higher detection probability compared to the same amount of single measurements.

Even though some of the assumptions made may be too general, and not necessarily realistic, we give an optimization strategy for monitoring the seafloor. The method is very general and can be used on all types of areas where the probability distribution of leakage is given, provided that the leakage assumed to be independent of the leakage site. It might readily be extended to combinations of different instrumentations, and for location dependent footprint due to e.g. topography or leakage rates. For instance the method can easily assist in finding the optimal layout if a combination of fixed installations and single measurement is required.

The investment and operational cost of the monitoring program can be included in the cost function for the optimization, as can other constrains. A natural next step in this research will be to find a way to optimize when the leakage rate may change, for instance assuming that a leakage through a well will be generally larger than through sediments.

5.3 Further work and perspective

The work in this thesis contributes in both end in the problem of designing a monitoring program. The first works studies the uncertainties in the coupled differential equation for rise and dissolution of a droplet in the marine waters. In the other end, Paper C and D aims to design a monitor program for optimizing the leakage detection problem. The rise velocity equation we have studied is a simple sub-model within a more complex model simulating CO_2 seeps to, and transport in, marine waters. This seep model will have to work in assembly with other models in order to assess environmental impact imposed by a leak from a CO_2 geological storage site. For instance, it will have to rely on a porous media model simulating CO_2 transport through sediments as input, and the output from the model will serve as input to ecosystem models. Event though the theoretical examples studied in Paper B show that the parameter uncertainty in the rise have little influence on the results, it may be that the outcome is different for a more complex model.

Ideally we would have done an uncertainty assessment of the parameters in rise velocity for a CO_2 bubble. This equation is incorporated in the plume model, which is used to simulate leakage scenarios for different current conditions. The resulting vertical CO_2 concentration distribution from these simulations is used as the source term in BOM, the model used to simulate the leakage footprint. First we could have investigated how the uncertainties propagate through the whole model process and impose uncertainties in the model output. Thereafter we can study how the uncertainty in the final leakage footprint will influence on the leak detection probability. However for complex model system the uncertainty propagation can be hard to analyze, often statistics have to be built through e.g. Monte Carlo methods. Such models are often computationally demanding, making studies of uncertainty propagation time consuming and often impossible.

According to [51], the non-engineering part of the CO_2 storage has the largest

gaps in environmental information and understanding and the validation of models for CO_2 transport is at an infant stage. Due to this it is not possible to perform a confident risk assessment of storage projects. To close this gap it is essential that modelers and experimentalists work together in establishing and designing new experiments and measurement campaigns. Aiming to assemble invaluable data for model calibration and for model validation. With the many different kind of models needed to study impact of a leak; there is a need for many different experiments and campaigns.

Both measurements and models have uncertainties and errors that are important to assess. As illustrated in the papers, both the measurement errors and the experimental setup will influence directly on the uncertainties in the model parameters. However, with a proper design of experiments it is possible to assess these uncertainties and reduce them. Hence, it is important that experimentalists and modelers communicate and work together when planning data gathering. This is a win-win situation; the parameter uncertainties can be reduced and the cost of the data gathering can be lowered through demand for less data for the same accuracy. Since uncertainties in integrated sub models will influence the output the complex system it will always be of interests to minimize these. In this respect experimental design offers an inventible tool in assessing and minimizing these uncertainties.

The two main components of the monitoring design in paper C and D are the probability map, and footprint statistics. To obtain these it is evident that a proper survey of the overburden and the marine environment is important. Due to the amount of data collected during a monitoring program, the quality of the baseline will increase during site operations. How the monitoring program can best assist in this quality improvement is not discussed here, it is an issue that needs to be addressed. For instance long time series of current taken at strategic locations will surely make the uncertainties in the footprint assessment smaller.

If CCS is to be commercially executed with the potential of decreasing CO_2 emission it must be carefully monitored, both in the geological formation and in the water column. To be able to estimate spatial and temporal distribution of a leakage footprint a proper baseline of the marine environment is needed. Therefore, the monitoring program for the marine environment must be incorporated in the characterization and assessment of the storage site, and continue for decades after the injection has ended. In that respect an optimization strategy for the monitor design will an invaluable tool.

Bibliography

[1] PMEL Carbon Program: Pacific Marine Environment Laboratory, visited: April 2015. http://pmel.noaa.gov/co2/story/Ocean+Acidification. 1.1, 1.1

[2] ALENDAL, G., BERNTSEN, J., ENGUM, E., FURNES, G. K., KLEIVEN, G., AND EIDE, L. I. Influence from 'Ocean Weather' on near seabed currents and events at Ormen Lange. *Marine and Petroleum Geology 22*, 1 (2005), 21–31. 3

[3] ALENDAL, G., DEWAR, M., ALI, A., EVGENIY, Y., VIELSTÄDTE, L., AVLESEN, H., AND CHEN, B. 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. 2.1, 2.2, 2.3, 5.2

[4] ALENDAL, G., AND DRANGE, H. Two-phase, near-field modeling of purposefully released CO_2 in the ocean. *Journal of Geophysical Research 106*, C1 (2001), 1085–1096. 2.1, 3, 3.3

[5] ANNUNZIATELLIS, A., BEAUBIEN, S., CIOTOLI, G., FINOIA, M., GRAZIANI, S., AND LOMBARDI, S. Development of an innovative marine monitoring system for CO_2 leaks: system design and testing. *Energy Procedia 1*, 1 (2333–2340), 2333–2340. 2.2

[6] ASTER, R. C., BORCHERS, B., AND THURBER, C. H. *Parameter Estimation and Inverse Problems*. Elsevier Academic Press, 2005. ISBN 0-12-065604-3. 4.1.1

[7] ATKINSON, A., DONEV, A., AND TOBIAS, R. *Optimum Experimental Designs, with SAS*. Oxford University Press New York, 2007. ISBN: 9780199296590. 4.1.2

[8] AYA, I., YAMANE, K., AND NARIAI, H. Solubility of CO_2 and density of CO_2 hydrate at 30 MPa. *Energy 22*, 2 (1997), 263–271. 3, 3.2

[9] AZAR, C., LINDGREN, K., LARSON, E., AND MÖLLERSTEN, K. Carbon capture and storage from fossil fuels and biomass–Costs and potential role in stabilizing the atmosphere. *Climatic Change 74*, 1–3 (2006), 47–79. 1.2

[10] AZAR, C., LINDGREN, K., OBERSTEINER, M., RIAHI, K., VAN VUUREN, D. P., DEN ELZEN, K. M. G., MÖLLERSTEN, K., AND LARSON, E. D. The feasibility of low CO_2 concentration targets and the role of bio-energy with carbon capture and storage (BECCS). *Climatic Change 110*, 1 (2010), 195–202. 1.2

[11] BATES, D., AND WATTS, D. Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological) 42* (1980), 1–25. 4.1.3, 4.1.5, 4.1.5

[12] BATES, D. M., AND WATTS, D. G. *Nonlinear Regression Analysis and its Applications*. John Wiley & Sons, 1988. ISBN 0-471-81643-4. 4.1, 4.1.5, 4.1.5, 4.1.5, 5.1

[13] BECK, J., AND WOODBURY, K. Inverse problems and parameter estimation: integration of measurements and analysis. *Measurement Science and Technology 9* (1998), 839–847. 4.1.4

[14] BIGALKE, N., ENSTAD, L., REHDER, G., AND ALENDAL, G. Terminal velocities of pure and hydrate coated CO_2 droplets and CH_4 bubbles rising in a simulated oceanic environment. *Deep Sea Research Part I: Oceanographic Research Papers 57*, 9 (2010), 1102–1110. 3, 3.1, 3.1, 3.1, 3.1, 3.3, 3.1, 5.1

[15] BIGALKE, N., REHDER, G., AND GUST, G. Experimental Investigation of the Rising Behavior of CO_2 Droplets in Seawater under Hydrate-Forming Conditions. *Environmental Science & Technology 42*, 14 (2008), 5241–5246. 3.1, 3.1, 5.1

[16] BLACKFORD, J. C. Predicting the impacts of ocean acidification: Challenges from an ecosystem perspective. *Journal of Marine Systems 81*, 1 (2010), 12–18. 2.1, 2.2

[17] BOX, G., AND LUCAS, H. Design of experiments in non-linear situations. *Biometrika 46*, 1/2 (1959), 77–90. 4.1.4, 4.1.5

[18] BOYD, A. D., LIU, Y., STEPHENS, J. C., WILSON, E. J., POLLAK, M., PETERSON, T. R., EINSIEDEL, E., AND MEADOWCROFT, J. Controversy in technology innovation: Contrasting media and expert risk perceptions of the alleged leakage at the weyburn carbon dioxide storage demonstration project. *International Journal of Greenhouse Gas Control 14* (may 2013), 259–269. 2.1

[19] BOZZANO, G., AND DENTE, M. Shape and terminal velocity of single bubble motion: a novel approach. In *European Symposium on Computer Aided Process Engineering-10*, S. Pierucci, Ed., vol. 8 of *Computer Aided Chemical Engineering*. Elsevier, 2000, pp. 649 – 654. 3.1

[20] BREWER, P., CHEN, B., WARZINKI, R., BAGGEROER, A., PELTZER, E., DUNK, R., AND WALZ, P. Three-dimensional acoustic monitoring and modeling of a deep-sea CO_2 droplet cloud. *Geophys. Res. Lett 33* (2006), L23607. 3

[21] BREWER, P., PELTZER, E., FRIEDERICH, G., AND REHDER, G. Experimental Determination of the Fate of Rising CO_2 Droplets in Seawater. *Environmental Science & Technology 36*, 24 (2002), 5441–5446. 3.1, 3.2

[22] CALDEIRA, K., AND WICKETT, M. E. Oceanography: anthropogenic carbon and ocean pH. *Nature 425*, 6956 (2003), 365–365. 1.1

[23] CALDEIRA, K., AND WICKETT, M. E. Ocean model predictions of chemistry changes from carbon dioxide emissions to the atmosphere and ocean. *Journal of Geophysical Research: Oceans (1978–2012) 110*, C9 (2005). 1.1, 2.1

[24] CHEN, B., SONG, Y., NISHIO, M., AND AKAI, M. Large-eddy simulation of double-plume formation induced by CO_2 dissolution in the ocean. *Tellus B 55*, 2 (2003), 723–730. 3.2, 3.3

[25] CLIFT, R., GRACE, J. R., AND WEBER, M. E. *Bubbles, drops, and particles*. Academic Press, New York:, 1978. ISBN 0-12-176950-x. 3, 3.1, 3.1

[26] COOK, R., AND WITMER, J. A note on parameter-effects curvature. *Journal of the American Statistical Association 80* (1985), 872–878. 4.1.5

[27] CROWE, C., SOMMERFELD, M., AND TSUJI, Y. *Multiphase Flow with Droplets and Particles*. CRC Press LLC, 1998. 3, 3.2

[28] DEWAR, M., SELLAMI, N., AND CHEN, B. Dynamics of rising CO_2 bubble plumes in the QICS field experiment: Part 2 – Modelling. *International Journal of Greenhouse Gas Control*, 0 (2014), –, doi: http://dx.doi.org/10.1016/j.ijggc.2014.11.003. 2.1, 3

[29] DEWAR, M., WEI, W., MCNEIL, D., AND CHEN, B. Small-scale modelling of the physiochemical impacts of CO_2 leaked from sub-seabed reservoirs or pipelines within the north sea and surrounding waters. *Marine Pollution Bulletin* (2013). 2.3

[30] DONALDSON, J., AND SCHNABEL, R. Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics 29* (1987), 67–82. 4.1.3, 4.1.5

[31] DRANGE, H., ALENDAL, G., AND JOHANNESSEN, O. Ocean release of fossil fuel CO_2: A case study. *Geophysical Research Letters 28* (2001), 2637–2640. 3.3

[32] DREW, D. A., AND PASSMAN, S. L. *Theory of Multicomponent Fluids*. Springer, 1999. 3.3

[33] EMERY, A. F., AND NENAROKOMOV, A. V. Optimal experiment design. *Meas. Sci. Technol. 9* (1998), 864–976. 4.1.4

[34] EUROPEAN COMMUNITIES . *Implementation of Directive 2009/31/EC on the Geological Storage of Carbon Dioxide: CO_2 Storage Life Cycle Risk Management Framework*. ISBN-13978-92-79-19833-5, DOI: 10.2834/9801. 2011. 2

[35] FRIEDLINGSTEIN, P., ANDREW, R., ROGELJ, J., PETERS, G., CANADELL, J., KNUTTI, R., LUDERER, G., RAUPACH, M., SCHAEFFER, M., VAN VUUREN, D., ET AL. Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nature Geoscience 7*, 10 (2014), 709–715. 1.1, 1.2

[36] GANGSTØ, R., HAUGAN, P., AND ALENDAL, G. Parameterization of drag and dissolution of rising CO_2 drops in seawater. *Geophysical research letters 32* (2005), L10612. 3.1, 3.1, 3.2

[37] GOLDBERG, D. *Genetic Algorithms in optimization, search and machine learning*. Addison-Wesley Publishing Company, 1989. 4.2

[38] GRIMSTAD, A., KOLLTVEIT, K., MANNSETH, T., AND NORDTVEDT, J. Assessing the validity of a linearized accuracy measure for a nonlinear parameter estimation problem. *Inverse Problems 17* (2001), 1373–1390. 4.1.5

[39] GRIMSTAD, A. A. *On nonlinearities and covariance analysis for the reservoir fluid flow model*. PhD thesis, University of Bergen, 1999. 4.1.2

[40] HAUGAN, P. M., AND JOOS, F. Metrics to assess the mitigation of global warming by carbon capture and storage in the ocean and in geological reservoirs. *Geophysical Research Letters 31*, 18 (2004), L18202. 2, 2.1

[41] HAUPT, R. L., AND HAUPT, S. E. *Practical genetic algorithms*. John Wiley & Sons, 2004. 4.2, 4.2.1

[42] HIRAI, S., OKAZAKI, K., TABE, Y., HIJIKATA, K., AND MORI, Y. Dissolution rate of liquid CO_2 in pressurized water flows and the effect of clathrate films. *Energy 22*, 2 (1997), 285–293. 3.2

[43] HOLLAND, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. The University Michigan Press, 1975. 4.2

[44] IEA. Technology Roadmap: Carbon capture and storage. Tech. rep., INTERNATIONAL ENERGY AGENCY, 2013. 1.2, 2

[45] IEAGHG. Quantification Techniques For CO_2 Leakage. Tech. rep., IEAGHG, 2012/01. 2.1

[46] IPCC 2013: SUMMARY FOR POLICYMAKERS: CLIMATE CHANGE 2013: THE PHYSICAL SCIENCE BASIS. CONTRIBUTION OF WORKING GROUP I TO THE FIFTH ASSESSMENT REPORT OF THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE 2013. *Stocker, TF and Qin, D and Plattner, GK and Tignor, M and Allen, SK and Boschung, J and Nauels, A and Xia, Y and Bex, V and Midgley, PM*. Cambridge University Press, Cambridge, UK, and New York. 1

[47] IPCC, 2014: CLIMATE CHANGE 2014: IMPACTS ADAPTATION AND VULNERABILITY. PART A: GLOBAL AND SECTORAL ASPECTS. CONTRIBUTION OF WORKING GROUP II TO THE FIFTH ASSESSMENT REPORT OF THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. *Field, C.B. and Barros V.R. and Dokken, D.J. and Mach, K.J. and Mastrandrea, M.D. and Bilir, T.E. and Chatterjee, M. and Ebi, K.L. and Estrada, Y.O. and Genova, R.C. and Girma, B. and Kissel, E.S. and Levy, A.N. and MacCracken, S. and Mastrandrea, P.R. and White, L.L. (eds.)*. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014. 1.1

[48] IPCC 2014, SUMMARY FOR POLYCYMAKERS. IN: CLIMATE CHANGE 2014: MITIGATION OF CLIMTE CHANGE. CONTRIBUTION OF WORKING GROUP III TO THE FIFTH ASSESMENT REPORT OF THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. *Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)*. Cambridge University Press, Cambridge, UK, and New York. 1.1

[49] JHA, M. K. 16 - Metaheuristic Applications in Highway and Rail Infrastructure Planning and Design: Implications to Energy and Environmental Sustainability. In *Metaheuristics in Water, Geotechnical and Transport Engineering*, X.-S. Y. H. G. T. H. Alavi, Ed. Elsevier, Oxford, 2013, pp. 365 – 384. 4.2.4

[50] KELBALIYEV, G. Drag coefficients of variously shaped solid particles, drops, and bubbles. *Theoretical Foundations of Chemical Engineering 45*, 3 (2011), 248–266. 3.1

[51] KOORNNEEF, J., RAMÍREZ, A., TURKENBURG, W., AND FAAIJ, A. The environmental impact and risk assessment of CO_2 capture, transport and storage-an evaluation of the knowledge base. *Progress in Energy and Combustion Science 38*, 1 (2011), 62–86. 5.3

[52] KORRE, A., IMRIE, C. E., MAY, F., BEAUBIEN, S. E., VANDERMEIJER, V., PERSOGLIA, S., GOLMEN, L., FABRIOL, H., AND DIXON, T. Quantification techniques for potential CO_2 leakage from geological storage sites. *Energy Procedia 4* (2011), 3413–3420. 2.2

[53] KRIEGLER, E., WEYANT, J. P., BLANFORD, G. J., KREY, V., CLARKE, L., EDMONDS, J., FAWCETT, A., LUDERER, G., RIAHI, K., RICHELS, R., ET AL. The role of technology for achieving climate policy objectives: overview of the EMF 27 study on global technology and climate policy strategies. *Climatic Change 123*, 3–4 (2014), 353–367. 1.2

[54] LE QUÉRÉ, C., PETERS, G., ANDRES, R., ANDREW, R., BODEN, T., CIAIS, P., FRIEDLINGSTEIN, P., HOUGHTON, R., MARLAND, G., MORIARTY, R., ET AL. Global carbon budget 2013. *Earth System Science Data Discussions 6*, 2 (2013). 1.1

[55] LEVENBERG, K. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathmatics II*, 2 (1944), 164–168. 4.1.1

[56] MAROTO-VALER, M. M., Ed. *Developments and innovation in carbon dioxide CO_2 capture and storage technology, Volume 1, Carbon dioxide (CO_2) capture, transportation and industrial applications*. 9781613443880. CRC Press, 2010. 2.1

[57] MARQUARDT, D. W. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics 11*, 2 (1963), 431–441. 4.1.1

[58] MCDOUGALL, T. J., AND BARKER, P. M. *Getting started with TEOS-10 and the Gibbs Seawater (GSW) oceanographic toolbox*, vol. 127. Trevor J. McDougall, 2011. 3.3

[59] MCGLADE, C., AND EKINS, P. The geographical distribution of fossil fuels unused when limiting global warming to 2^{O}C. *Nature 517*, 7533 (2015), 187–190. 1.1

[60] MORI, Y., AND MOCHIZUKI, T. Dissolution of liquid CO_2 into water at high pressures: A search for the mechanism of dissolution being retarded through hydrate-film formation. *Energy Conversion and Management 39*, 7 (1998), 567–578. 3, 3.2

[61] NOBLE, R. R. P., STALKER, L., WAKELIN, S. A., PEJCIC, B., LEYBOURNE, M. I., HORTLE, A. L., AND MICHAEL, K. Biological monitoring for carbon capture and storage – a review and potential future developments. *International Journal of Greenhouse Gas Contro 10*, 520–535 (sep 2012). 2.2

[62] NORDBOTTEN, J., FLEMISCH, B., GASDA, S., NILSEN, H., FAN, Y., PICKUP, G., WIESE, B., CELIA, M., DAHLE, H., EIGESTAD, G., ET AL. Uncertainties in practical simulation of CO_2 storage. *International Journal of Greenhouse Gas Control 9* (2012), 234–242. 2.2

[63] OGASAWARA, K., YAMASAKI, A., AND TENG, H. Mass transfer from CO_2 drops traveling in high-pressure and low-temperature water. *Energy & Fuels 15*, 1 (2001), 147–150. 3.2

[64] OLDENBURG, C. M., AND LEWICKI, J. L. On leakage and seepage of co2 from geologic storage sites into surface water. *Environmental Geology 50*, 5 (jul 2006), 691–705. 2.2

[65] OZAKI, M., MINAMIURA, J., KITAJIMA, Y., MIZOKAMI, S., TAKEUCHI, K., AND HATAKENAKA, K. CO_2 ocean sequestration by moving ships. *Journal of Marine Science and Technology 6*, 2 (2001), 51–58. 3.2

[66] PEARSON, P. N., AND PALMER, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature 406*, 6797 (2000), 695–699. 1.1

[67] RAVEN, J., CALDEIRA, K., ELDERFIELD, H., HOEGH-GULDBERG, O., LISS, P., RIEBESELL, U., SHEPHERD, J., TURLEY, C., AND WATSON, A. *Ocean acidification due to increasing atmospheric carbon dioxide*. The Royal Society, 2005. 1.1

[68] RENKSIZBULUT, M., NAFZIGER, R., AND LI, X. A mass transfer correlation for droplet evaporation in high-temperature flows. *Chemical Engineering Science 46*, 9 (1991), 2351 – 2358, doi: 10.1016/0009-2509(91)85133-I. 3.2

[69] RINGROSE, P., MATHIESON, A., WRIGHT, I., SELAMA, F., HANSEN, O., BISSELL, R., SAOULA, N., AND MIDGLEY, J. The in salah {CO2} storage project: Lessons learned and knowledge transfer. *Energy Procedia 37*, 0 (2013), 6226 – 6236, doi: http://dx.doi.org/10.1016/j.egypro.2013.06.551. GHGT-11. 2

[70] ROTHLAUF, F. *Representations for genetic and evolutionary algorithms*. Springer, 2006. 4.2

[71] SCHONFELD, P., AND JONG, J.-C. *Intelligent road design*, vol. 19. WIT Press, 2006. 4.2.4

[72] SEBER, G. A. F., AND WILD, C. J. *Nonlinear Regression*. John Wiley & Sons, 1989. ISBN 0-471-47135-6. 4.1, 4.1.1, 4.1.1, 4.1.5

[73] SHITASHIMA, K., MAEDA, Y., AND OHSUMI, T. Development of detection and monitoring techniques of CO_2 leakage from seafloor in subseabed CO_2 storage. *Applied Geochemistry 30*, 0 (2013), 114 – 124, doi: http://dx.doi.org/10.1016/j.apgeochem.2012.08.001. Geochemical Aspects of Geologic Carbon Storage. 2.2

[74] STEINBERG, D., AND HUNTER, W. Experimental design: review and comment. *Technometrics 26* (1984), 71–97. 4.1.4

[75] STOCKER, T., QIN, D., PLATTNER, G., TIGNOR, M., ALLEN, S., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V., AND MIDGLEY, P. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change 2013. *Cambridge University Press, Cambridge, UK, and New York* (2013). 1.1

[76] SYLTE, A., EBELTOFT, E., GRIMSTAD, A., KULKARNI, R., NORDTVEDT, J., AND WATSON, A. Design of two-phase displacement experiments. *Inverse Problems in Engineering 10*, 1 (2002), 65–84. 4.1.5

[77] TENG, H., MASUTANI, S., KINOSHITA, C., AND NIHOUS, G. Solubility of CO_2 in the ocean and its effect on CO_2 dissolution. *Energy conversion and management 37*, 6 (1996), 1029–1038. 3.3

[78] THORKILDSEN, F., AND HAUGAN, P. Numerical model for plumes of dissolving CO_2 droplets in seawater. *The Nansen Environmental and Remote Sensing Center* (1993). 3.3

[79] TIKHONOV, A. N. *Numerical methods for the solution of ill-posed problems*, vol. 328. Springer, 1995. 4.1.1

[80] TORP, T. A., AND GALE, J. Demonstrating storage of CO_2 in geological reservoirs: The sleipner and SACS projects. *Energy 29*, 9–10 (2004), 1361 – 1369, doi: http://dx.doi.org/10.1016/j.energy.2004.03.104. 6th International Conference on Greenhouse Gas Control Technologies. 2

[81] WEGENER, G., SHOVITRI, M., KNITTEL, K., NIEMANN, H., HOVLAND, M., AND BOETIUS, A. Biogeochemical processes and microbial diversity of the Gullfaks and Tommeliten methane seeps (Northern North Sea). *Biogeosciences Discussions 5*, 1 (Feb. 2008), 971–1015. 2.2

Part II

Included Papers

A

Paper A

Assessing model parameter uncertainties for rising velocity of CO_2 droplets through experimental design

Hilde Kristine Hvidevold, Guttorm Alendal, Truls Johannessen and Trond Mannseth

International Journal of Greenhouse Gas Control, **11**, 283–289 (2012)

International Journal of Greenhouse Gas Control 11 (2012) 283–289

Contents lists available at SciVerse ScienceDirect

International Journal of Greenhouse Gas Control

journal homepage: www.elsevier.com/locate/ijggc

Assessing model parameter uncertainties for rising velocity of CO_2 droplets through experimental design

Hilde Kristine Hvidevold[a,*], Guttorm Alendal[a,b], Truls Johannessen[c,d], Trond Mannseth[e,a]

[a] University of Bergen, Department of Mathematics, Johannes Brunsgt 12, 5008 Bergen, Norway
[b] Uni Computing, Norway
[c] University of Bergen, Geophysical Institute, Norway
[d] Uni Bjerknes, Norway
[e] Uni CIPR, Norway

ARTICLE INFO

Article history:
Received 17 April 2012
Received in revised form 29 August 2012
Accepted 11 September 2012
Available online 4 October 2012

Keywords:
Geological storage
Leakage to marine environment
CO_2 droplets
Parameter uncertainties
Experimental design

ABSTRACT

The risk and environmental impact assessments required for geological CO_2 storage projects will have to rely on different types of numerical models, which will have to be calibrated and validated against measurements. Available measurements from ongoing demonstration projects are limited, hence it is necessary to turn to analog processes or laboratory experiments to estimate model parameters. In any case, parameter estimates will have uncertainties that will be important to assess when predicting future scenarios.

We study a model for the rise velocity of droplets in the ocean, an important process sub-model for simulating gas seeps into marine waters. As the origin we use the parameters estimation study by Bigalke et al. (2010) based on a tank experiment. We illustrate how Linearized Covariance Analysis (LCA) can be used to assess the parameter uncertainties, and how to design a similar experiment that reduces these uncertainties. The linearity assumption underlying LCA is assessed using curvature measures. It is shown that up to ~63% reduction in uncertainties is achieved by choosing the right droplet size distribution; by extending the range of droplet sizes to include larger droplets the uncertainties are reduced by another ~88%.

1. Introduction

Storage is at the end of the chain composing the Carbon Capture and Storage (CCS) technology (Pires et al., 2011), and regulations (Bäckstrand et al., 2011) and guidelines (Carpenter et al., 2011) for CO_2 geological storage projects are under development. Throughout the project period; from site selection, the operational phase and after closure, numerical simulation tools will be an integral part of the toolbox for the operator and governmental bodies. The range of models necessary will include porous media migration (Jiang, 2011), monitoring design (Seto and McRae, 2011; Yang et al., 2011), environmental risk assessment models (Koornneef et al., 2011) and economical models (Trabucchi et al., 2010). The range of spatial and temporal scales involved complicate the scope even further (Celia and Nordbotten, 2011).

All models have uncertainties in the model output. These uncertainties originates from different sources, such as uncertainties in input variables, numerical inaccuracy, and uncertainties in model parameters (Oberkampf and Roy, 2010). It is challenging to assess these uncertainties in model output and the challenge increases with model complexity. Hence, there is a need for data for model calibration, including parameter estimation, and validation for a large variety of models. The monitoring programs connected to the ongoing demonstration projects, for instance Sleipner, In-Salah and Snøvit (Eiken et al., 2011) provide invaluable data. They are, however, designed not to leak and do not provide data for models that simulates migration of CO_2 out of the storage reservoir, especially leaks to marine waters.

A CO_2 seep to marine waters will produce individual rising droplets if between ~3000 and 500 m, and bubbles if shallower than ~500 m (Brewer et al., 2002). A plume will be generated by entraining water if the flux of droplets/bubbles is high enough (Alendal and Drange, 2001). The dissolution of CO_2 content in the droplets/bubbles will gradually acidify the surrounding water masses, with potential environmental impacts (Haugan and Drange, 1992). The environmental impacts depends on how fast the droplets/bubbles dissolve and how fast the CO_2 concentration is diluted by local currents and mixing. The regional current conditions will determine whether the dissolved CO_2 is transported to the surface, and then exchange with the atmosphere, or to deeper

* Corresponding author.
E-mail address: HildeKristine.Hvidevold@math.uib.no (H.K. Hvidevold).

1750-5836/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijggc.2012.09.008

water, hence serving as a buffer for the retention time (Drange et al., 2001).

Modeling of CO_2 seeps to the water column contains two very important sub-models; the rise velocity of the droplets/bubbles and the mass-transfer rate (Alendal and Drange, 2001; Sato, 2004; Chen et al., 2005). We will focus on the rise velocity and do not discuss the mass-transfer rate further. Rise velocity models have been built by adopting general models given in the literature (Clift et al., 7695; Kelbaliyev, 2011). Available experimental and in situ measurements of CO_2 behavior in seawater for parameter estimation in these general models have so far been sparse. The in situ experiments performed off the Californian coast (Brewer et al., 2002) have been used (Gangstø et al., 2005) and has later been improved by the experiments performed by Bigalke et al. (2010, 2008).

Lately the focus has turned toward using natural CO_2 seeps, like in the Okinawa trough Shitashima et al. (2008), in the Mediterranean (Caramanna et al., 2011), and in the North Sea (Mcginnis et al., 2011), as a laboratory for gaining better understanding of processes involved. But how do we add to the existing data in a way that assures reduction in the model parameters uncertainties? In other words, how do we design new experiments in order to reduce the uncertainties in the model parameters?

For us the objective is to assess the uncertainties in the parameter estimation for a process model of CO_2 droplet rising velocity, using data from a pressurized tank experiment (Bigalke et al., 2008). We will use Linearized Covariance Analysis (LCA) (see, e.g., Seber and Wild, 1989) to analyze parameter uncertainties in the results of Bigalke et al. (2010), and to design a similar experiment that would result in reduced parameter uncertainties.

2. The rise velocity model

The rise velocity model used by Bigalke et al. (2010) are obtained by balancing buoyancy and a quadratic friction law, hence assuming fast relaxation time; the vertical rise velocity u_T of a droplet/bubble is then (Clift et al., 7695):

$$u_T = \sqrt{\frac{8gr(\rho_c - \rho_d)}{3C_d(u_T)\rho_c}}. \tag{1}$$

Here ρ_c and ρ_d is the density of seawater and CO_2, respectively, and g is the gravitational constant. To account for neglected forces, e.g. deformation effects, the radius r is replaced by the equivalent radius, r_e, which is the radius of a sphere having the equivalent volume as the droplet studied. The drag coefficient, C_d, account for neglected frictional forces.

The drag coefficient is usually split into the friction factor f_c and the deformation factor D, $C_d = f_c D$. A number of expressions for f_c and D exist. They are generally complex and dependent on several factors (Kelbaliyev, 2011; Clift et al., 7695). We will adopt the expressions used by Bigalke et al. (2010), where they did parameter estimation using data from the experiments by Bigalke et al. (2008). The experiments were conducted in a stainless steel tank with constant temperature and pressure conditions. Bubbles of CH_4 and droplets of CO_2 of different sizes were released and cameras were recording shape, size and velocity for each bubble or droplet. Bigalke et al. (2010) distinguish between droplet with and without hydrate coating, using different equations for the two regimes. In the following we will focus on CO_2 droplets without hydrate coating.

Under the assumption that bubbles attain shapes that minimize their energies, Bigalke et al. (2010) derived the following expression for the deformation factor:

$$D = \left[\frac{2}{(3.974\times10^{-3}(\mathrm{We}-12.62)^2 - 7.186\times10^{-4}(\mathrm{Eo}-17.87)^2 + 3.280\times10^{-5}\mathrm{EoWe}((\mathrm{Eo}-27.77)(\mathrm{We}-8.405)+67.08)+1.130)}\right]^2. \tag{2}$$

Here, $\mathrm{We} = 2r_e\rho_c u_T^2/\sigma$ and $\mathrm{Eo} = g(\rho_c - \rho_d)4r_e^2/\sigma$ are the Weber number and Eötvös number, respectively. Surface tension is denoted σ.

To estimate the friction factor, known asymptotic behavior was used as first approximation, and subsequent adjustments were added by parameter fitting. For particles without hydrate skin the calculations gave the friction factor as

$$f_c = \frac{48}{\mathrm{Re}} + 0.9\frac{0.75\mathrm{Eo}^2 + 0.6}{0.75\mathrm{Eo}^2 + 14.5} = \frac{48}{\mathrm{Re}} + \lambda\frac{\alpha\mathrm{Eo}^2 + \beta}{\gamma\mathrm{Eo}^2 + \omega}, \tag{3}$$

where $\mathrm{Re} = 2r_e\rho_c u_T/\mu$ is the Reynolds number, μ is the dynamic viscosity and λ, α, β, γ and ω are the parameters estimated by parameter fitting. However, the five parameters in Eq. (3) do not represent a minimal set, and can be reduced to three parameters, $\boldsymbol{\Theta}^{\mathrm{T}} = [\Theta_1, \Theta_2, \Theta_3]$, in the following way

$$f_c = \frac{48}{\mathrm{Re}} + \lambda\frac{\alpha}{\gamma}\frac{(\mathrm{Eo}^2 + (\beta/\alpha))}{(\mathrm{Eo}^2 + (\omega/\gamma))} = \frac{48}{\mathrm{Re}} + \Theta_1\frac{(\mathrm{Eo}^2 + \Theta_2)}{(\mathrm{Eo}^2 + \Theta_3)}.$$

We will consider the rise velocity as a function of the equivalent radius and the parameters, while the other variables, such as densities, are assumed to be constant and known. Hence, we neglect uncertainties in the measurements of bubble volume, and subsequent calculations of equivalent radiuses.

Both the friction factor $f_c = f_c(u_T, \Theta_i)$, Eq. (3), and the deformation factor $D = D(u_T)$, Eq. (2), are dependent on the terminal velocity, through the Reynolds number and the Weber number, respectively. Therefore, Eq. (1) may be expressed as $P(u_T) = 0$ where $P(u_T)$ is a polynomial of order eight. The roots of $P(u_T)$ can be written

$$u_T = u_T(r_e, \boldsymbol{\Theta}).$$

This equation can be used to estimate $\boldsymbol{\Theta}$ by comparing u_T for different r_e's to corresponding experimental observations. The parameter estimates, $\boldsymbol{\Theta}^*$, as well as the uncertainties in these estimates, will depend on the experimental conditions. The dependency of the parameter uncertainties on the experimental conditions allows for optimization of these conditions (here, the selected equivalent radiuses) aiming to minimize parameter uncertainties. This process is termed experimental design.

3. Assessing parameter uncertainties by linearized covariance analysis

An important part of a parameter estimation problem is to assess the uncertainty in the estimated parameters. The covariance matrix of the parameter estimation errors, $\boldsymbol{P}$, is useful for this purpose. For a linear model, or through linearization of a nonlinear model, one can obtain an informative explicit expression for $\boldsymbol{P}$, see Section 3.2. If the model is nonlinear the validity of the linearization will have to be checked, and if not valid, other methods, e.g. Monte-Carlo estimation, will have to be used.

After stating the parameter estimation problem, we briefly outline Linearized Covariance Analysis (LCA) and experimental design based on LCA. Finally, we discuss the assumptions underlying LCA and how to assess if these assumptions are valid. Most of the theory is adapted from Seber and Wild (1989) and Bates and Watts (1988) but, we will use notation introduced previously for the particular model considered in this study.

3.1. Parameter estimation problem

In a classical parameter estimation setting, it is assumed that there exist a true parameter set, $\hat{\Theta}$, and a physically correct model, u_T, such that the experimental measurement U_j may be expressed as

$$U_j = u_T(r_j, \hat{\Theta}) + \epsilon_j \quad j = 1, \ldots, n. \tag{4}$$

(For a Bayesian view of parameter estimation, see, e.g., Stuart (2010) and references therein.) The measurement errors, ϵ_j, are assumed to be independent and normally distributed with zero mean and a diagonal covariance matrix, $\boldsymbol{\Sigma}$, with diagonal elements $\Sigma_{j,j} = \vartheta_j^2$, where ϑ_j is the standard deviation of measurement j. Gathering the quantities U_j, $u_T(r_j, \hat{\Theta})$, and ϵ_j ; $j = 1, \ldots, n$, into n-vectors, $\boldsymbol{U}$, $\boldsymbol{u}(\hat{\Theta})$, and $\boldsymbol{\epsilon}$, the measurements of the nonlinear model studied can be written as

$$\boldsymbol{U} = \boldsymbol{u}(\hat{\Theta}) + \boldsymbol{\epsilon}.$$

Due to measurement errors, the aim is to find a set of parameters $\boldsymbol{\Theta}^*$ so that the match of $\boldsymbol{u}$ to $\boldsymbol{U}$ is appropriate for the given measurement error model. The most commonly used method to find such a set of parameters is to minimize a least squares functional,

$$S(\boldsymbol{\Theta}) = (\boldsymbol{U} - \boldsymbol{u}(\boldsymbol{\Theta}))^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{U} - \boldsymbol{u}(\boldsymbol{\Theta})) + R(\boldsymbol{\Theta}), \tag{5}$$

with respect to $\boldsymbol{\Theta}$, and terminate the minimization when the appropriate error level has been attained (Seber and Wild, 1989). Here, $R(\boldsymbol{\Theta})$ is a regularization term.

In this study, parameter estimation is not performed. Instead we simply adopt the parameter estimates obtained by Bigalke et al. (2010). Our interest lies in studying parameter uncertainties, and in particular, to utilize parameter uncertainties to perform experimental design on the experiment of Bigalke et al. (2008) underlying the parameter estimates obtained by Bigalke et al. (2010). Due to the low number of parameters, we further assume $R(\boldsymbol{\Theta}) = 0$.

3.2. Linearized Covariance Analysis

By definition, the covariance matrix of the parameter estimation errors is $\boldsymbol{P} = \mathrm{E}[(\boldsymbol{\Theta}^* - \hat{\Theta})(\boldsymbol{\Theta}^* - \hat{\Theta})^T]$, where E denotes statistical expectation. Inserting $\boldsymbol{u}(\boldsymbol{\Theta})$, linearized around $\hat{\Theta}$, into the equation $\partial S/\partial \boldsymbol{\Theta} = 0$, valid at an unconstrained minimum of S, leads to (Atkinson et al., 2007; Grimstad, 1999)

$$\boldsymbol{P} = (\boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{A})^{-1}. \tag{6}$$

Here $\boldsymbol{A}$ is the sensitivity matrix of the simulated values, with columns $\boldsymbol{a}_i = \partial \boldsymbol{u}/\partial \Theta_i$; $i = 1, 2, \ldots, p$. (p is the number of parameters.) Note that Eq. (6) is only valid if the sensitivity matrix $\boldsymbol{A}$ has full rank.

A confidence region is a region in parameter space that with a given probability level contains the true parameters. A $100(1-\alpha)\%$ confidence region around the estimate $\boldsymbol{\Theta}^*$ is given by (Atkinson et al., 2007; Grimstad, 1999)

$$(\boldsymbol{\Theta} - \boldsymbol{\Theta}^*) \boldsymbol{P}^{-1} (\boldsymbol{\Theta} - \boldsymbol{\Theta}^*) \leq \chi^2_{p,\alpha}, \tag{7}$$

where $\chi^2_{p,\alpha}$ is the upper α quantile for the χ^2 distribution with p degrees of freedom. If the model $\boldsymbol{u}(\boldsymbol{\Theta})$ is linear, the confidence region will be exact.

If the covariance matrix, $\boldsymbol{P}$, exists, it is symmetric and positive definite, and $\boldsymbol{P}^{-1}$ can be diagonalized as

$$\boldsymbol{P}^{-1} = \boldsymbol{W}^T \boldsymbol{\Lambda} \boldsymbol{W}, \tag{8}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of positive eigenvalues and the columns of $\boldsymbol{W}$ are orthonormal eigenvectors. The confidence region can be visualized as a p-dimensional ellipsoid where the ith major semi-axis direction is defined by $\boldsymbol{W}_{,i}$ and where the semi-axis length is given by $\sqrt{(\chi^2_{p,\alpha}/\boldsymbol{\Lambda}_{i,i})}$.

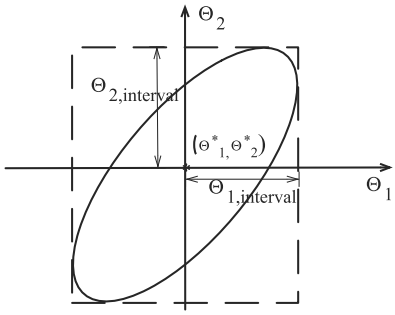


Fig. 1. A 2 dimensional illustration of a projection of the confidence region to obtain a $100(1-\alpha)\%$ joint confidence interval for the parameters. Note that the box include many points that lie outside of the $100(1-\alpha)\%$ confidence region.

By projecting the $100(1-\alpha)\%$ confidence region into the $(\Theta_1, \Theta_2, \ldots, \Theta_p)$ coordinate system we obtain a $100(1-\alpha)\%$ confidence interval for the parameters considered jointly. The value obtained by projecting the confidence region onto axes i will be denoted $\Theta_{i,\mathrm{interval}}$, see Fig. 1. The Cartesian product $\prod_{i=1}^{p}(-\Theta_{i,\mathrm{interval}}, \Theta_{i,\mathrm{interval}})$ gives a conservative estimate of the confidence region, as illustrated in Fig. 1.

3.3. Experimental design based on linearized covariance analysis

The LCA-based covariance matrix, $\boldsymbol{P} = (\boldsymbol{A}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{A})^{-1}$, depends on the measurement errors through $\boldsymbol{\Sigma}$, and on the experimental conditions (the selected equivalent radiuses, $\{r_j\}$) through $\boldsymbol{A}$. $\boldsymbol{P}$ will be viewed as a function of the r_j's only, and later we derive an estimate of the measurement errors based on the experimental setup.

The LCA based $\boldsymbol{P}$ is independent of the actual measurement values, and can therefore be calculated prior to the experiment for a given set of experimental conditions. This facilitates LCA-based experimental design, where the general aim is to minimize some scalar measure of $\boldsymbol{P}$.

There exists several different scalar measures based on different properties of $\boldsymbol{P}$, leading to different optimality criteria (Steinberg and Hunter, 1984). The rationale for the D-criterion, which is one of the most commonly used criteria in experimental design (Beck and Woodbury, 1998), is that the determinant of $\boldsymbol{P}$ is proportional to the volume of the confidence region (Box and Lucas, 1959). We will select optimal experimental conditions, $\boldsymbol{r}^* = (r_1^*, \ldots, r_n^*)^T$, utilizing the D-optimality criterion,

$$\boldsymbol{r}^* = \arg\min(\det \boldsymbol{P}(\boldsymbol{r})). \tag{9}$$

The genetic algorithm in the global optimization toolbox in MATLAB<*ce* : *sup*> ® </*ce* : *sup*> will be used to solve this minimization problem.

3.4. Assessment of LCA validity

Since the model we consider is nonlinear and LCA is based on model linearization, an assessment of the validity of the linearization is required. The current section briefly summarizes the most relevant features of some computational tools – curvature measures of nonlinearity (Bates and Watts, 1980) – that will be applied for this purpose. For a more in-depth description, see Seber and Wild (1989), Bates and Watts (1988, 1980).

Table 1
Pressure (P) and temperature (T) conditions (P/T), and corresponding liquid densities of pure liquid CO_2 droplets, ρ_d, and seawater density, ρ_c, r_{min} is the smallest droplet and r_{max} is the largest droplet in the experiments by Bigalke et al. (2008).

P/T	P [MPa]	T [°C]	ρ_d [kg m^{-3}]	ρ_c [kg m^{-3}]	r_{min} [mm]	r_{max} [mm]
1	24.8	13.1	989.5	1037.0	0.8	5.9
2	22.0	13.9	973.9	1036.3	0.7	5.2
3	18.3	13.1	958.7	1034.2	0.8	5.5

Basically, curvature measures of nonlinearity assess the deviation from linearity in a particular model. The curvature c_h in the direction $\boldsymbol{h}$ at a point is defined as the ratio of the length of the acceleration vector to the squared length of the tangent vector

$$c_h^{\Theta} = \frac{\|\ddot{\boldsymbol{u}}_h^{\Theta}\|}{\|\dot{\boldsymbol{u}}_h\|^2}, \tag{10}$$

$$c_h^{I} = \frac{\|\ddot{\boldsymbol{u}}_h^{I}\|}{\|\dot{\boldsymbol{u}}_h\|^2}. \tag{11}$$

Here, c_h^{Θ} denotes the parameter-effects curvature while c_h^{I} denotes the intrinsic curvature. The former depends on the selected model parameterization while the latter does not.

The tangent and acceleration vectors can be calculated as

$$\dot{\boldsymbol{u}}_{\boldsymbol{h}} = \sum_i \frac{\partial \boldsymbol{u}(\boldsymbol{\Theta})}{\partial \Theta_i}\bigg|_{\Theta^*} h_i = \boldsymbol{A}\boldsymbol{h}, \tag{12}$$

$$\ddot{\boldsymbol{u}}_{\boldsymbol{h}} = \sum_i \sum_j \frac{\partial^2 \boldsymbol{u}(\boldsymbol{\Theta})}{\partial \Theta_i \partial \Theta_j}\bigg|_{\Theta^*} h_i h_j = \boldsymbol{h}^{\boldsymbol{T}}\ddot{\boldsymbol{A}}\boldsymbol{h}, \tag{13}$$

respectively. For a single measurement, $\ddot{\boldsymbol{A}}$ is a $p \times p$ matrix, for n measurements $\ddot{\boldsymbol{A}}$ will be an $n \times p \times p$ array.

To make the curvatures measures invariant under scaling, c_h^{Θ} and c_h^{I} must be converted to standardized *relative curvatures*. This is done by normalizing the curvature with respect to the number of parameters used and the magnitude of the model function values. If the standard deviation of the measurements is assumed to be known a convenient scaling is to divide the output of the model function, and thus all derivatives thereof, by the appropriate standard deviation, resulting in $\gamma_{\boldsymbol{h}}^{\Theta}$ and $\gamma_{\boldsymbol{h}}^{I}$. Thereafter the *maximum* relative parameter effects curvature is defined as

$$\gamma^{\Theta} = \max_{\boldsymbol{h}} \gamma_{\boldsymbol{h}}^{\Theta}, \tag{14}$$

and the *maximum* relative intrinsic curvatures as

$$\gamma^{I} = \max_{\boldsymbol{h}} \gamma_{\boldsymbol{h}}^{I}. \tag{15}$$

The $\boldsymbol{h}$ that maximize relative curvature measures is found using the algorithm in Bates and Watts (1980).

To assess the impact of nonlinearity on a particular confidence region Bates and Watts (1988) suggested to compare the relative curvature measures with an appropriate cut-off value, $\gamma_c = 1/(\chi^2_{p,\alpha})^{1/2}$. According to their analysis, LCA-based confidence regions should not be trusted if

$$\frac{\gamma^{\Theta,I}}{\gamma_c} > 1.$$

However, this cut-off value does not represent a sharp division between acceptable and unacceptable nonlinearity. Different authors (Donaldson and Schnabel, 1987; Cook and Witmer, 1985; Sylte et al., 2002) have studied how well warnings about use of LCA obtained from curvature measures correspond with comparisons of LCA results to Monte-Carlo results. Results in Sylte et al. (2002) indicate that the cut-off value may be too strict for some types of problem, at least if one is content with LCA providing conservative results. Grimstad et al. (2001) elaborates further on this observation, and points out a connection to a link between low sensitivity and high nonlinearity for certain directions in parameter space for a large class of parameter estimation problems. Keeping these reservations in mind, curvature measures of nonlinearity will be carefully applied to check the validity of LCA in this paper.

4. Designing the droplet experiment

In Bigalke et al. (2008) they conducted experiments to measure the rise velocity for CO_2 droplets, and data sets were collected both within and outside the field of hydrate stability. For practical reasons this study only considers the equation for rise of a CO_2 droplet without a hydrate skin. Three experiments with different pressures and temperatures were conducted for CO_2 droplets without hydrate skin, see Table 1. In the conducted experiment the nozzle orifice for release of droplets was fixed, which led to a range of droplet radiuses between 0.7 and 5.9 mm.

It is assumed that the experiment is set up such that it is possible to release a droplet with a predefined equivalent radius, that is each r_j is considered as given exact. This is a simplification; the problem of achieving this in the laboratory is acknowledged. All other variables are assumed to be constant and known, and the estimated parameters values are assumed to be those found by Bigalke et al. (2010), $\boldsymbol{\Theta}^{\mathrm{T}} = [0.9, 0.8, 19.33]$. The number of measurements in each experiment is denoted n.

The droplet rise rates were determined optically by two cameras at fixed vertical locations both with a framerate 25 Hz, see Fig. 2. The picture where the droplet is closest to the center line is chosen as the time stamp. In these photos the droplets will be located with a uniform distribution within a distance $\pm u_T/50$ mm from the center line. Hence, the standard deviation of the measurement error for one camera is $u_T/25\sqrt{12}$. Combining the two cameras, the standard deviation of the error of measurement j is $\vartheta_j = u_{Tj}/25\sqrt{6}$, which will be used in the covariance matrix for the measurement errors $\boldsymbol{\Sigma}$ in Eq. (6).

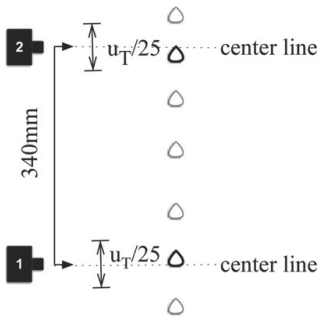


Fig. 2. The two cameras are placed in a distance 340 mm from each other. Since the framerate of each camera is 25 Hz, the droplet can move a distance $u_T/25$ between each shot and the droplet closest to the camera (bold droplets) would be spotted within a distance $\pm u_T/50$ mm from the center line.

Table 2
Values associated to the LCA 95% confidence region for the conducted and optimized experiments. The relative error of each parameter, $\Theta_{\text{rel},i}$, and the confidence regions for the conducted experiment, V_C, and the optimized, V_O, is shown together with the improvement in percent $I_{O,C} = (1 - (V_O/V_C)) \cdot 100$. P/T denotes the pressure and density condition (Table 1) and n is the number of measurements in the experiment.

P/T	n	$\Theta_{\text{rel},1}$	$\Theta_{\text{rel},2}$	$\Theta_{\text{rel},3}$	V_C
Conducted					
1	89	0.50	0.21	0.63	0.0594
2	45	0.52	0.26	0.69	0.1063
3	81	0.26	0.18	0.37	0.0288

P/T	n	$\Theta_{\text{rel},1}$	$\Theta_{\text{rel},2}$	$\Theta_{\text{rel},3}$	V_O	$I_{O,C}$
Optimized						
1	89	0.27	0.17	0.38	0.0221	63
2	45	0.36	0.23	0.50	0.0587	45
3	81	0.14	0.16	0.23	0.0139	52

4.1. Evaluating and reducing uncertainties in the conducted experiment

To compare the conducted experiment with an experiment with optimized experimental setting, the 95% confidence region (Eq. (7)) has been used. In the tables we present a relative improvement in percent, $I_{k,l} = (1 - \frac{V_k}{V_l}) \cdot 100$, where V_k and V_l are two confidence regions. In addition to confidence regions we will also consider an approximate measure for the relative uncertainty in individual parameters, by normalizing $\Theta_{\text{i,interval}}$ (Eq. (1)),

$$\Theta_{\text{rel},i} = \frac{\Theta_{i,\text{interval}}}{\Theta_i}, \quad i = 1, 2, 3.$$

In Table 2, a 95% confidence region for the experiment conducted by Bigalke et al. (2010) is shown together with the 95% confidence region obtained using optimized experimental settings. The same number of measurements and the same range of radiuses have been applied for the two experimental settings.

Optimizing the experiment lead to improvements of between 45 and 63 percent, for the confidence regions considered. For all the experiments in Table 2 the relative error for each parameter is below one. We notice for instance that the second parameter, Θ_2, is the least uncertain while parameter three, Θ_3, is the most uncertain. In the following we will focus only on the confidence regions.

Fig. 3 shows the placement of the radiuses in the different experiments discussed above. The optimized results indicate that the smallest radiuses have little influence on the confidence region, and it seems important to do measurements for larger values.

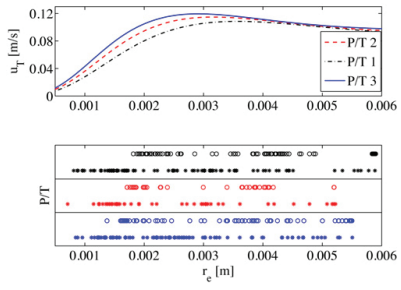


Fig. 3. The uppermost figure shows the rise velocity for the three different pressure and temperature conditions. The lowermost figure shows placement of measurements for conducted experiment, stars, and optimized experiment, circles.

Table 3
The table show a comparison of the confidence region for optimized experiment with large (V_L) and small (V_O) measurement region.

P/T	n	V_L	V_O	$I_{L,O}$
1	89	0.0032	0.0221	86
2	45	0.0073	0.0587	88
3	81	0.0028	0.0139	80

Table 4
The 95% confidence region of experiments with equally spaced measurements, $\hat{V}_E$ is compared to the confidence region obtained using optimal experimental settings, $\hat{V}_O$, for P/T 3 with $r \in [0.1, 12]$ mm.

n	$\hat{V}_O$	$\hat{V}_E$
40	0.0081	0.0140
60	0.0044	0.0076
80	0.0028	0.0049
100	0.0021	0.0035

In addition there is a tendency of high concentration of measurements around the inflection points of the rise velocity.

We now investigate the effect on the confidence region of allowing a larger radius interval, $r \in [0.1–12]$ mm. Although a change in the interval of r could lead to changes in the values of the estimated parameters, we will assume that the estimated parameters values are the same as above.

In Table 3 we study how a widening of the measurement range will influence the confidence region. V_O is the confidence region of the optimized experiment above, see Table 2, and V_L is the confidence region of the optimized experiment when the radiuses is between 0.1 mm and 12 mm. The table shows that V_L is between 80 and 88 percent smaller than V_O. Note that all the improvement is because larger droplets are included; the optimization do not recommend droplets smaller than 1.4 mm.

To emphasize how an optimized experimental setting can decrease the number of measurements needed to achieve desirable results, we have used P/T 3 (Table 1) to compare optimized experiments with experiments where the radiuses are equally spaced over the interval, see Table 4. As an example, an optimized experiment with 60 and 80 measurements results in a smaller confidence regions than equidistant experiments with 80 and 100 measurements, respectively. Fig. 5 illustrates how an increase in number of measurements will decrease the confidence region.

From Fig. 4 we note that the experiment with 40 measurements have three clusters around the inflection points in the rise velocity. When the number of measurements increases these clusters

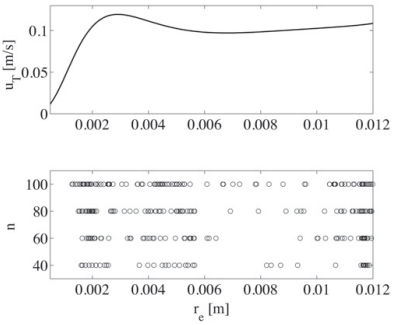


Fig. 4. The uppermost figure shows the velocity profile of P/T 3 with $r \in [0.1, 12]$ mm. The lower figure shows the placement of measurements for optimized experiments, using P/T 3.

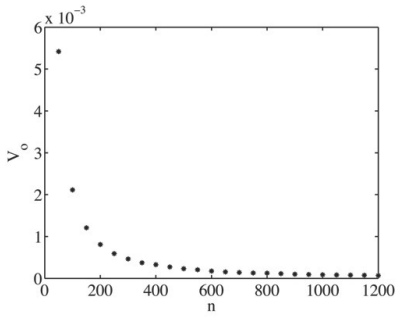


Fig. 5. The confidence region, for an optimized experiment with P/T 3 and $r \in [0.1, 12]$ mm, plotted against the number of measurements. The number of measurements increases with 50 in each step.

widens, probably because important regions have been sufficiently covered.

4.2. Validation of the linearization assumption

Before accepting the results, the validity of the linearity assumption must be assessed.

Table 5 shows scaled parameter and intrinsic curvature measures, that is $\gamma^{\Theta,I}/\gamma_c$, for the conducted experiments (Bigalke et al., 2008) and the optimized experiments for both narrow and wide radius ranges, corresponding to Tables 2 and 3. For both experiments, the intrinsic curvature is below the critical value, 1. However, the parameter-effect curvature is above the critical value for the experiments with $r \in [0.7, 5.9]$ mm, which indicates that these LCA results might not be accurate. (Note that this situation motivates a reparameterization, since a reparameterization can only change the parameter-effect curvature. Unfortunately, it is not clear how to select a reparameterization that will reduce γ^{Θ} (Bates and Watts, 1988), and we will make no such attempt here.) When the droplet radius range is increased, however, also the parameter-effect curvatures are below one.

Table 6 shows relative curvatures for the optimized experiment with $r \in [0.1–12]$ mm, see also Table 4. Here, both the intrinsic and parameter-effect curvature are clearly below the critical value.

Table 5
Scaled maximum parameter effect (P) and intrinsic (I) curvature for conducted and optimized experiments.

P/T	n	Conducted $r \in [0.7, 5.9]$		Optimized $r \in [0.7, 5.9]$		Optimized $r \in [0.1, 12]$	
		P	I	P	I	P	I
1	89	7.69	0.25	3.33	0.07	0.32	0.09
2	45	6.56	0.23	3.86	0.08	0.36	0.13
3	81	2.92	0.20	1.24	0.09	0.24	0.09

Table 6
Scaled maximum parameter effect (P) and intrinsic (I) curvature for optimized experiments for P/T 3, the first three columns uses all measurements from 0.1 mm to 12 mm, the last three columns show how many of the measurement which fell in the range between 0.8 mm and 5.5 mm and the curvature belonging to these measurements.

n	P large	I	n	P small	I
40	0.34	0.12	23	4.44	0.26
60	0.28	0.11	33	3.16	0.15
80	0.24	0.09	49	2.80	0.16
100	0.22	0.08	62	2.86	0.16

Curvature-measure values therefore indicate that LCA results in Table 4 are accurate, and that the experimental setup shown in the lower plot in Fig. 4 could serve as basis for new experiments. Since the curvature measures are below one for the optimized experiment, it is interesting to investigate if the placement of the radiuses in the conducted measurement region is crucial for the linearity measure. Therefore the curvature using only those measurements from the optimized experiment that fell within this region is included in Table 6. The table shows that the parameter effect curvature is above one if you isolate the measurements between 0.8 mm and 5.5 mm, showing that the placement of the radiuses within the conducted measurement region is not important for the validity of LCA.

Conclusions about preferential experimental setups based on the LCA results where the associated parameter-effect curvature values are above the critical value are very similar to those drawn from the LCA results where the associated parameter-effect curvature values are clearly below the critical value. This increases the belief in the accuracy of the former LCA results.

5. Discussion

Model calibration and validation are vital parts of model quality assessments. The type of data required differs; validation requires input and boundary conditions for the scenarios being validated and the model output is compared with measured data. Data used for calibration will have the purpose to estimate internal parameters in the model, which are often empirical equations, as accurate as possible. In both of these model quality activities the modelers have to depend on measurements, either from laboratory experiments or in situ.

Both measurements and models have uncertainties and errors that are important to assess. As illustrated here, both the measurement errors and the experimental setup will influence directly on the uncertainties in the model parameters. However, with a proper design of experiments it is possible to assess these uncertainties and reduce them. Hence, it is important that experimentalists and modelers communicate and work together when planning data gathering. This is a win-win situation; the parameter uncertainties can be reduced and the cost of the data gathering can be lowered through demand for less data for the same accuracy.

The parameter uncertainties will propagate through the model and impose uncertainties in the model output. If the model system is complex the uncertainty propagation can be hard to analyze, often statistics have to be built through e.g. Monte Carlo methods. Such models are often computationally demanding, making studies of uncertainty propagation time consuming and often impossible.

The model for droplet rise velocity studied here is extremely simple and is a sub-model within a more complex model simulating CO_2 seeps to, and transport in, marine waters. This seep model will have to work in assembly with other models in order to assess environmental impact imposed by a leak from a CO_2 geological storage site. For instance, it will have to rely on a porous media model simulating CO_2 transport through sediments as input, and the output from the model will serve as input to ecosystem models.

Koornneef et al. (2011) found that the non-engineering part of the CO_2 storage has the largest gaps in environmental information and understanding, claiming that the validation of models for CO_2 transport is at an infant stage. Due to this it is not possible to perform a confident risk assessment of storage projects. To close this gap it is essential that modelers and experimentalists work together in establishing and designing new experiments and measurement campaigns. The aim of these activities will be to assemble invaluable data for model calibration and for model validation.

With the many different kind of models needed to study impact of a leak; there is a need for many different experiments and campaigns. The ongoing and planned CO_2 storage demonstration projects, with proper monitoring program, will have a role. But they cannot serve the need alone.

Acknowledgements

We would like to thank Nikolaus Bigalke for giving us access to the experimental data. This work has been funded by SUCCESS centre for CO_2 storage under grant 193825/S60 from Research Council of Norway (RCN). In addition, the research leading to these results has received funding from the European Union Seventh Framework Program (FP7/2007-2-13) under grant agreement no. 265847.

References

Alendal, G., Drange, H., 2001. Two-phase, near-field modeling of purposefully released CO_2 in the ocean. Journal of Geophysical Research 106, 1085–1096.

Atkinson, A., Donev, A., Tobias, R., 2007. Optimum Experimental Designs, with SAS. Oxford University Press, New York, ISBN 9780199296590.

Bäckstrand, K., Meadowcroft, J., Oppenheimer, M., 2011. The politics and policy of carbon capture and storage: framing an emergent technology. Global Environmental Change 21, 275–281.

Bates, D., Watts, D., 1980. Relative curvature measures of nonlinearity. Journal of the Royal Statistical Society, Series B: Methodological 42, 1–25.

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and its Applications. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc, ISBN 0-471-81643-4.

Beck, J., Woodbury, K., 1998. Inverse problems and parameter estimation: integration of measurements and analysis. Measurement Science and Technology 9, 839–847.

Bigalke, N., Enstad, L., Rehder, G., Alendal, G., 2010. Terminal velocities of pure and hydrate coated CO_2 droplets and CH_4 bubbles rising in a simulated oceanic environment. Deep Sea Research Part I: Oceanographic Research Papers 57, 1102–1110.

Bigalke, N., Rehder, G., Gust, G., 2008. Experimental investigation of the rising behavior of CO_2 droplets in seawater under hydrate-forming conditions. Environmental Science & Technology 42, 5241–5246.

Box, G., Lucas, H., 1959. Design of experiments in non-linear situations. Biometrika 46, 77–90.

Brewer, P., Peltzer, E., Friederich, G., Rehder, G., 2002. Experimental determination of the fate of rising CO_2 droplets in seawater. Environmental Science & Technology 36, 5441–5446.

Caramanna, G., Voltattorni, N., Maroto-Valer, M., 2011. Eight years of research on a marine natural analogue for sub-seabed CO_2 storage seepage. In: OCEANS, 2011 IEEE-Spain, IEEE, pp. 1–7.

Carpenter, M., Kvien, K., Aarnes, J., 2011. The CO2QUALSTORE guideline for selection, characterization and qualification of sites and projects for geological storage of CO_2. International Journal of Greenhouse Gas Control 5, 942–951.

Celia, M., Nordbotten, J., 2011. How simple can we make models for CO_2 injection, migration, and leakage? Energy Procedia 4, 3857–3864.

Chen, B., Song, Y., Nishio, M., Someya, S., Akai, M., 2005. Modeling near-field dispersion from direct injection of carbon dioxide into the ocean. Journal of Geophysical Research 110, C09S15.

Clift, R., Grace, J.R., Weber, M.E., 1978. Bubbles, Drops, and Particles. Academic Press, New York, ISBN 0-12-176950-x.

Cook, R., Witmer, J., 1985. A note on parameter-effects curvature. Journal of the American Statistical Association 80, 872–878.

Donaldson, J., Schnabel, R., 1987. Computational experience with confidence regions and confidence intervals for nonlinear least squares. Technometrics 29, 67–82.

Drange, H., Alendal, G., Johannessen, O., 2001. Ocean release of fossil fuel CO_2: a case study. Geophysical Research Letters 28, 2637–2640.

Eiken, O., Ringrose, P., Hermanrud, C., Nazarian, B., Torp, T., Høier, L., 2011. Lessons learned from 14 years of CCS operations: Sleipner, in salah and snøhvit. Energy Procedia 4, 5541–5548.

Gangstø, R., Haugan, P., Alendal, G., 2005. Parameterization of drag and dissolution of rising CO_2 drops in seawater. Geophysical Research Letters 32, L10612.

Grimstad, A., Kolltveit, K., Mannseth, T., Nordtvedt, J., 2001. Assessing the validity of a linearized accuracy measure for a nonlinear parameter estimation problem. Inverse Problems 17, 1373–1390.

Grimstad, A.A., 1999. On nonlinearities and covariance analysis for the reservoir fluid flow model. Ph.D. Thesis. University of Bergen.

Haugan, P., Drange, H., 1992. Sequestration of CO_2 in the deep ocean by shallow injection. Nature 357, 318–320.

Jiang, X., 2011. A review of physical modelling and numerical simulation of long-term geological storage of CO_2. Applied Energy 88, 3557–3566.

Kelbaliyev, G., 2011. Drag coefficients of variously shaped solid particles, drops, and bubbles. Theoretical Foundations of Chemical Engineering 45, 248–266.

Koornneef, J., Ramírez, A., Turkenburg, W., Faaij, A., 2011. The environmental impact and risk assessment of CO_2 capture, transport and storage—an evaluation of the knowledge base. Progress in Energy and Combustion Science 38, 62–86.

Mcginnis, D.F., Schmidt, M., Delsontro, T., Themann, S., Rovelli, L., Reitz, A., Linke, P., 2011. Discovery of a natural CO_2 seep in the German North Sea: implications for shallow dissolved gas and seep detection. Journal of Geophysical Research 116, C03013.

Oberkampf, W., Roy, C., 2010. Verification and Validation in Scientific Computing. Cambridge University Press, New York, ISBN 978-0-521-11360-1.

Pires, J., Martins, F., Alvim-Ferraz, M., Sim oes, M., 2011. Recent developments on carbon capture and storage: an overview. Chemical Engineering Research and Design 89, 1446–1460.

Sato, T., 2004. Numerical simulation of biological impact caused by direct injection of carbon dioxide in the ocean. Journal of Oceanography 60, 807–816.

Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression. Wiley Series in Propability Statistics, John Wiley & Sons, Inc, ISBN 0-471-47135-6.

Seto, C., McRae, G., 2011. Reducing risk in basin scale CO_2 sequestration: a framework for integrated monitoring design. Environmental Science & Technology 45, 845–859.

Shitashima, K., Maeda, Y., Koike, Y., Ohsumi, T., 2008. Natural analogue of the rise and dissolution of liquid CO_2 in the ocean. International Journal of Greenhouse Gas Control 2, 95–104.

Steinberg, D., Hunter, W., 1984. Experimental design: review and comment. Technometrics 26, 71–97.

Stuart, A., 2010. Inverse problems: a Bayesian perspective. Acta Numerica 19, 451–559.

Sylte, A., Ebeltoft, E., Grimstad, A., Kulkarni, R., Nordtvedt, J., Watson, A., 2002. Design of two-phase displacement experiments. Inverse Problems in Engineering 10, 65–84.

Trabucchi, C., Donlan, M., Wade, S., 2010. A multi-disciplinary framework to monetize financial consequences arising from CCS projects and motivate effective financial responsibility. International Journal of Greenhouse Gas Control 4, 388–395.

Yang, Y., Small, M., Ogretim, E., Gray, D., Bromhal, G., Strazisar, B., Wells, A., 2011. Probabilistic design of a near-surface CO_2 leak detection system. Environmental Science & Technology 45, 6380–6387.