Advanced Inverse Modeling of Sediment Thermal Diffusion Processes

Reconstructing Temporal Variant Boundary Conditions for the One-Dimensional Heat Equation

submitted by
Frederieke Miesner, M. Sc.

in fulfillment of the requirements
for the degree of Doctor rerum naturalium (Dr. rer. nat.)

Center for Industrial Mathematics
University of Bremen
April 2018

Thesis advisor: Prof. Dr. Armin Lechleiter

First reviewer: Prof. Dr. Emily King Second reviewer: Dr. Tapio Helin

Date of the colloquium: 5 July 2018

Abstract

Temperature in marine sediments is driven by the bottom water temperature deviation and the geothermal heat flow from below. The partial differential equations describing this setting are well established, mathematically analyzed and broadly used in geophysical modeling. We present a parametrization that focuses on one-year periods in the bottom water temperature function and results in a low dimensional but non-linear forward model. We therefore cannot build on the inversion schemes used in climate reconstruction but need to introduce regularization methods.

In this work, we present three Newton-like and two Markov Chain Monte Carlo methods, as well as introduce a new heuristic linear fitting approach. All methods are tested on artificial data as well as three measured data sets. The measured data sets build the basis for three example areas with different characteristics in the thermal properties of the sediment, the temperature range of the bottom water, and the absolute value of the geothermal gradient. The data sets contain up to 22 temperature and thermal diffusivity measurements at a sediment depth of up to 6 m. We approximate the bottom water temperature deviation with a shifted cosine, whose parameters are the mean temperature, the amplitude, and the day of the annual minimum.

We conclude that the inversion of measured temperature- and thermal diffusivity-depth profiles for the parameters of the bottom water temperature and the geothermal gradient is – in general – possible. We observe that the mean temperature is generally approximated very well, while the day of the annual minimum cannot be reliably reconstructed better than ± 14 days. The accuracy of the reconstructed amplitude increases with decreasing relative difference to the mean temperature, and the accuracy of the geothermal gradient increases with decreasing amplitude and increasing measurement depth. While measurements deeper than 10 m are preferable, 6 m should be sufficient in most areas.

The methods all perform very well on artificial data, with up to 70 measurements at a depth of up to 10 m. For realistic data with only 22 measurements at a depth of up to 6 m, only the simple Newton method, the heuristic linear fitting, and the delayed rejection adaptive metropolis (DRAM) algorithm yield reliable results. The other methods could possibly be fine-tuned to fit the specific data sets, but that is not realistically usable for real data sets. Of these methods, the heuristic linear fitting has the best accuracy in reasonable computing time, while the DRAM algorithm has proven convergence for enlarging ensembles.

Acknowledgments

First of all, I wish to thank Prof. Dr. Armin Lechleiter. I am deeply grateful to have had him as my supervisor. I have grown – both mathematically and personal – in my time at the university, in his lectures, and in the workgroup. He encouraged me to pursue this path, and always asked the right questions and had the right answers.
He will be truly missed.

I want to thank Prof. Dr. Emily King, Prof. Dr. Samuli Siltanen, and Dr. Tapio Helin for their advices on form and content of this work in the last few months. I could not have finished this thesis without their short-term supervision and guidance.
Special thanks also to my colleagues Marcel, Florian, Stefan, Thies, and Alexander for the scientific exchange, the discussions, and the laughter, and to Ebba for keeping us all together and organized.
Dr. Christian Müller has been greatly supportive and gave helpful comments from the geophysical point of view. I also want to thank the colleagues from FIELAX for the great time, especially Dr. Regina Usbeck for the opportunity to operate the heatflow lance in the field, and Aline for showing me how to survive off-shore.
I would not be here without my family who enabled me to study whatever I wanted, and Raimund Albers who motivated me to study mathematics in the first place. Family and friends were always a great support and endured all the math-talk. I especially want to thank my lovely flat mates Anna and Johannes for their time, all the silly and deep discussions, and the home we shared. Lastly, I want to thank Erik with all my heart for his encouragement, the interested questions, the puns, and everything else.

Contents

1. Introduction **15**

I. The Forward Problem 19

2. Mathematical Introduction of the Forward Model Operator **21**

2.1. The Underlying Partial Differential Equation 21

2.2. Existence and Uniqueness 24

2.3. The Forward Operator 29

3. Temperature Measurements and Basic Regions **31**

3.1. Temperature Measurements in Marine Sediments 31

3.2. Example Data Sets 36

3.3. Characterization of the Geothermal Gradient 43

4. Numerical Implementation of the Forward Operator **45**

4.1. The Method of Lines 45

4.2. Choosing a Reasonable Mesh Size 48

5. Modeling the Bottom Water Temperature **51**

5.1. Different Bottom Water Temperature Functions 51

5.2. Analysis of Water Temperatures 54

5.3. The Measurement Operator and Artificial Data 58

II. The Inverse Problem 65

6. Theoretical Attributes of the Inverse Problem 67
6.1. A Short Introduction to the Theory of Inverse Problems 67
6.2. Eigenvalues of the Forward Operator 69

7. Newton-like Inversion Methods 73
7.1. Basic Newton Algorithm 75
7.2. A Statistical Newton Method 78

8. Linear Fitting 83
8.1. Linear Fitting with Tikhonov Regularization 83
8.2. Heuristic Linear Fitting with Day Reconstruction 90

9. Markov Chain Monte Carlo Methods 95
9.1. Monte Carlo Integration and Markov Chains 96
9.2. The Metropolis-Hastings Algorithm 97
9.3. Delayed Rejection and Adaptive Metropolis 99

10. Inversion Results 103
10.1. Results for Artificial Data 103
10.2. Results for the Basic Regions 109

III. Discussion 115

11. Generalization and Applicability 117

List of Figures

1. Schematic Diagram of Marine Sediment. 22

2. FIELAX HetaFlowProbe Device 33
3. Recorded Temperatures from the Thermistors 35
4. Example 1: Processed Data 38
5. Example 2: Processed Data 40
6. Example 3: Processed Data 42

7. Numerical Error in Dependence on the Lower Boundary 49
8. Numerical Error in Dependence on the Time Steps and Mesh Size 50
9. Numerical Error in Dependence on the Pre-Modeling Interval 50

10. Differences in the Sediment's Temperature in Dependence on A, B, and d 53
11. Bottom Water Temperature with Inflow 53
12. Water Temperature and Data for Different Phases 55
13. Error in the Data for Varying Second Phase 56
14. Water Temperatures, German Bight, 2011 57
15. Analysis of Given Water Temperatures 59
16. Temperature Field of Basic Region 1 over One Year. 62
17. Temperature Field of Basic Region 2 over One Year. 63
18. Temperature Field of Basic Region 3 over One Year. 64

19. Relative Errors for Varying B in Relation to A 77

20. Basis of Measurements for the shifted Cosine 85
21. Errors in Dependence on the Depth of the Measurement 88
22. Errors in Dependence on the Guessed Day of the Measurement 89

23. Discrepancies and Errors for Varying Regularization Parameter α 91
24. Linear Fitting: Reconstruction Error over Data Error 93

25. Results for Realistic Artificial Data for Example 1 105
26. Results for Realistic Artificial Data for Example 3 106
27. Example 1: Reconstruction Results 111
28. Example 2: Reconstruction Results 112
29. Example 2: Reconstruction Results 113

List of Tables

1. Characteristics of the Basic Regions 41
2. Thermal Lengths for the Basic Regions 43

3. Compare Inversion Results for Realistic Artificial Data with 0.5 % Noise . . 107
4. Compare Inversion Results for Realistic Artificial Data with 1 % Noise . . . 108
5. Compare Inversion Results for Real Data Sets 110

6. Newton Inversion Results: 72 sensors, 10 m depth, 0.5 % noise. 122
7. Newton Inversion Results: 72 sensors, 10 m depth, 1 % noise. 123
8. Newton Inversion Results: 22 sensors, 5.95 m depth, 0.5 % noise. 124
9. Newton Inversion Results: 22 sensors, 5.95 m depth, 1 % noise. 125
10. Newton Inversion Results: Measured Data 126
11. Statistical Newton Inversion Results: 72 sensors, 10 m depth, 1 % noise. . . 127
12. Statistical Newton Inversion Results: 72 sensors, 10 m depth, 1 % noise. . . 127
13. Statistical Newton Inversion Results: 22 sensors, 5.95 m depth, 0.5 % noise. 128
14. Statistical Newton Inversion Results: 22 sensors, 5.95 m depth, 1 % noise. . 128
15. Staistical Newton Inversion Results: Measured Data 129
16. Linear Fitting Inversion Results: 72 sensors, 10 m depth, 0.5 % noise. . . . 130
17. Linear Fitting Inversion Results: 72 sensors, 10 m depth, 1 % noise. 131
18. Linear Fitting Inversion Results: 22 sensors, 5.95 m depth, 0.5 % noise. . . 131
19. Linear Fitting Inversion Results: 22 sensors, 5.95 m depth, 1 % noise. . . . 132
20. Linear Fitting Inversion Results: Measured Data 132
21. MCMC Inversion Results: 72 sensors, 10 m depth, 0.5 % noise. 133
22. MCMC Inversion Results: 72 sensors, 10 m depth, 1 % noise. 134
23. MCMC Inversion Results: 22 sensors, 5.95 m depth, 0.5 % noise. 135

24. MCMC Inversion Results: 22 sensors, 5.95 m depth, 1 % noise. 136
25. MCMC Inversion Results: Measured Data 137

List of Algorithms

1. The Newton Algorithm 75
2. A Statistical Newton Algorithm. 80
3. Linear Fitting with Tikhonov Regularization 87
4. Heuristic Linear Fitting 90
5. The Metropolis-Hastings Algorithm 98
6. The DRAM Algorithm 100

1. Introduction and Geophysical Motivation

Over 70 % of the Earth's surface is covered with water, of which roughly 97 % is accumulated in the oceans. Temperature and sea level changes, flora and fauna of the oceans, and the possibilities of traveling over seas and rivers has influenced human life since ever. Thus, mankind has been exploring the oceans for a very long time.

Still, more is unknown than known about the dynamics of the oceans. One of the many topics which are not fully understood yet is thermodynamics. Behavior of ocean water temperatures is complex and driven by solar radiation, bathymetry, salinity, the Earth's rotation, global currents, and heat dissipation from plate tectonics. The water in the oceans forms so-called *water masses*, which are defined by their temperature and salinity. For an introduction to water masses we refer to Mamayev (1975). Due to the specific density, depending on salinity and temperature, the water masses typically do not mix without outer forces like strong currents or heavy storms.

To understand the dynamics of the whole system it is necessary to understand small subsystems first. It is easier to accurately describe a small subsystem by mathematical equations. Additionally, the ability to measure all of the involved factors is limited, which is less essential the less factors are involved in a subsystem.

In this study we focus on the bottom water temperature in areas where no strong currents dominate the temperature evolution of the sea body and where no advective heat flow occurs in the sediment. Such areas can be found e.g. in the German North and Baltic Sea and in deep sea basins. We will focus on the one-year period of the bottom water temperature deviation and neglect all long-term trends.

Knowledge about the temperature evolution in small sea basins, such as the German North

and Baltic Sea, are of interest mainly for oceanological questions: Which temperature ranges are vital for benthic communities (Crespo et al., 2017; Gray and Christie, 1983; Hancke et al., 2014)? How do cargo shipping, oil platforms, and cable routes to offshore wind parks influence the temperature evolution (Merck and Wasserthal, 2009; Zucco et al., 2006)? Additionally, such knowledge can be used to determine the geothermal heat flow from temperature measurements in the seabed. The geothermal heat flow is the amount of energy that dissipates from the Earth's crust (Stein, 1995). In the oceanic lithosphere it originates solely from the cooling of the convection system in the mantle (Sclater et al., 2014). Thus, the heat flow is a measure for the age of the crust. In the continental lithosphere, additional heat comes from radioactive decay (Stein, 1995). The measurable geothermal heat flow can also be influenced by topographic variations, advective processes, phase changes of gas hydrates, sedimentation rates and anisotropy of the thermal properties (Stein, 1995).

Thermistors as measurement devices for (water-) temperatures are cheap and accurate, while the ability to monitor water temperatures over a longer period of time is limited. In shallow seas, the cargo traffic necessitates extra effort in securing measurement systems to a specific location; in deep basins the communication with the device is problematic. Launching and retrieving for data transfer or maintenance is only possible with an adequate vessel and therefore expensive.

Thus, instead of installing a monitoring device for a period of time, we use Fourier's Law and the heat equation and determine the water temperature evolution mathematically from one single temperature-depth-profile. The advantages of one single measurement over steady monitoring are clear: it lowers the costs as the vessel has to be rented only once and it provides quick response on the success rate. Additionally, on many surveys the temperature-depth-profile is measured anyway as a by-product of thermal diffusivity and thermal conductivity measurements.

In this work, we present a rigorous analysis of the forward model, i.e. calculating the sediments' temperature from water temperatures, derived from Fourier's law. In Section 2.1 we will introduce the governing equations and obtain the forward operator. While the forward operator is well defined for a large class of bottom water temperature functions, we will focus on one-year periodic functions. For piecewise constant functions this forward operator is widely used in climate modelling and ground temperature reconstruction for geological

timescales from deep boreholes (Clow, 1992; Pollack et al., 1996; Shen and Beck, 1991). This leads to a high-dimensional but linear forward operator and results in inversion findings with broad temporal resolution over aforementioned timescales. Our forward operator only covers the last year, but has a higher temporal resolution. In fact, the parameterization of the bottom water temperature function, which introduces the non-linear terms into the forward operator, yields continuous reconstruction of the bottom water temperature.
In Section 5.2 we introduce the functions used for the bottom water temperature, which assumptions need to be made, and how realistic they are. It will be shown in Section 3.2 which information is stored in the sediment's temperature field and if they are significantly measurable. Here, we will also introduce three measurements and the prevailing oceanographic settings in the chosen example areas.
In Section 4.1 we will introduce the numerical methods to calculate solutions to the forward operator and discuss the approximation qualities.
In Section 6 we finally introduce the inverse problem, i.e. calculating the water temperature from the sediments' temperature, and discuss existence and uniqueness of a bottom water temperature function to given (i.e. measured) temperature-depth-profiles.
We then introduce different numerical approaches to solve the inverse problem under certain assumptions in Sections 7 – 9. The methods will be analyzed and tested on (realistic) artificial data sets and real measured data from the example regions with their different characteristics. We use classical inversion schemes, e.g. the Newton method, Tikhonov regularization, and Markov Chain Monte Carlo methods and also introduce a new heuristic scheme to obtain information on the non-linear parameter of the forward operator from a linear fitting method.
Lastly in Section 11, we will discuss the potential of application of the shown methods to the presented data sets in particular and to measured data in general. We conclude with an overview on the capabilities and disadvantages of the presented methods and give an outlook on possible further research.

Part I.

The Forward Problem

2. Mathematical Introduction of the Forward Model Operator

2.1. The Underlying Partial Differential Equation

In order to determine bottom water temperatures from sediment temperatures we need to connect the two mathematically. We consider the sediment to be horizontally laminated with homogeneous layers and no intrusions. The lamination occurs naturally as the sediment builds up over time. We also restrict our model to regions without advective material or fluid transport and heat sources (or sinks).

With these assumptions a three-dimensional cube of sediment can mathematically be described as a one-dimensional interval of finite depth. This setup is illustrated in Fig. 1.

Accordingly, the temperature distribution is influenced solely by the thermal properties of the sediment, the water temperature at the sediment-water-interface, and the heat flow from the Earth's crust.

The thermal properties, i.e. the thermal diffusivity κ as well as the thermal conductivity λ and the volumetric heat capacity ρc_v, are measured at the same time as the sediment's temperature (see Section 3.2) and hence can be considered as *known* parameters in this model. The geothermal gradient may also be determined from the measurements or is known from literature. However, the determination of the geothermal gradient from depth-dependent temperature measurements is not always possible for measurements shallower than 10 m (Sclater et al., 2014). Thus, we will also try to reconstruct the geothermal gradient g together with the bottom water temperature function $W(t) \in L^2([0, T])$.

As stated before, the sediment is modeled as a one-dimensional region $[0, x_E]$ with finite depth and the time is given as an interval $[0, T]$, where $t = 0$ marks 1 January at timestamp 00:00.

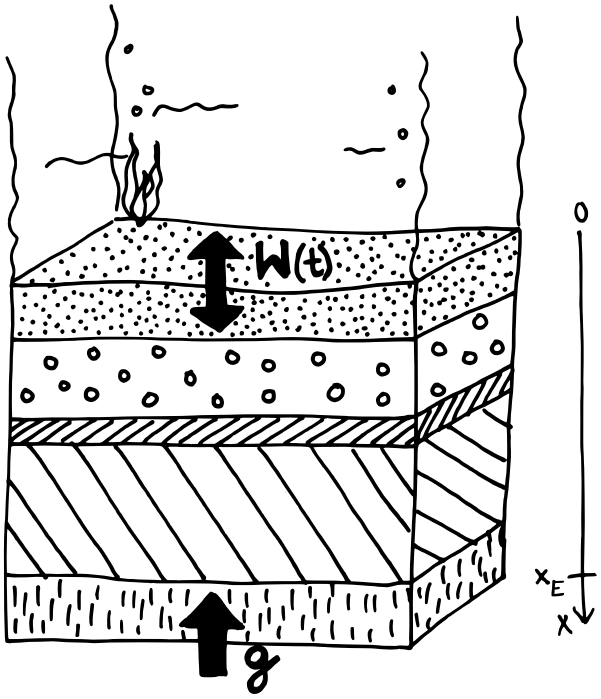

Figure 1.: Schematic diagram of marine sediment. The sediment is laminated such that each layer is horizontally homogeneous. Such a finite cube of sediment exchanges no heat horizontally with the surrounding sediment, but with the sediment below, represented by the geothermal gradient g. Additionally, there is heat flow over the sediment-water interface; the water temperature is denoted with $W(t)$.

Now, for the temperature field $u(x,t)$, the heat equation $\partial_t u(x,t) - \partial_x\left(\kappa(x)\partial_x u(x,t)\right) - v(x,t)u(x,t) = H(x,t)$ holds for every $x \in [0, x_E]$ and $t \in [0, T]$. Here, the velocity field $v(x,t)$ accounts for advective heat transport, whereas $H(x,t)$ represents inner sources or sinks. As we will not cover both phenomena in this work, these two terms are set to zero. Note that neither the bottom water temperature function nor the geothermal heat flow are involved in the heat equation. The influence of both terms is given by the boundary conditions at $x = 0$ and $x = x_E$. At the sediment-water-interface, $x = 0$, we want the condition to reflect that the heat exchange between the sediment and the water is proportional to the temperature difference. Thus, we formulate a Robin boundary condition $\kappa(0)\partial_x u(0,t) = \tilde{h}\left(u(0,t) - W(t)\right)$, where we assume that the heat transfer coefficient $\tilde{h}$ is known. At the lower boundary $x = x_E$, the geothermal gradient g prevails, such that $\partial_x u(x_E, t) = g$.
Putting this together, we get the following boundary value problem:

$$\partial_t u(x,t) - \partial_x\left(\kappa(x)\partial_x u(x,t)\right) = 0 \qquad \forall x \in [0, x_E], \forall t \in [0, T], \tag{2.1}$$

$$\kappa(0)\partial_x u(0,t) = \tilde{h}\left(u(0,t) - W(t)\right) \qquad \forall t \in [0, T], \tag{2.2}$$

$$\kappa(x_E)\partial_x u(x_E,t) = \kappa(x_E)\, g \qquad \forall t \in [0, T]. \tag{2.3}$$

Definition 2.1.1. *The Temperature Model.*
For given κ, g, x_E, $\tilde{h}$ and $W(t)$ we refer to Equations (2.1) – (2.3) *as the* temperature model.

The influence of the heat transfer coefficient $\tilde{h}$ is broadly discussed by Brakelmann and Stammen (2006). In areas with changing water coverage, i.e. mudflat, the coefficient is time-dependent with a period of up to 6 hours. However, the influence is very small and the use of an average value of $\tilde{h} = 150\,\mathrm{W/(m^2\,K)}$ is proposed. They also discuss the influence of the Robin boundary condition all together and rewrite the boundary condition at the sediment-water interface as an inhomogeneous Dirichlet condition

$$u(0,t) = W(t) \qquad \forall t \in [0, T]. \tag{2.4}$$

This simplifies the formulation of the analytic solution for constant thermal properties.
Next, we will show that the temperature model as in Definition 2.1.1 with initial conditions

$u_0 = u(\cdot, 0)$ has a unique solution in $H^1([0, x_E])$ for every piece-wise continuous function $W(t)$ and then formulate the definition of the *forward model*.

2.2. Existence and Uniqueness of a Solution to the PDE Formulation

Following the standard procedures for solving partial differential equations (PDE), see Evans (2010), we are going to show existence and uniqueness of a solution to the temperature model defined in Definition 2.1.1 in this section.

A Classical Solution

Our model described in Definition 2.1.1 has a classical solution when the thermal diffusivity is constant and the bottom water temperature function has certain regularity.
As we are looking to reconstruct the bottom water temperature for the preceding year, let $W(t) = A + B\cos(\omega t + \varphi(d))$, where $\omega = \frac{2\pi}{365}$, and $\varphi(d) = \omega(\frac{365}{2} - d)$. This shifted cosine is one-year periodic with mean temperature A and amplitude B.

Theorem 2.2.1. *A classical solution (Lowrie, 2007).*
Let, $W(t)$ be defined as above, $\kappa, g > 0$ and $a(\kappa) = \sqrt{\frac{\omega}{2\kappa}}$. The classical solution to the temperature model from Definition 2.1.1 is then given by

$$u(x, t) = A + B\exp(-a(\kappa)x)\cos(\omega t - a(\kappa)x + \varphi(d)) + g\,x, \tag{2.5}$$

This solution can be easily found by separation of variables and variation of constants (Lowrie, 2007).
Such classical solutions are used in geothermal modeling with greater time-scales or larger depths, where the influence of the variations in the thermal diffusivity is small. A comprehensive introduction can be found in (Carslaw and Jaeger, 1959) or in (Evans, 2010) for the mathematical point of view. However, as marine sediments are laminated, the thermal diffusivity can be best approximated with a piecewise constant function. For models with this short periodicity the influence of the depth-dependent variations are significant

and thus we will only use this classic solution to approximate the accuracy of the numerical implementation in Section 4.2.

Weak Solutions

The function $\kappa(x)$ appears in the heat equation with its first derivative, which is not smooth for a piecewise constant function. Thus, we need to broaden our definition of derivatives. Details on weak derivatives and weak solutions for general problems as well as the procedure of finding such solutions can be reread in (Evans, 2010, Ch. 7).

For the variational formulation of the initial-boundary-value problem, we state the setting in different function spaces. We therefore define

$$\mathbf{u}\colon [0, T] \to H^1([0, x_E]), \qquad \mathbf{u}[t] : x \mapsto u(x, t). \tag{2.6}$$

For a given $t \in [0, T]$ we abbreviate $\mathbf{u}^t = u(\cdot, t)$. From now on we also assume, that $\kappa(x) = \kappa_j$ for $x \in I_j$, where $I_j, j = 1, \dots, s$ are disjoint subintervals such that $\dot{\bigcup}_{j=1}^{s} I_j = [0, x_E]$.

To obtain the variational formulation, we multiply the heat equation (2.1) with a test-function $\psi(x) \in H^1([0, x_E])$ and integrate over $[0, x_E]$ to obtain

$$0 = \partial_t \int_0^{x_E} \mathbf{u}^t(x)\psi(x)\, dx - \int_0^{x_E} \partial_x \left(\kappa(x)\partial_x \mathbf{u}^t(x)\right)\psi(x)\, dx.$$

Now, Green's first identity (Allaire, 2007) yields

$$\begin{aligned} 0 &= \partial_t \int_0^{x_E} \mathbf{u}^t(x)\psi(x)\, dx - \left[\left[\kappa(x)\partial_x \mathbf{u}^t(x)\psi(x)\right]_0^{x_E} - \int_0^{x_E} \kappa(x)\partial_x \mathbf{u}^t(x)\partial_x\psi(x)\, dx\right] \\ &= \partial_t \int_0^{x_E} \mathbf{u}^t(x)\psi(x)\, dx - \kappa(x_E)\partial_x \mathbf{u}^t(x_E)\psi(x_E) + \kappa(0)\partial_x \mathbf{u}^t(0)\psi(0) \\ &\quad + \int_0^{x_E} \kappa(x)\partial_x \mathbf{u}^t(x)\partial_x\psi(x)\, dx. \end{aligned}$$

Inserting the boundary values (2.2) – (2.3), we end with

$$0 = \partial_t \int_0^{x_E} \mathbf{u}^t(x)\psi(x)\,dx - \kappa(x_E)\,g\psi(x_E) + \tilde{h}\,(\mathbf{u}^t(0) - W(t))\,\psi(0) + \int_0^{x_E} \kappa(x)\partial_x \mathbf{u}^t(x)\partial_x\psi(x)\,dx.$$

For further analysis we introduce the bilinearform

$$a\colon (\mathbf{u}^t, \psi) \mapsto \int_0^{x_E} \kappa(x)\partial_x \mathbf{u}^t(x)\partial_x\psi(x)\,dx + \tilde{h}\mathbf{u}^t(0)\psi(0) \tag{2.7}$$

and end up with the variational formulation

$$\partial_t\langle \mathbf{u}^t, \psi\rangle + a[\mathbf{u}^t, \psi] = \kappa(x_E)g\psi(x_E) + \tilde{h}\,W(t)\psi(0). \tag{2.8}$$

Here $\langle\cdot,\cdot\rangle$ denotes the scalar product in $L^2([0, x_E])$.

Definition 2.2.2. *The Weak Formulation of the Model.*
A function $\mathbf{u} \in L^2([0, T], H^1([0, x_E]))$ *with derivative* $\partial_t\mathbf{u} \in L^2([0, T], H^{-1}([0, x_E]))$ *is called a weak solution of the initial-boundary-problem of Definition 2.1.1, if it solves the integral equation* (2.8) *for every test function* $\psi(x) \in H^1([0, x_E])$ *and fulfills the initial condition* $\mathbf{u}[0] = u_0(x)$ *for almost every* $x \in [0, x_E]$ *and given* $u_0(x) \in H^1([0, x_E])$ *and* $W(t) \in L^2([0, T])$.

Existence and Uniqueness of the Weak Solution

The existence and uniqueness of such a weak solution can be easily adapted from the proofs provided by literature for general or special cases. Here, we will only show the main ideas of the proof, the details can be seen in Evans (2010).

Theorem 2.2.3. *Existence of a Weak Solution.*
There exists a weak solution of (2.8).

Proof. (for details see (Evans, 2010, Ch. 7.1.2, Thms. 1-3)) The idea is to first construct solutions of certain finite-dimensional approximations to a solution to the model from Definition 2.1.1 and then passing to limits. Let $w_k = w_k(x)$, $k = 1, 2, \ldots$ be an orthogonal basis

of $H^1([0, x_E])$ and an orthonormal basis of $L^2([0, x_E])$. For a fixed $m > 0$, the approach is to look for a function of the form

$$\mathbf{u}_m^t(x) := \sum_{k=1}^{m} d_m^k(t)\, w_k(x), \tag{2.9}$$

that satisfies the weak formulation in Equation (2.8) for every w_k, $k \leq m$.

Thus we are looking for coefficients d_m^k, such that for the initial values

$$d_m^k(0) = \int_{[0,x_E]} u_0(x)\, w_k(x)\, dx \tag{2.10}$$

holds. Inserting the w_k as test functions into the the weak formulation (2.8) we get

$$\partial_t \langle \mathbf{u}_m^t, w_k \rangle + a[\mathbf{u}_m^t, w_k] = \kappa(x_E) g w_k(x_E) + \tilde{h}\, W(t) w_k(0) \tag{2.11}$$

for every $k \leq m$. With the representation (2.9) it follows that

$$\partial_t d_m^k(t) + \sum_{l=1}^{m} a[w_l, w_k] d_m^l(t) = \kappa(x_E) g w_k(x_E) + \tilde{h}\, W(t) w_k(0) \tag{2.12}$$

for every $k \leq m$. Now, this linear system of ordinary differential equations (2.9) and initial condition (2.10) can be uniquely solved for every $m \in \mathbb{N}$ by standard existence theory of ordinary differential equations, i.e. with the Picard-Lindelöf Theorem (Evans, 2010, Ch. 7, Thm. 1).

For the next step we show that the sequences $\{\mathbf{u}_m^t\}_{m=1}^{\infty}$ and $\{\partial_t \mathbf{u}_m^t\}_{m=1}^{\infty}$ are bounded in $H^1([0, x_E])$ and $H^{-1}([0, x_E])$, respectively (Evans, 2010, Ch. 7, Thm. 2). For the bilinear form a of Equation (2.7) *continuity* and *ellipticity* can be shown, i.e. there exist constants $\alpha, \beta > 0$, such that

$$|a[u, w]| \leq \alpha \|u\|_{H^1([0,x_E])} \|w\|_{H^1([0,x_E])} \text{ and } \beta \|u\|^2_{H^1([0,x_E])} \leq a[u, u]$$

hold (Evans, 2010, Sec. 6.2.2).

Multiplying Equation (2.11) with d_k^m and sum over $k = 1, \dots, m$, we find that

$$\begin{aligned}\kappa(0)\tilde{h}W(t)\,\mathbf{u}_m^t(0) + \kappa(x_E)\,g\,\mathbf{u}_m^t(x_E) =& \partial_t \int_0^{x_E} \mathbf{u}_m^t(x)\,\mathbf{u}_m^t(x)\,dx \\ &+ \int_0^{x_E} \kappa(x)\partial_x \mathbf{u}_m^t(x)\,\partial_x \mathbf{u}_m^t(x)\,dx \\ &+ \kappa(0)\tilde{h}\mathbf{u}_m^t(0)\,\mathbf{u}_m^t(0) \\ =& \partial_t \frac{1}{2}\|\mathbf{u}_m^t\|^2_{L^2([0,x_E])} + a[\mathbf{u}_m^t, \mathbf{u}_m^t],\end{aligned}$$

for every $t \in [0, T]$. Furthermore,

$$\kappa(0)\tilde{h}\,W(t)\,\mathbf{u}_m^t(0) + \kappa(x_E)\,g \cdot \mathbf{u}_m^t(x_E) \leq c_1\|W\|_{L^2([0,x_E])} + c_2|g| + c_3\|\mathbf{u}_m^t\|_{L^2([0,x_E])}$$

for almost every $t \in [0, T]$. With the ellipticity of the bilinear form a, the above equation then yields

$$\partial_t \frac{1}{2}\|\mathbf{u}_m^t\|^2_{L^2([0,x_E])} + \beta\|\mathbf{u}_m^t\|^2_{H^1([0,x_E])} \leq c_1\|W\|_{L^2([0,x_E])} + c_2|g| + c_3\|\mathbf{u}_m^t\|_{L^2([0,x_E])}$$

for almost every $t \in [0, T]$.
This yields boundedness of the sequences $\{\mathbf{u}_m^t\}_{m=1}^\infty$ and $\{\partial_t \mathbf{u}_m^t\}_{m=1}^\infty$. From these boundedness results, often referred to as the *energy estimates* (Evans, 2010, Ch. 7, Thm. 2) it follows that sub-sequences exist, which converge in the weak sense to functions $\mathbf{u}^t \in H^1([0, x_E])$ and $\partial_t \mathbf{u}^t \in H^{-1}([0, x_E])$ respectively, as the spaces $H^1([0, x_E])$ and $H^{-1}([0, x_E])$ are complete metric spaces. For these functions it can then be shown, that they satisfy the given boundary-conditions (2.2) – (2.3) and therefore are a solution to the given problem (Evans, 2010, Ch. 7, Thm. 3). □

Theorem 2.2.4. *Uniqueness of the Weak Solution.*
A weak solution of (2.8) *is unique.*

Proof. (for details see (Evans, 2010, Ch. 7, Thm. 4))
Without loss of generality, we only check that for $u_0 = W(t) \equiv g \equiv 0$ the only solution to (2.8) is $\mathbf{u}^t(x) \equiv 0$.
First, we insert $\mathbf{u}^t(x) \in H^1([0, x_E])$ as a test function in (2.8). In terms of the bilinear form

(2.7), this yields

$$\begin{aligned} 0 =& \partial_t \int_0^{x_E} \mathbf{u}^t(x) \cdot \mathbf{u}^t(x)\, dx + \int_0^{x_E} \kappa(x) \partial_x \mathbf{u}^t(x) \cdot \partial_x \mathbf{u}^t(x)\, dx + \kappa(0) \tilde{h} \mathbf{u}^t(0) \cdot \mathbf{u}^t(0) \\ =& \partial_t \left(\frac{1}{2} \|\mathbf{u}^t\|^2_{L^2([0,x_E])} \right) + a[\mathbf{u}^t, \mathbf{u}^t]. \end{aligned}$$

By its coercivity we get

$$0 \geq \partial_t \left(\frac{1}{2} \|\mathbf{u}^t\|^2_{L^2([0,x_E])} \right) + \beta \|\mathbf{u}^t\|^2_{H^1([0,x_E])}.$$

Since norms are only zero if and only if the argument is zero, it follows with Gronwall's Lemma (Evans, 2010, Appendix B) that $\mathbf{u}^t \equiv 0$ for all $t \in [0, T]$. □

2.3. Definition and Derivative of the Forward Operator

Definition 2.3.1. *The Forward Operator.*
For given $u_0 \in H^1([0, x_E])$ and $\kappa(x)$, and a family of functions $W[f](t) \in L^2([0, T])$, we call

$$\Phi : (f, g) \mapsto u(x, t) \tag{2.13}$$

the forward operator*, where $u(x, t)$ is the unique solution to* (2.8)*.*
We will use the short form $\Phi(p)$ with a combined parameter vector $p = (f, g)$.

Remark 2.3.2.
Regularity results for the weak solution can only be obtained if the functions $\kappa(x)$, $\mathbf{u}_0(t)$ and $W[f](t)$ are also more regular. This topic is covered in (Evans, 2010, Sec. 7.1.3). However, in our case point-wise evaluation is possible and continuity in x and t is provable and therefore the definition of the forward model operator is well defined.

The forward operator as described in Definition 2.3.1 is non-linear. However, the solution $u(x, t)$ of (2.8) depends linearly on the boundary conditions.

Definition 2.3.3.
For given $u_0 \in H^1([0, x_E])$, $\kappa(x)$, and $\tilde{h} > 0$ we call

$$S : L^2([0, T]) \times \mathbb{R} \to L^2([0, T], H^1([0, x_E]))$$
$$(W, g) \mapsto u(x, t)$$

the solution operator*, where $u(x, t)$ is the unique solution to* (2.8)*.*

Definition 2.3.4.
For a family of functions $W[f](t) \in L^2([0, T])$ with parameters $f \in \mathbb{R}^k$ and $g \in \mathbb{R}$, we call

$$N : \mathbb{R}^k \times \mathbb{R} \to L^2([0, T]) \times \mathbb{R}$$
$$(f, g) \mapsto (W[f], g)$$

the boundary condition operator*.*

Thus, we can write the forward operator as the concatenation of the boundary operator and the solution operator

$$\Phi(f, g) : \mathbb{R}^k \times \mathbb{R} \to L^2([0, T], H^1([0, x_E]))$$
$$\Phi(f, g) = S(N(f, g)). \tag{2.14}$$

Theorem 2.3.5. *Fréchet Derivative of the Forward Operator.*
The forward operator as defined in 2.3.1 is Fréchet differentiable, if the bottom water temperature function is differentiable with respect to the parameters f.

Proof. A definition of the Fréchet derivative can be found in (Allaire, 2007, Ch. 10.1.2). For the forward operator we have the representation $\Phi(f, g) = S(N(f, g))$ (2.14), where S is linear in both arguments. Therefore, the Fréchet derivative S' of S is the operator itself. We can now use the chain rule to obtain the Fréchet derivative of the forward operator as

$$\Phi'(f, g) = \frac{\partial}{\partial N} S \left(\frac{\partial}{\partial (f, g)} N(f, g) \right) \tag{2.15}$$

□

3. Temperature Measurements and Basic Regions

This section is dedicated to the temperature measurements in marine sediments, which are the input to the inverse problem. We will first introduce a method used to determine depth-dependent thermal properties and in-situ temperature-depth profiles. Then, we will take a closer look at three measured temperature-depth profiles in three different regions. The locations of these three measurements will be the anchor points for three *basic regions* with different thermal properties and different bottom water temperature functions, which we will use later on to generate artificial data sets.

3.1. Temperature Measurements in Marine Sediments

The principle for the measurements of depth-dependent thermal parameters originates from the classical method of determining steady state heat flow values for the oceanic crust from deep sea sediments. Heat flow values are determined based on Fourier's law of thermal conduction from the steady state (undisturbed) temperature gradient and thermal conductivities.

The design of deep sea Lister-type heat flow probes follows the concept by Hyndman et al. (1979), where a thermistor string parallel to a massive strength member penetrates into the sediment by its own weight. A total of 22 thermistors record the sediment temperature during the whole process. The in-situ temperature is determined from the decay of the frictional heat accompanying the penetration, while the thermal conductivity and diffusivity

values are calculated from the temperature decay of an artificial, exactly defined, calibrated heat pulse which heats up the sediment (Hartmann and Villinger, 2002).

This method is normally used in deep sea soft sediments, where the heat flow probe penetrates by its own weight. Shallow water sediments in the North and Baltic seas are characterized by more shear resistant sediments such as sands, tills, and clays, where this classical method of penetration by gravity alone is not applicable. For this reason, a thermistor string has been combined with a standard VKG vibrocorer (Dillon et al., 2012), where a barrel is forced into the sediment with a vibration motor. In Fig. 2 the FIELAX heat flow probe is shown during operation.

The measuring procedure is the same for both systems: the system is lowered towards the sediment, penetrates the sediment by vibrocoring and rests in the sediment for in-situ temperature measurement and heat pulse decay recording. With this system, a penetration depth of up to 6 m is possible. The accuracy of the thermistor strings is 2 mK and the resolution is 1 mK for both systems.

The procedure is illustrated by an excerpt of the raw data in Fig. 3. The time-dependent temperature is shown for the uppermost sensor (in green), the lowermost sensor (in red) and the sensor in the middle of the probe (in blue). There is a temperature rise from the penetration, which is highest for the lowermost sensor. From the temperature decay after this temperature rise, the in-situ temperature is calculated in the inversion process by Hartmann and Villinger (2002). After approximately 8 min the defined heat pulse is fired and from the decay the thermal properties of the sediment are obtained.

The processing of the raw data from both measuring devices is handled with the same software tool which allows determining in-situ temperatures and thermal material properties with an inversion algorithm following Hartmann and Villinger (2002). Based on an assumption of thermal decay around a cylindrical symmetric infinite line source, steady state in-situ temperatures, thermal conductivity and thermal diffusivity are determined in an iterative inversion procedure. From thermal conductivity and diffusivity the volumetric heat capacity can also be determined by $\rho c_v = \lambda/\kappa$. All thermal properties can be obtained with less than 0.5 % error.

The in situ geothermal gradient can also be determined directly from these measurements via Bullard's method (Bullard, 1939). For measured temperature $\Theta(x)$ and thermal con-

Figure 2.: The FIELAX HetaFlowProbe in during operation. This device has a 6 m long strength member which leads the thermistor string (on the right hand side of the barrel) while the device penetrates the sediment. On top of the strength member, the electronics and power supply are housed in two cylindrical high strength titanium pressure cases. On the thermistor string the position of the uppermost (green), middle (blue) and lowermost (red) sensors are marked.

ductivity $\lambda(x)$ the cumulative thermal resistance is defined as $\Omega = \int_0^{x_e} 1/\lambda(x)\, dx$ and the heat flow is then calculated as $q = \delta\Theta(x)/\delta\Omega(x)$ (Beardsmore and Cull, 2001; Stranne and O'Regan, 2016). If the resulting heat flow is not constant over depth, the measurements are not deep enough to obtain the undisturbed background heat flow and thus the determination of the in-situ geothermal gradient is not possible with Bullard's method. The global heat flow database collects data sets and can therefore be used to find an appropriate regional approximation (International Heat Flow Commission (IHFC), 2014). Background heat flow and geothermal gradient are related via $q = \lambda\, g$. Therefore the geothermal gradient is calculated as $g = 1/\lambda(x_e)\, q(x_e)$. Calculating the in-situ geothermal gradient from measurements where Bullard's method fails would clearly improve said database.

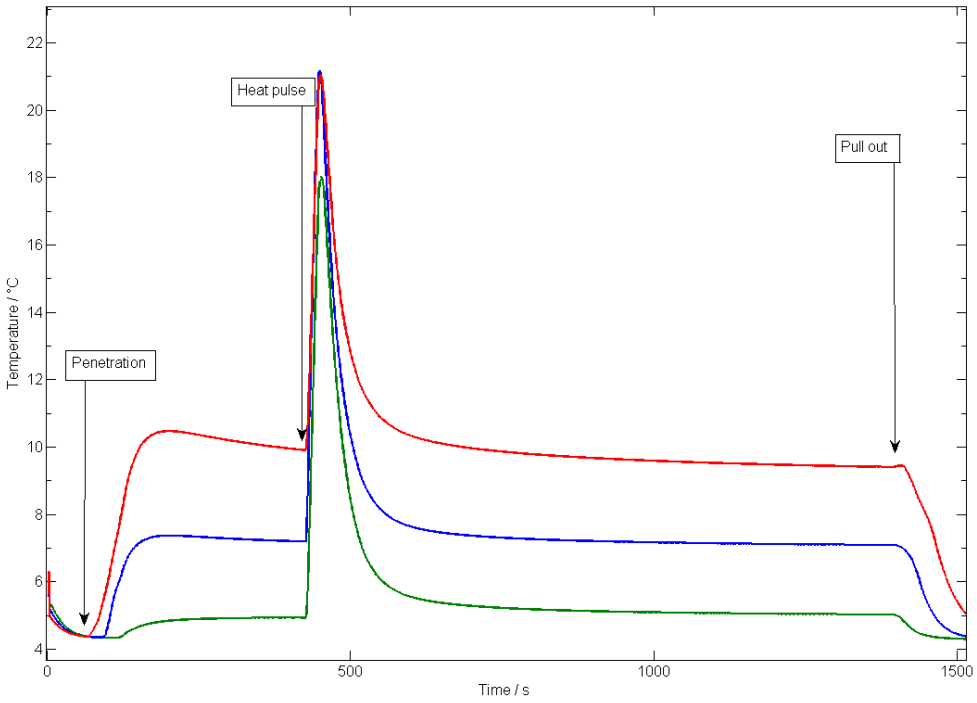

Figure 3.: A typical sequence of the recorded temperatures from the thermistor string. Only 3 of the 22 sensors are depicted, the uppermost in green, the lowest in red and a middle one in blue, as marked in Fig. 2.
The three events during the measurement are marked. First, the temperature rises from frictional heating accompanying the penetration, second the sudden temperature rise induced by an exactly known heat pulse and last another small temperature rise occurring during pull out before the temperature drops again to sea water level.

3.2. Example Data Sets

For the analysis of the proposed methods, we will use artificial data with known input parameters and additive noise. In order to be able to assess the applicability to real data sets, it is necessary to use parameter sets and noise levels that can be observed in real areas before analyzing the methods also on real measured data sets.

Therefore, we will introduce three real data sets from different regions with different typical bottom water temperatures and thermal properties. For each region we will give a short overview over the regions' oceanography and propose a typical parameter set for the bottom water temperature. As we focus on one-year periodic functions, the parameter sets contain the mean temperature A, the amplitude B and the day of the annual minimum d. This leads to a temperature in the range of $[A - B, A + B]$, where the minimal value is taken on at day d of each year. These typical parameters and the measured geothermal gradient will be used to generate artificial data sets from the measured thermal diffusivity values throughout this work. The example areas are color-coded for clarity. Note, that the proposed parameter are only examples and we do not know if they are a good representation for the specific bottom water temperatures at the measurements, i.e. the measured data probably can not be explained by a water temperature function with these parameters.

Example 1. High mean temperature, high amplitude, highly heterogeneous sediment, average geothermal gradient

The first example data set was measured using the VibroHeat method (Dillon et al., 2012) in the middle of March 2012, north-west of the island of Helgoland in the German North Sea at 20 m water depth and is already published by Miesner et al. (2015); Müller et al. (2013). The uppermost three sensors did not penetrate the sediment and were therefore not used. Thus, this measurement contains 19 depth-dependent temperature and thermal diffusivity measurements at a depth of up to 2.5 m. The post-processing estimates the relative noise level on the data as 2.0 %. It was not possible to obtain the geothermal gradient from the measurements and therefore a literature value of 0.03 K/m (Bonté et al., 2012; Davies, 2013; Madsen, 1974; Sundvor and Myhre, 1987), will be used for this area.

The measured depth-dependent temperature and thermal diffusivity values are shown in Fig. 4. We see that the temperature is small at the sediment-water-interface and increases with

increasing depth. This relates to the bottom water temperature, which is the lowest at the end of February and thus, the deeper sediment still carries the information of the last warmer summer.

Temperature and salinity of the North Sea are mainly governed by a general cyclonic circulation, which renews the water in the timescale of 1 year (Rhode et al., 2007). The freshwater input from rivers is comparably small. The central part of the North Sea becomes stratified due to heating in the summer but gets vertically mixed during winter. At the western and southern coasts, the vertical stratification is prevented by strong tidal currents (for detail on the oceanography of the German North Sea see (Rhode et al., 2007)). In conclusion, a simple shifted cosine with a one-year period is a sufficient approximation to the bottom water temperature function.

In the right panel of Fig. 4 the measured thermal diffusivity is shown. The values are distributed over a wide range from $2 \times 10^{-7}\,\mathrm{m}^2/\mathrm{s}$ to $9 \times 10^{-7}\,\mathrm{m}^2/\mathrm{s}$. Highly heterogeneous thermal diffusivity profiles and bottom water temperatures ranging between 2 °C and 20 °C are typical for the German North and Baltic Sea.

For generating artificial data in this example area, we will approximate the bottom water temperature function with a mean temperature of 7.5 °C, an amplitude of 5.5 K and assume that the water is the coldest on 25 February (day 55 of the year).

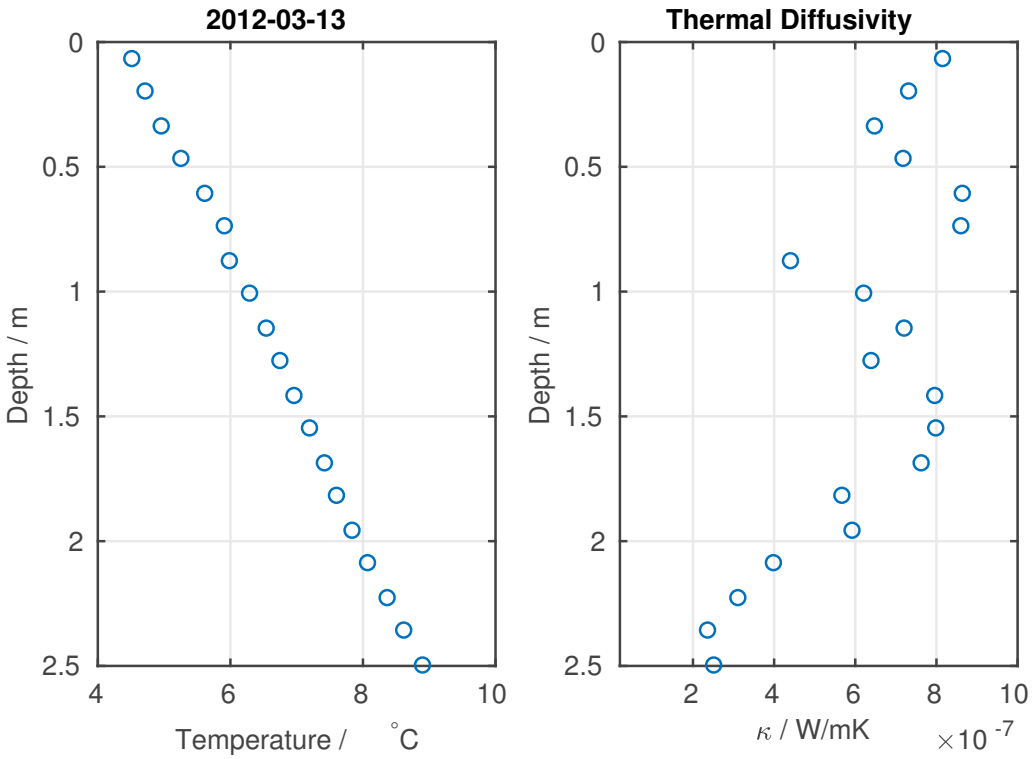

Figure 4.: Data set acquired in the middle of March 2012 northwest of the island of Helgoland in 20 m water depth. On the left hand side we see the recorded temperatures and on the right the calculated thermal diffusivity.
19 of 22 total sensors were used. The vertical distance between the sensors is 13.5 cm.

Example 2. High mean temperature, low amplitude, nearly homogeneous sediment, high geothermal gradient

The second example was acquired west of Mexico in the Wagner Basin at the beginning of May in 2015 at 200 m water depth. The temperature- and diffusivity-depth profiles are shown in Fig. 5. The measurement contains only 21 depth-sensors at a depth of up to 5.06 m, as one sensor was omitted due to technical problems. In the post-processing the geothermal gradient could be obtained to 0.084 K/m, which is higher than the global average. The relative noise level on the data was approximated to 0.1 %.

The sediment temperature is again increasing with depth, but the uppermost two sensors also measured higher temperatures. This is due to the fact that the measurement took place roughly two months after the coldest day of the year.

The diffusivity values show only slight variations around $3 \times 10^{-7}\,\mathrm{m^2/s}$.

The Wagner Basin is located in the north of the Gulf of California and has a maximum depth of 225 m. The water is mainly Eastern Tropical Pacific Water (Robles and Marinone, 1987) with most seasonal temperature and salinity changes in the surface layer. The seasonal changes are produced by seasonality of the Pacific Ocean, i.e. the wind system and the fluxes of heat and moisture (Lavìn and Marinone, 2003). The surface layer covers a range from 16 °C in February to 31 °C in August Robles and Marinone (1987). Because of the low-pass filter attributes of the water itself, the amplitudes of any seasonal deviation decreases with increasing depth, but seasonal cycle of the temperature is also present in deeper layers.

For the approximation of the water temperature for this area and the depth of the measurement, we use a mean temperature of 10.5 °C, an amplitude of 0.3 K and 14 February (day 45 of the year) as the coldest day.

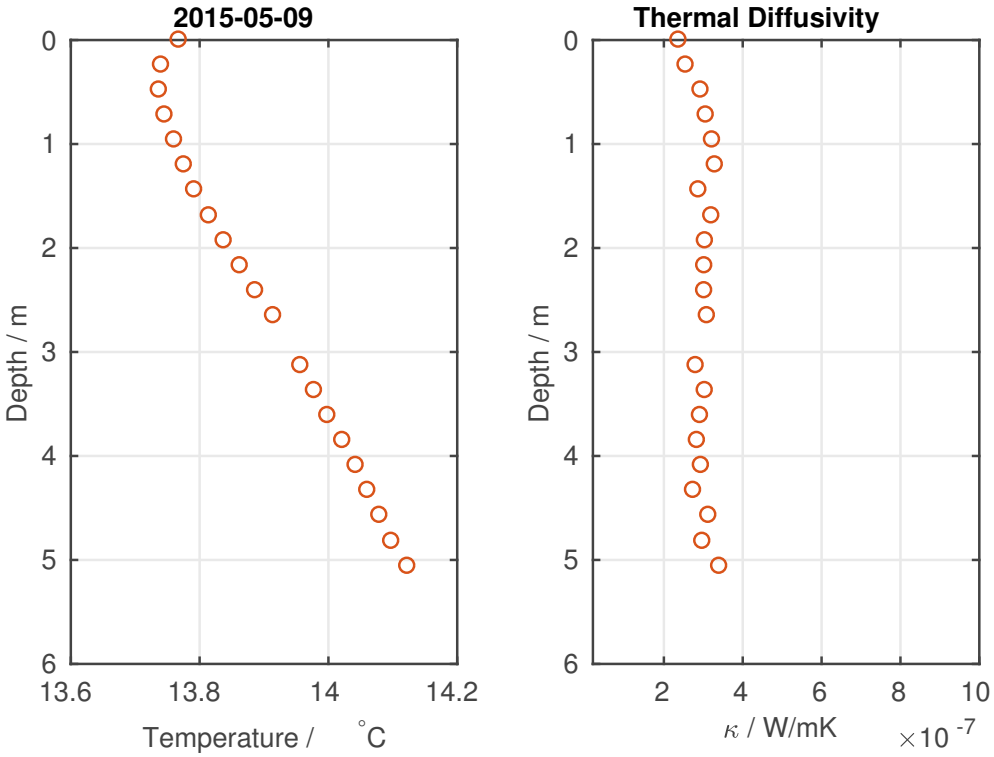

Figure 5.: Data set acquired at the beginning of May 2015 west of Mexico in the Gulf of California in 200 m water depth. On the left hand side we see the recorded temperatures and on the right hand side the calculated thermal diffusivity.
21 of 22 total sensors were used. The vertical distance between the sensors is 25 cm.

Example 3. Low mean temperature, low amplitude, heterogeneous sediment, slightly higher geothermal gradient

The third example was acquired at the end of August in 2006 southwest of Greenland at 1304 m water depth and was also already published by Miesner et al. (2015).

Baffin Bay and Davis Strait are characterized by the northwards flowing West Greenland Current moving temperate saline Irminger Water from the Atlantic Ocean in the top layers and cold low-salinity Polar Water in the bottom layers (Ribergaard, 2011). The data set from the Davis Strait is obtained from 1300 m depth, where the seasonal influence is mostly damped by the water column and the cold Polar Water is dominant.(Ribergaard, 2011)

In Fig. 6 the measured temperature and thermal diffusivity values are shown. We again see the highest temperatures at the lowermost sensors. The data consists of 19 sensors at a depth of up to 5.27 m. The geothermal gradient is near the global average with its 0.038 K/m and the relative noise level was approximated to 0.1 %.

The typical bottom water temperature will be approximated with a mean temperature of 3 °C, an amplitude of 0.5 K and 7 March (day 65 of the year) as the day of the annual minimum temperature.

In Table 1 the main characteristics of the three basic regions are summed up.

Table 1.: Main characteristics of the basic regions.
Note that the bottom water temperature parameters are guessed example parameters, while the other properties are measured data.

Example No.	1	2	3
Location	German North Sea	Wagner Basin	Baffin Bay
Water Depth in m	20	200	1304
Number of Sensors	19	21	19
Lowermost Sensor in m	2.5	5.06	5.27
Thermal Diffusivity in $10^{-7}\,m^2/s$	2.1–9.0	2.9 – 3.1	3.8 – 6.5
Geothermal Gradient in K/m	0.030	0.084	0.038
Mean Temperature in °C	7.5	10.5	3.0
Amplitude in K	5.5	0.3	0.5
Day of Annual Minimum	55	45	65

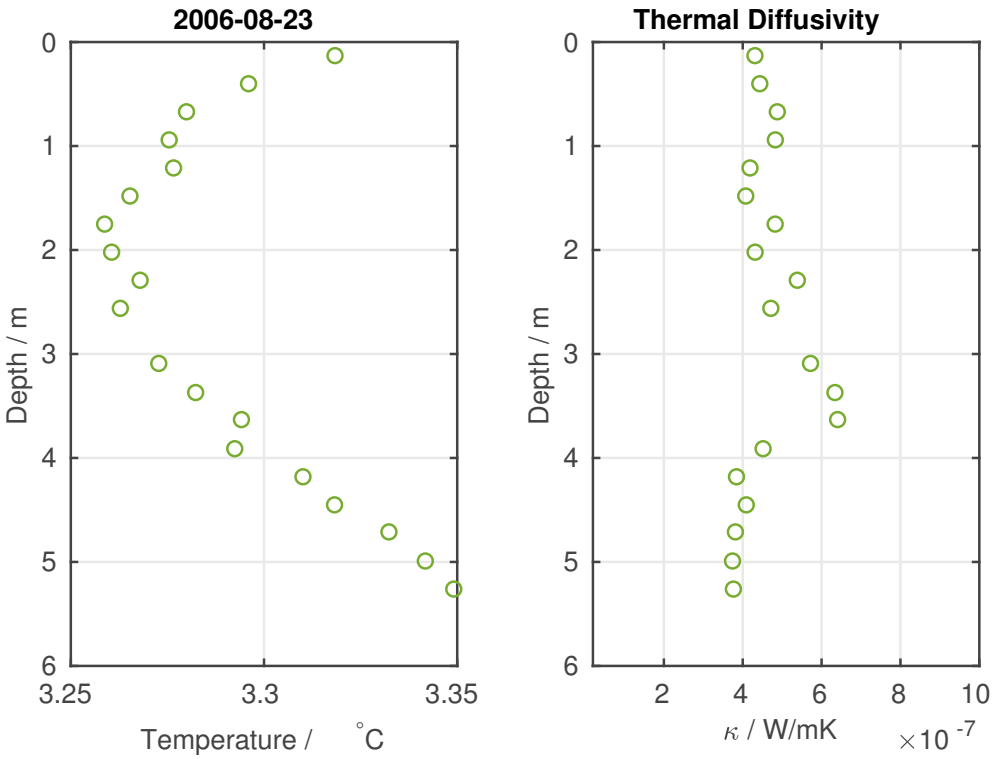

Figure 6.: Data set acquired at the end of August 2006 southwest of Greenland in 1304 m water depth. On the left hand side we see the recorded temperatures and on the right hand side the calculated thermal diffusivity.
19 of 22 total sensors were used. The vertical distance between the sensors is 27 cm.

3.3. Characterization of the Geothermal Gradient

As stated before, the geothermal gradient can be obtained from the depth-dependent temperature and thermal diffusivity values via Bullard's method, as described in Section 3.1, and thus be treated as an input parameter to the forward operator Φ (see Definition 2.3.1). However, Bullard's method only yields accurate results if no seasonal fluctuations overlay the background (steady state) heat flow.

All the data sets we are analyzing in this work are not in the steady state and the heat flow probe does not penetrate the sediment to a depth where the steady state prevails. The depth to which a thermal perturbation can be transmitted over a certain amount of time, is calculated with Fick's second Law of Diffusion and is denoted by the thermal length $l = \sqrt{4\kappa t}$ (Irene, 2005). For the maximum values of the thermal diffusivity of the three example areas and a time span t of one year, the thermal length is shown in Table 2. As the thermal diffusivity is not constant in our basic regions, we also show the *apparent depth*. Therefor, we calculated the temperature-depth profile numerically and obtained the depth in which the temperature difference to the geothermal gradient drops below 0.01 K, i.e. in which the influence of the seasonal fluctuations is less than said 0.01 K.

Table 2.: Thermal lengths l for the example regions for the mean and maximum values of the thermal diffusivity κ. The apparent depth is approximated from the forward model operator as the depth in which the difference between the temperature field and the geothermal gradient is smaller than 0.01 K.

Example	mean κ	l	max κ	l	apparent depth
	$10^{-7}\,\mathrm{m}^2/\mathrm{s}$	m	$10^{-7}\,\mathrm{m}^2/\mathrm{s}$	m	m
1	6.2	8.9	8.7	10.5	8.3
2	3.0	6.1	3.4	6.6	1.8
3	4.7	7.7	6.4	9.0	3.8

We can observe that the thermal length to which a temperature signal *could* be transmitted (the thermal length l) is more than the depth to which the chosen cosine-functions propagate measurably (the *apparent depth*) in our example regions. The depth in which the seasonal fluctuation fades depends not only on the thermal properties of the sediment, but also on the amplitude. Thus, we have the smallest apparent depth at Example 2, where the thermal

diffusivity and the amplitude are small, and the highest apparent depth at Example 1, where both have higher values.
In conclusion, it can be possible to retrieve reliable calculations of the geothermal heat flow from 6 m deep measurements, although it depends on the thermal diffusivity and the amplitude of the bottom water temperature function.
Therefore, it seems reasonable to invert the forward operator given in Definition2.3.1 for both, the parameters of the bottom water temperature function as well as the geothermal gradient.

4. Numerical Implementation of the Forward Operator

Now, we know that there is a unique solution to the model given in Definition 2.1.1. However, with the exception of some special cases, we can not directly calculate this solution. That is why we will introduce the numerical methods to find a good approximation to the solution and analyze their convergence properties.

4.1. The Method of Lines

Starting with the integral equation (2.8)

$$\partial_t \langle \mathbf{u}^t, \psi \rangle + a[\mathbf{u}^t, \psi] = g\psi(x_E) + \tilde{h}W(t)\psi(0),$$

we have to discretize the equation in time and space.

First the space discretization is done with the one-dimensional Finite Element Method (FEM). For a thorough introduction see Allaire (2007). We therefore lay a mesh of n equidistant distributed points over $[0, x_E]$, such that $0 = x_1 < x_2 < \cdots < x_n = x_E$. Upon this mesh we define basis functions v_l with support on the l^{th} subinterval $I_l := [x_l, x_{l+1}]$, $l = 1, \dots, n-1$.

The solution over this mesh is denoted with $u_n(x, t)$. Following the Galerkin approach, this solution can now be represented in the finite-element-space as $u_n(x, t) = \sum_l u_l^n(t) v_l(x)$.

Substituting this representation into the integral equation (2.8) and replacing the test func-

tions $v(x)$ with their finite-element-space representation v_i yields

$$\begin{aligned}\kappa(0)\tilde{h}(t)W(t)v_k(0) + g\, v_k(x_E) = &\int_0^{x_E} \partial_t \sum_l u_l^n(t) v_l(x)\, v_i(x)\, dx \\ &+ \int_0^{x_E} \kappa(x)\partial_x \sum_l u_l^n(t) v_l(x)\, v_i(x)\, dx \\ &+ \kappa(0)\tilde{h}(t) u_1^n(0,t) v_1(0) \\ = &\partial_t \sum_l u_l^n(t) \int_0^{x_E} v_l(x)\, v_i(x)\, dx \\ &+ \sum_l^n u_l^n(t) \int_0^{x_E} \kappa(x)\partial_x v_l(x)\, \partial_x v_i(x)\, dx \\ &+ \kappa(0) h(t) u_k^n(0,t) v_k(0),\end{aligned}$$

for all $i = 1, \dots, k$. The sums and integrals can be interchanged as both are finite. In terms of the scalar product in L^2 and the bilinear form (2.7) we end with

$$\kappa(0)\tilde{h}(t)W(t)v_1(0) + g\, v_k(x_E) = \sum_l \frac{d}{dt} u_l^n(t) \langle v_l(x), v_i(x)\rangle + \sum_l u_l^n(t) a[v_l(x), v_i(x)], \quad (4.1)$$

for all $i = 1, \dots, k$.

Assembling the scalar products and bilinear form terms into $(n-1) \times k$ matrices we obtain a system of linear ordinary differential equations,

$$G^n(t) = M^n \frac{d}{dt} u^n(t) + K^n u^n(t). \quad (4.2)$$

Here, G^n contains the left hand side $\kappa(0)\tilde{h}(t)W(t)v_1(0) + g\, v_k(x_E)$, M^n the scalar products $\langle v_l(x), v_i(x)\rangle$, and K^n the bilinearform $a[v_l(x), v_i(x)]$. With associated initial values $u_j^n(0) = u^n(x_l, 0)$ this ODE is solvable in every mesh point x_j ($j = 1, \dots, n$) and for every time $t \in [0, T]$ (Allaire, 2007). This method is called (vertical) method of lines, because in this first step of the discretization we get a function of t for every discrete mesh point x_i, and thus resulting in vertical lines in a t-over-x plot.

For increasing n, and thus decreasing mesh size δx, the approximative solution of this ODE

converges to the weak solution of (2.8). We cite an estimation from Hanke-Bourgeois (2009) without proof.

Theorem 4.1.1. *(from (Hanke-Bourgeois, 2009, page 727))*
Provided $u_0 \in H^1([0, x_E]) \cap H^2([0, x_E])$*, the weak solution* $\mathbf{u}(t) \in H^1([0, x_E])$ *of* (2.8) *has a derivative with respect to time* $\partial_t \mathbf{u}(t) \in H^2([0, x_E])$ *for almost all* $t \geq 0$ *and* $\mathbf{u}^k(t)$ *is the solution to* (4.2)*, then*

$$\begin{aligned}\|u^n(t) - \mathbf{u}(t)\|_{L^2([0,x_E])} \leq \|u^k(0) - u_0\|_{L^2([0,x_E])} \\ + 2c_0 \cdot (\delta x)_k^2 \left(\|u_0\|_{H^2([0,x_E])} + \int_0^t \|\partial_t u^n(\tau)\|_{H^2([0,x_E]\times\mathbb{R}_{\geq 0})}\, d\tau \right)\end{aligned}$$

holds for every $t \in [0, T]$*.*
Here, c_0 *is a constant and* $(\delta x)_n := \frac{x_E}{n-1}$ *is the mesh size associated with the discretization of* $[0, x_E]$ *into* $n-1$ *subintervals* $0 = x_1 < x_2 < \ldots < x_n = x_E$*.*

Using also an equidistant mesh in the time domain $0 = t_1 < t_2 < \ldots < t_n = T$ and a suitable method for the initial value problem over each time interval, we get $u_{m,n} \in \mathbb{R}^{k \times n}$ as the discrete solution to the model.
As the one-dimensional heat equation is not very complex, no specialized numerical scheme is needed and thus, we used the built-in MATLAB algorithm `pdepe` (MathWorks, 2017). For the spatial semi-discretization they use the Galerkin approach as illustrated above (see Skeel and Berzins (1990) for details). The resulting initial value problem (4.2) is then solved with `ode15s` (MathWorks, 2017). This ODE solver is based on the numerical differentiation formulas of orders 1 to 5, and chooses this order automatically (Berzins et al., 1989).

Summarizing this section, we can formulate the discrete forward operator as follows.

Definition 4.1.2. *Discrete Forward Operator*
For given $m, n \in \mathbb{R}$*, let* $\mathbf{x} = (x_1, \ldots, x_m)$ *with* $0 = x_1 < \cdots < x_m = x_E$ *and* $\mathbf{t} = (t_1, \ldots, t_n)$ *with* $0 = t_1 < \cdots < t_n = T$ *equidistant spatial and time discretization.*
Let then $u_0 \in H^1([0, x_E])$*,* $\kappa(x)$ *be a piece wise constant function and* $W[f](t) \in C([0, T])$*, i.e. the point-wise evaluations at* x_i *for* $i = 1, \ldots, m$ *or* t_j *for* $j = 1, \ldots, n$ *are well-defined.*

We then call

$$\hat{\Phi}_{m,n}\colon (f, g) \to u_{m,n}(\mathbf{x}, \mathbf{t}) \tag{4.3}$$

the discrete forward operator*, where $u_{m,n}$ is the $m \times n$ solution to* (4.2) *over the discretization.*

4.2. Choosing a Reasonable Mesh Size

While it would be desirable to use a very small mesh size to get approximations as accurate as possible, decreasing the mesh size directly increases memory requirements and calculation time.

Thus, the goal is to find a reasonable choice between good approximation and short calculation time.

As mentioned before in Section 2.2, there exists a classical solution (2.5) to the model 2.1.1, for the special case of constant thermal properties and a shifted cosine as model function for the bottom water temperature.

Thus, we can compare this classical solution to the weak solution on different meshes and choose the broadest mesh for which the approximation error is just small enough. The thermistors used for temperature measurements in marine sediments have an accuracy of 1 mK (Dillon et al., 2012), so we aim for an approximation error smaller than 10^{-3}.

The approximation error mainly has four sources, the maximum depth x_E, the mesh size (both in space and time) and the length of the pre-modeling interval. The pre-modeling interval denotes the time between the chosen (and presumably inaccurate) initial values and the model year. The default values for the presented tests were $x_E = 20\,\text{m}$, $\delta t = 0.05\,\text{d}$, $\delta x = 0.025\,\text{m}$ and a pre-modelling interval of 6 years.

In Smerdon et al. (2006) the influence of the location of the lower boundary is discussed. For models of the surface temperature with periods of about 1 year they suggest a minimal depth of 15 m. This was taken as a first start and further investigated in the testing.

In Figure 7 we see the relative error in dependence on the lower boundary location. The depicted error is the Euclidean norm of the $k \times n^*$-matrix with the absolute values of the differences between the classical and the approximated weak solution in every point in the space-time mesh.

The error plot shows the same characteristics for all three examples, with Example 1 showing

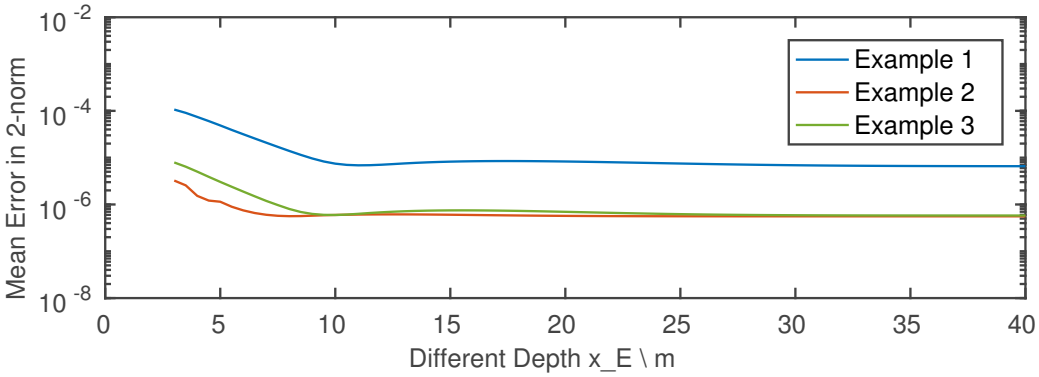

Figure 7.: Error between the classical and the approximated weak solution in the euclidean norm in dependence on the lower boundary location.

significantly higher values. The error decreases with increasing depth and we can observe a saturation for depths $x_E \geq 30\,\mathrm{m}$.

The error between the classical and the approximated weak solution in dependence on the mesh size and time steps is depicted in Figure 8. As the second derivative with respect to x is part of the model equation (2.1), the mesh size should be chosen smaller than the time steps. Thus, we changed the time steps alongside the mesh size with the relation $\delta t = \sqrt{\delta x}$ (see also Section 3.3).

Again, we see similar curves for all examples with the highest values for Example 1. The error decreases with decreasing mesh size and time steps. The curve indicates that smaller errors could be achieved if the mesh size was reduced below $\delta x = 0.005\,\mathrm{m}$.

Lastly, the approximation of the weak solution depends on the chosen initial values. The longer the model runs, the less influence of the initial conditions is visible. The pre-modeling interval is basically another model to calculate a good approximation to the initial values from an educated guess. In the setting with piecewise constant thermal diffusivity, we use the classical solution for the mean value of the measured (or given) thermal diffusivity as this educated guess. If the thermal diffusivity is highly heterogeneous, this initial condition is a rather rough approximation and the pre-modeling interval is needed to calculate better initial conditions. For the setting with constant thermal diffusivity, we used the classical solution with $\kappa_{\mathrm{init}} = 0.5 \cdot \kappa_{\mathrm{given}}$ as initial values and compared the model year to the classic solution with the given value. The resulting errors in the euclidean norm are shown in Fig.

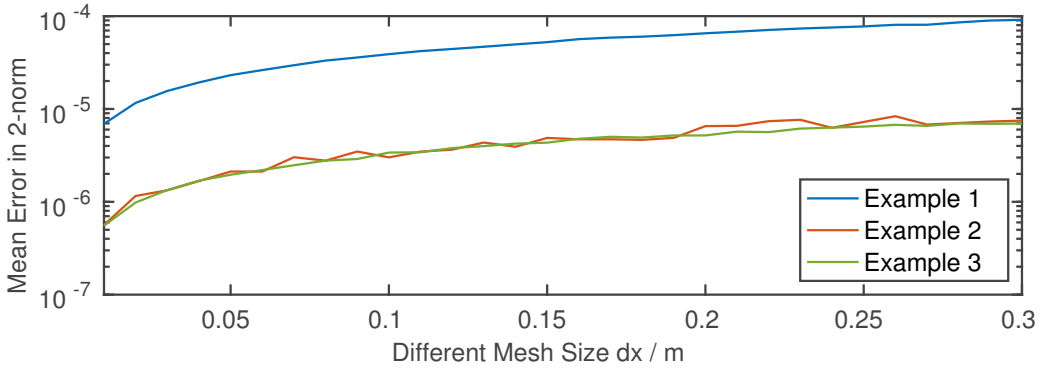

Figure 8.: Relative error between the classical and the approximated weak solution in dependence on the time steps and mesh size. The time steps are changed according to the relation $\delta t = \sqrt{\delta x}$.

9. The same similarities between the curves for the different examples can be made again. The error decreases with increasing pre-model interval. The change rate is highest between 1 and 5 years and comes to a saturation after 20 years. However, a pre-modeling interval of 20 years is physically not solid. There are periods in the scale of 10 years in the bottom water temperature, which we neglected as we are aiming for the temperatures in the last year only. If we consider a pre-modeling interval of more than 10 years, these longer periods need to be taken into account as well.

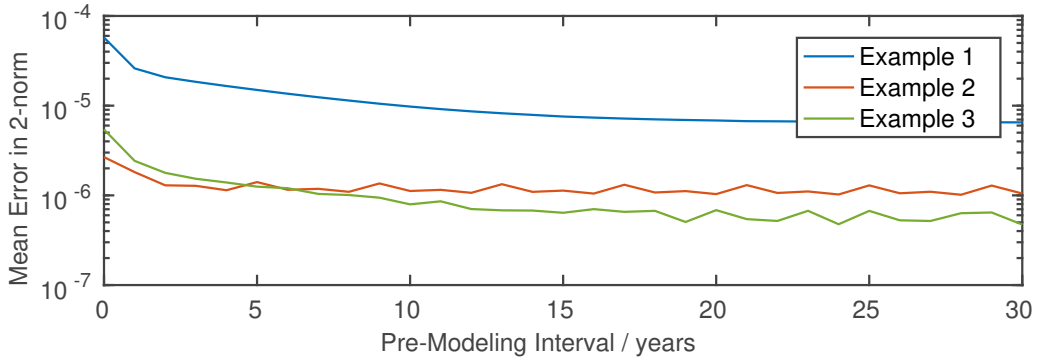

Figure 9.: Numerical error in dependence on the pre-modeling interval.

In conclusion, we can state that the numerical error of the model with the settings $x_E = 20\,\text{m}$, $\delta t = 0.05\,\text{d}$, $\delta x = 0.025\,\text{m}$, and a pre-modeling interval of 6 years is well below the accuracy of the measurement device.

5. Modeling the Bottom Water Temperature

5.1. Different Bottom Water Temperature Functions

In this section, the properties of the forward operator 2.3.1 are analyzed with respect to the bottom water temperature functions $W[f](t)$.

Shifted Cosine Function

We already saw a few examples (the classical solution (2.5) and the illustrations of the basic regions 1, 2 and 3), where the bottom water temperature was modelled using a shifted cosine,

$$W[f](t) = A + B\cos\left(\omega t + \varphi(d)\right), \tag{5.1}$$

where $f = (A, B, d)^T$, $\omega = \frac{2\pi}{365}$ and $\varphi(d) = \omega\left(\frac{365}{2} - d\right)$.

In Fig. 10, the errors in the sediment temperature are shown for varying parameters. The errors are highest for varying mean temperature A, variations in the day of the annual minimum d caused the smallest errors. For all three parameters, the relative errors are greatest for Example 1, where the temperature range is the highest.

Variations in the mean temperature A cause an error in the sediment temperature in roughly the same amount, i.e. 10 % variation in A causes 10 % error in the sediment temperature.

For the Examples 2 and 3, the influence of the amplitude B is of one magnitude lower, i.e. a change of 10 % in B causes 1 % error in the sediment temperature. For Example 1 the influence in greater; here we see a 5 % error for 10 % variation in B.

The influence of the day of the annual minimum d is generally smaller, 20 days difference causes less than 5 % for Example 1 and less than 1 % for the Examples 2 and 3.

Aperiodic Events

In areas with glaciers, such as the Baffin Bay where Example 3 was acquired, aperiodic events like the calving of a glacier can happen and thus, a short-term cold-water inflow would overlay the periodic bottom water temperature function.
We modelled such an aperiodic event as a Gaussian function with a base of thirty days and a minimum temperature of $-1\,^{\circ}\mathrm{C}$. The difference between the undisturbed bottom water temperature function (5.1) and the cold-water inflow is shown in Fig. 11. The resulting error in the data is about 0.2 %, which is in the same magnitude as the assumed measurement error.
We ran the forward model repeatedly with different combinations of inflow days and measurement days and determined the influence of the cold-water inflow in relation to the time δt between the day of the inflow and the day of the measurement. The relative difference between such two artificial measurements, which are both are taken at the same day, but one is generated with the pure cosine-function and one with a cold-water inflow is always below $2 \times 10^{-3}\,\mathrm{K}$. Although we can observe a general slight trend to a higher relative error, the closer the day of the measurement is to the inflow day, the differences are in the same magnitude of the measurement errors.
Thus, we can conclude that single cold-water inflows over less than 30 days cannot be reconstructed with this model.

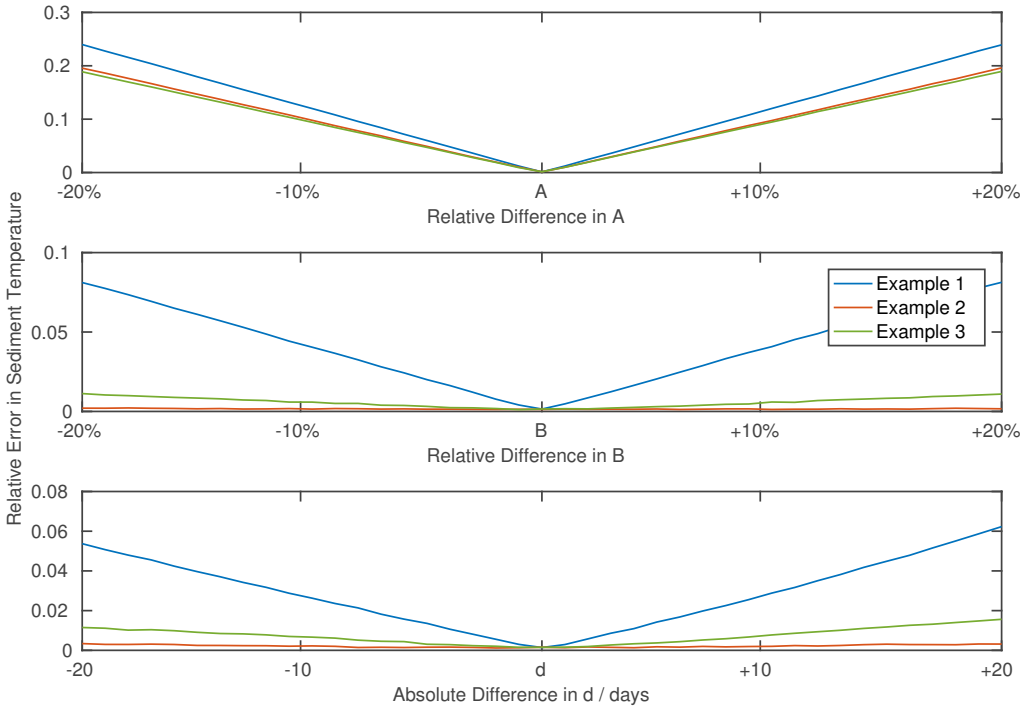

Figure 10.: Differences in the sediment temperature in dependence on variations in the water parameters A (top panel), B (middle panel) and d (lower panel). The variations in A and B are relative to the values as introduced with the examples, the variations in d are absolute days.

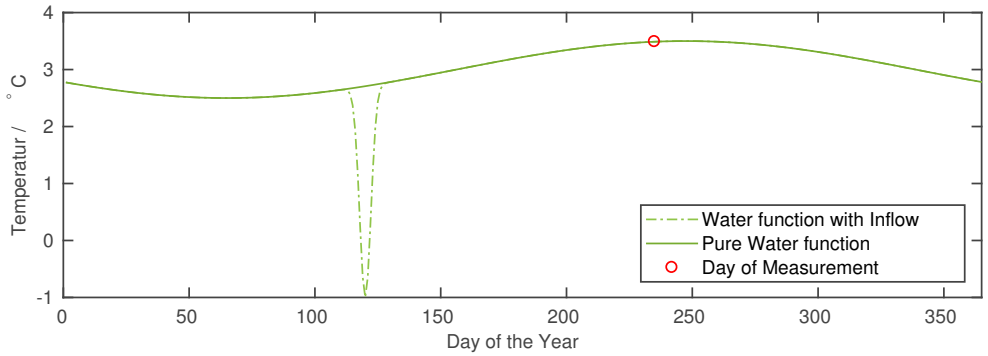

Figure 11.: A typical bottom water temperature function for the Baffin Bay as introduced in Example 3 with a cold-water inflow of 30 days with $-1\,°\mathrm{C}$.

Cut-Off Fourier Series

The shifted cosine in (5.1) can also be written as a Fourier series

$$W[f](t) = A + \sum_{n=0}^{N} B_n \cos\left(n\,\omega t + \varphi(d)\right). \tag{5.2}$$

for $N = 1$. We could expand our model to more periods, e.g. $N = 365$ would include the daily cycle. While we want to focus on the one-year period, including more periods could bring more information on the bottom water temperature.
In Fig. 12 data generated with only the first period, i.e. the bottom water temperature function is given in Equation (5.1), is compared to data generated with the first two phases, i.e. the temperature function from Equation (5.2) for $N = 2$.
In the upper panel of Fig. 12 the bottom water temperature functions are depicted over one year, in the lower panel the resulting temperature-depth-profiles are shown for a chosen day of the year.
In Fig 13, the relative difference between such data sets for varying B_2 is shown.
The difference between data with only the first period and data with the first and second phase increases as B_2 increases in relation to B_1. While $B_2 \leq 0.005 \cdot B_1$, the error stays well below 0.5 %. The greater B_1 is, the steeper the increase for $B_2 \geq 0.01 \cdot B_1$.

5.2. Analysis of Given Water Temperatures from the German Bight

The German *Federal Maritime and Hydrographic Agency* (BSH) runs a network of stations in the German North and Baltic Seas, where they measure e.g. wind velocity, water temperature and wave heights. In Fig. 14 the recorded water temperatures of the BSH station *German Bight* BSH (2014) are shown for the year 2011. Example 1 is measured in the same area as this station. Obviously the water temperature follows a quasi-periodical function with a minimum in the second half of February of approximately 3.5 °C and a maximum at the end of August of approximately 17.5 °C.
In Figure 15 the temperature measurements from the deepest sensor in 30 m water depth

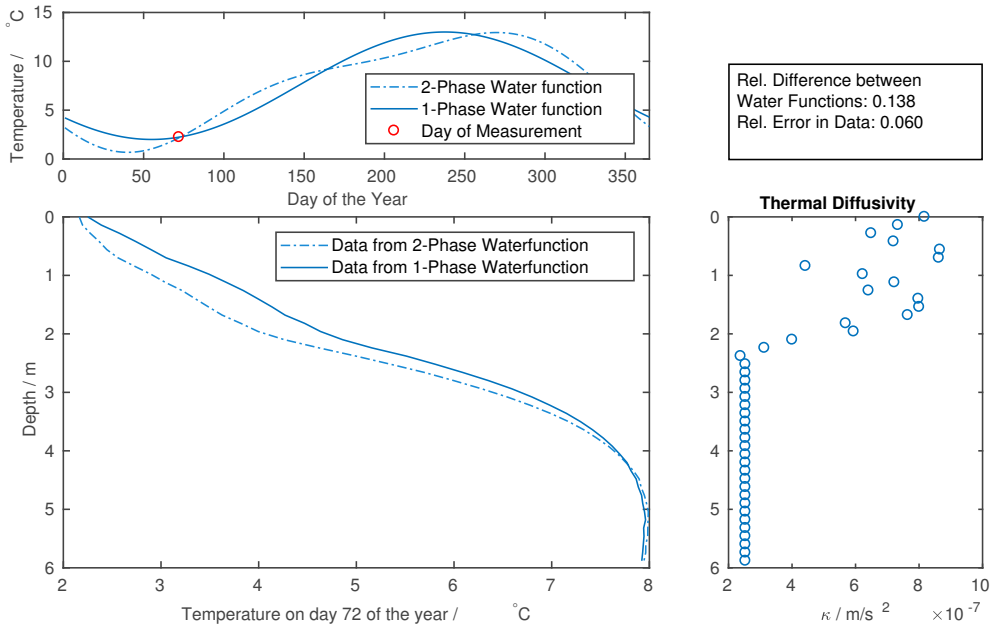

Figure 12.: Water temperature (upper panel) and resulting sediment temperature (lower panel) for a bottom water temperature function with only the first (solid line) and the first and second phase (dashed line). The parameters and the thermal diffusivity (right panel) are from Example 1 and the second phase B_2 is chosen as $B_2 = 0.3 \cdot B_1$

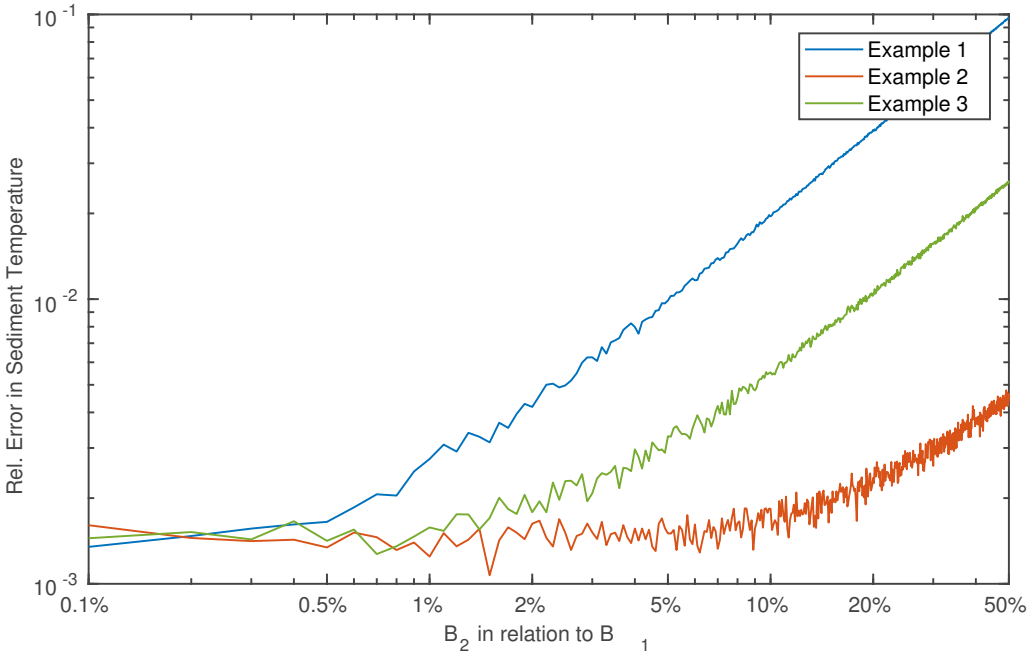

Figure 13.: Difference between data with only the first phase and data with the first two phases of the cut-off Fourier-Series (5.2) , plotted against the relative difference between the two amplitudes.

Figure 14.: Recorded water temperatures at the BSH station *German Bight* at different water depths in 2011. The dark blue lines belong to the deepest measurement 30 *m* below the sea surface and gives an idea of the temperatures on the sea floor.

from the data set depicted in Figure 14 is shown and analyzed. We therefore fitted the simple cosine function from Equation (5.1) and the Fourier series with 52 periods from Equation (5.2) with a simple Newton algorithm (see Section 7) to the data. The reconstructed parameter vector for the simple cosine function (5.1) is $f_1 = (10.5, 7.0, 51)^T$ and the reconstruction for the Fourier series (5.2) with 52 periods is $f_{52} = (10.6, 7.0, B_2, \dots, B_{52}, 52)^T$, where $|B_n| \leq 0.3\,\mathrm{K}$ for $i = 2, \dots, 52$. We see, that the measured water temperature has a sparse representation in the means of the Fourier series. The relative difference between the 3 *main* parameters A, B_1 and d in the reconstructions for one period and for 52 periods is 1.3 %. We can therefore conclude that the other periods of the Fourier series besides the one-year period B_1 are damped too much by the water column, such that they have no significant influence on the bottom water temperature, see also the oceanographic introduction in Section 3.2.

We can conclude, that the bottom water temperature function, at least in the German North Sea, can be approximated sufficiently by Equation (5.1). Therefore we will focus the reconstructions on this function, i.e. only the first two phases of the cut-off Fourier series.

5.3. The Measurement Operator and Artificial Data

The forward operator Φ from Definition 2.3.1 maps the bottom water temperature function, the geothermal gradient and the measured thermal diffusivity to a function that calculates the sediment temperature at any given depth $x \in [0, x_E]$ at time $t \in [0, T]$. The discrete forward operator $\hat{\Phi}$ from Definition 4.1.2 maps the boundary parameters $p = (A, B, d, g)$ to a matrix with temperature values for discrete depth points and discrete times.

However, the measured and processed data from the Heat Flow device (see Section 3.1) is a temperature vector $\Theta \in \mathbb{R}^s$, where $s \leq 22$ is the number of used sensors. Additionally the measurement consists of a corresponding depth vector $\hat{x} \in \mathbb{R}^s$, thermal diffusivity values $\hat{\kappa} \in \mathbb{R}^s$ and the day of the measurement t_m.

For the construction of artificial measurements, we therefore introduce the *measurement operator*.

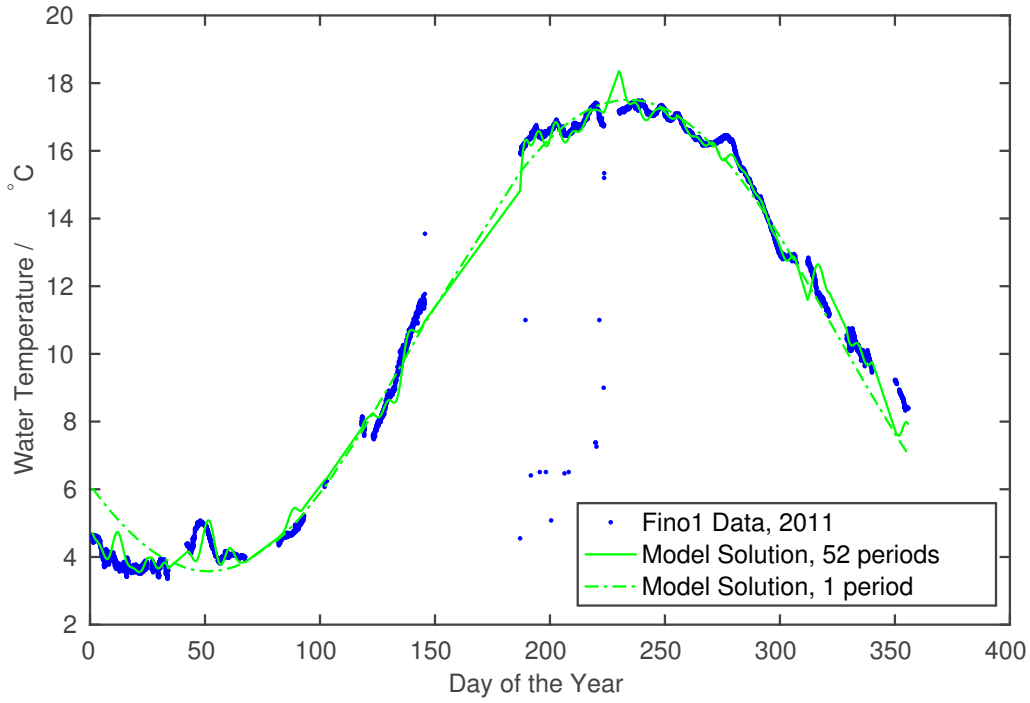

Figure 15.: Recorded water temperatures at the BSH station *German Bight* in 2011 (blue dots) and reconstructed Fourier series with 52 and one period(s) (green and dotted green lines, respectively). The relative difference between the 3 *main* parameters A, B_1 and d is 1.3 %.

Definition 5.3.1. *Measurement Operator.*
Let $\hat{x} \in [0, x_E]^s$ *with* $x_1 < x_2 < \cdots < x_s$ *and* $t_m \in [0, T]$ *be the depth-vector and time of a measurement.*
With the linear interpolation $\mathcal{P}[\hat{x}, t_m] : \mathbb{R}^{m \times n} \to \mathbb{R}^s$*, we call*

$$F(f, g) := \mathcal{P}[\hat{x}, t_m]\hat{\Phi}_{m,n}(f, g), \tag{5.3}$$

or $F(p)$ *for abbreviation, the* measurement operator*.*
Here, $\hat{\Phi}$ *is the discrete forward operator from Definition 4.1.2.*

To generate artificial data, we first generate *exact* artificial data Θ with the measurement operator 5.3.1, and then add white noise to simulate measurement errors.

$$\begin{aligned} \Theta^\epsilon &:= \Theta + e = F(f, g) + e, \\ E &\sim \mathcal{N}(0, \epsilon\|\Theta\|). \end{aligned} \tag{5.4}$$

This ensures, that the noisy data has a relative error of $\|\Theta^\epsilon - \Theta\|/\|\Theta\| \leq \epsilon$. When dealing with stochastic variable, we will use the common notation, where the capital letter, e.g. E for the error, denotes the random variable and lowercase, for the error e, denotes the realization. Normal distributed random variables with mean $\overline{e}$ and variance σ are denoted with $\mathcal{N}(\overline{e}, \sigma)$.

Realistic Artificial Data

When generating artificial data sets, we can choose the number of sensors, the depth of the lowermost sensor and the amount of added white noise as we like. To see how the methods perform on the forward model in general (see Sections 7 – 9), we used up to 50 sensors, with the lowermost sensor in 10 m and white noise between 0.5 and 5 %.

To simulate real measured data we decided to use 22 sensors, with the lowermost sensor in 5.95 m depth and 1 % white noise to generate *realistic artificial data*. These numbers are based on the configuration of the FIELAX heat flow lance and the documented measurement errors in the example data sets (see Section 3.2).

Artificial Data in the Basic Regions

In Figs. 16 – 18 we present artificial data for the basic regions. In the upper right panel of each figure, the used input parameters, i.e. the mean temperature A, the amplitude B, the day of the annual minimum d and the geothermal gradient g will be shown. The parameter sets contain the typical parameters for the bottom water temperature as introduced in Section 3.2 and the associated measured geothermal gradient. In the lower right panel, the measured thermal diffusivity values are plotted over depth. This also determines the discrete evaluation points of the measurement operator.

In the upper left panel, we see the bottom water temperature function (5.1) as defined by the used parameters. This is (together with the geothermal gradient in the lower panel) the outcome of the boundary operator $N((A, B, d), g)$ from Definition 2.3.4.

In the lower left panel the resulting temperature-depth-profile is depicted for different times of the year, which are the solutions to the weak formulation (2.8) for different fixed $t \in [0, T]$. The different t are color-coded (orange and yellow for the summer months, blue and green for the winter months) and match the colored circles in the upper left panel. The solid black line emphasizes the contribution of the geothermal gradient to the temperature field.

Additionally in the last panel, the outcome of the measurement operator (black crosses) without added noise is shown. The figures show how we get from parameters $p \in \mathbb{R}^4$ to artificial data $\Theta \in \mathbb{R}^s$, with $s \leq 22$, and what is mathematically happening in between.

We can observe, that Example 1 (Fig. 16) has the widest temperature range and that the deviation is this significantly present in 8 m depth (see also Table 2). Example 2 (Fig. 17) and 3 (Fig. 18) span a clearly smaller temperature range. In Example 2 we can additionally see how different the influence of the geothermal gradient can be. While the temperature in 10 m depth is smaller than the maximum temperature of the bottom water for Examples 1 and 3, it is approximately 0.5 K higher in Example 2.

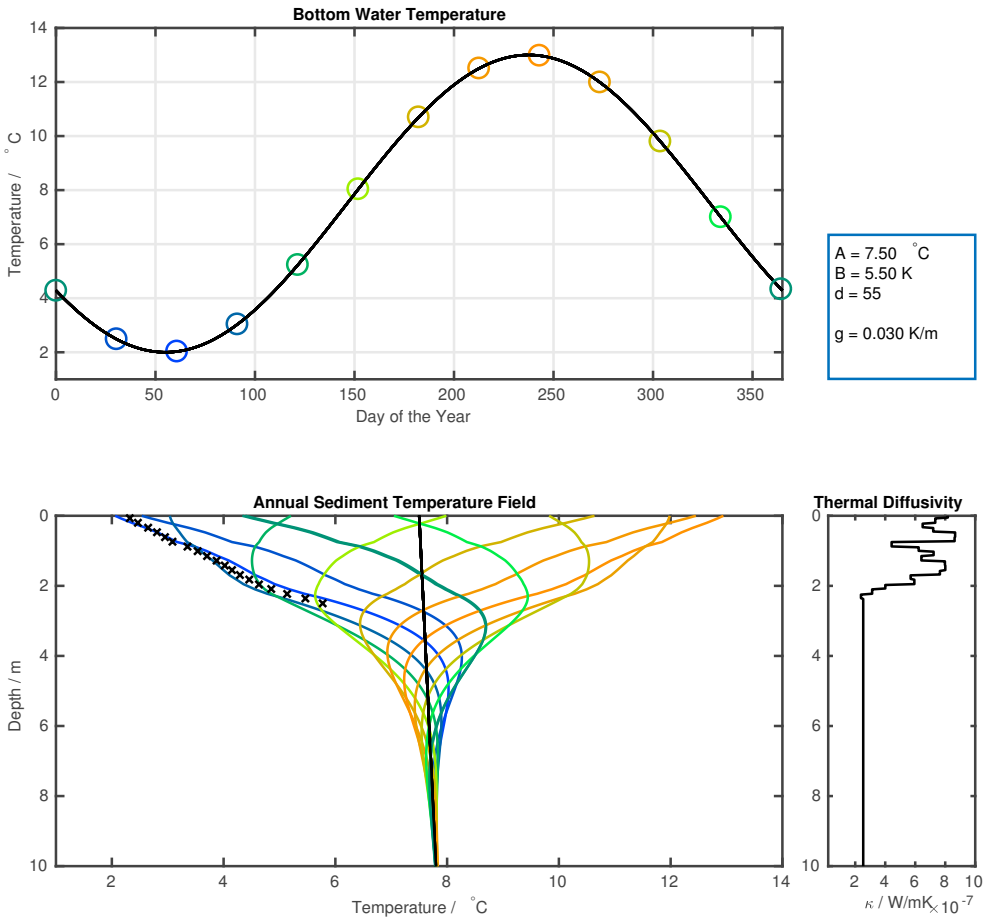

Figure 16.: Temperature Field of basic region 1 over one year.
In the top right panel, we see the input parameters, i.e. the mean temperature A, the amplitude B and the day of the annual minimum d, for the annual water temperature deviation as a shifted cosine function (top left panel). In the lower right panel, the measured (and interpolated) thermal diffusivity is shown. In the lower left panel, the temperature evolution of the sediment is shown. The colors of the temperature-depth-profiles refer to the time of the year. The solid black line indicates the geothermal gradient and the black crosses an artificial measurement.

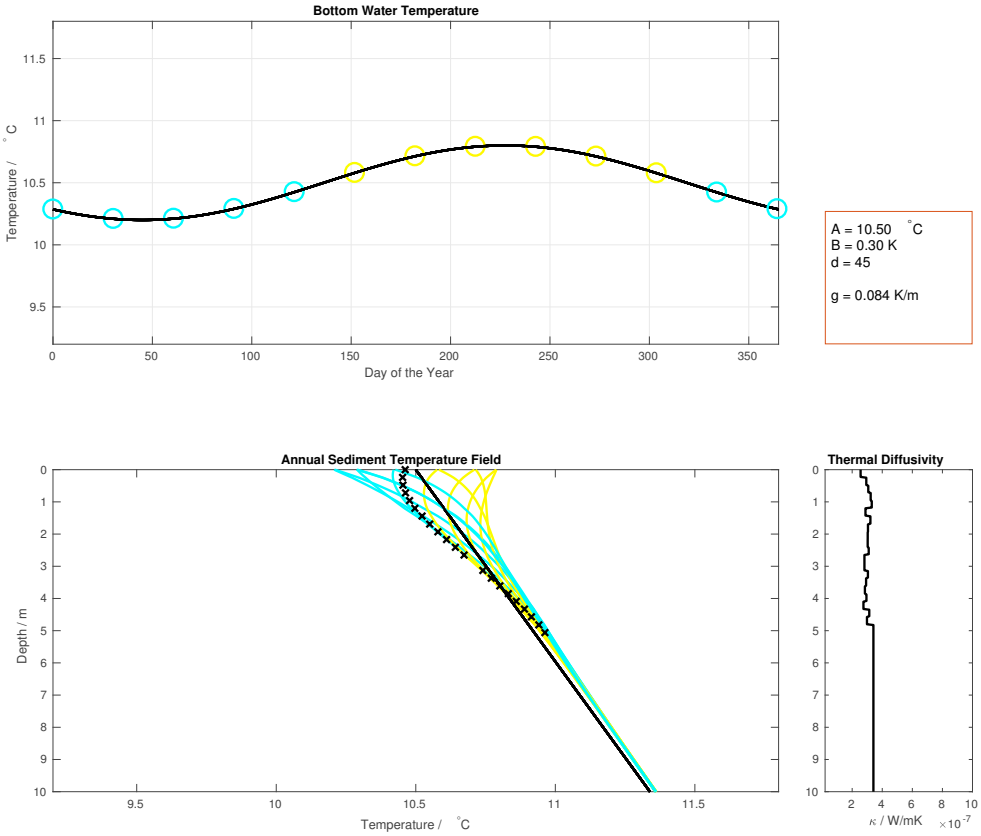

Figure 17.: Temperature Field of basic region 2 over one year.
In the top right panel, we see the input parameters, i.e. the mean temperature A, the amplitude B and the day of the annual minimum d, for the annual water temperature deviation as a shifted cosine function (top left panel). In the lower right panel, the measured (and interpolated) thermal diffusivity is shown. In the lower left panel, the temperature evolution of the sediment is shown. The colors of the temperature-depth-profiles refer to the time of the year. The solid black line indicates the geothermal gradient and the black crosses an artificial measurement.

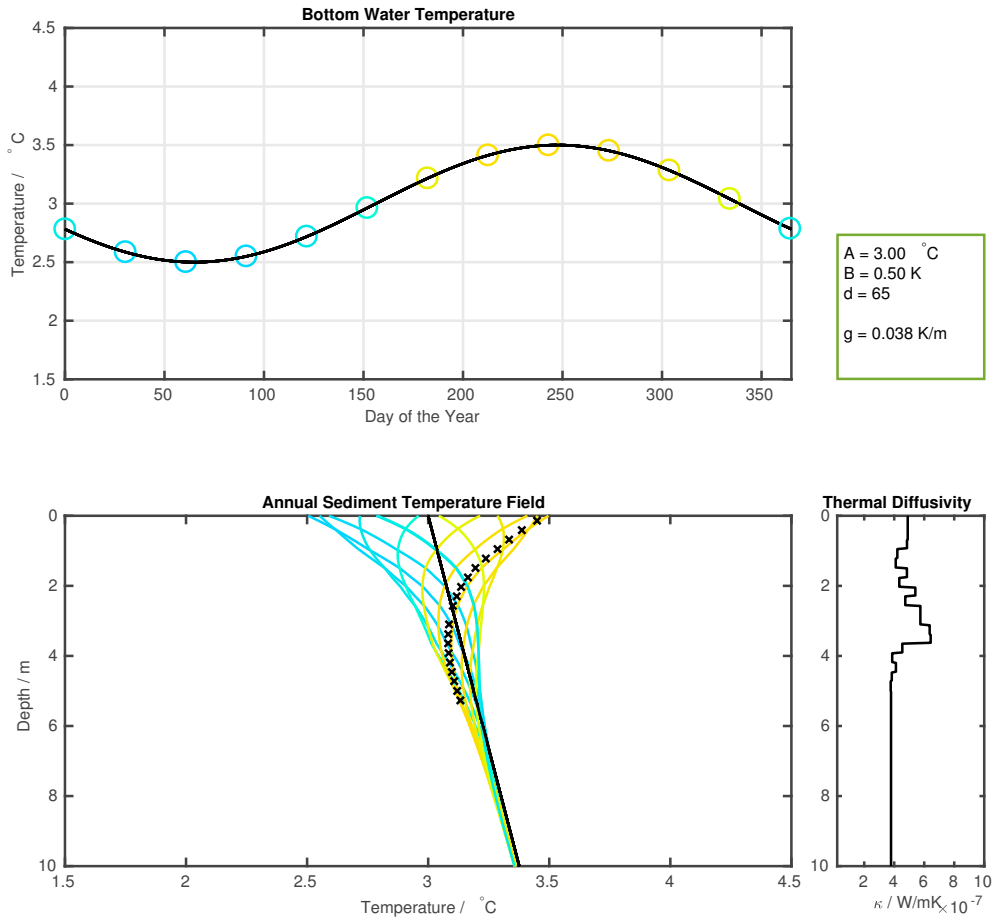

Figure 18.: Temperature Field of basic region 3 over one year.
In the top right panel, we see the input parameters, i.e. the mean temperature A, the amplitude B and the day of the annual minimum d, for the annual water temperature deviation as a shifted cosine function (top left panel). In the lower right panel, the measured (and interpolated) thermal diffusivity is shown. In the lower left panel, the temperature evolution of the sediment is shown. The colors of the temperature-depth-profiles refer to the time of the year. The solid black line indicates the geothermal gradient and the black crosses an artificial measurement.

Part II.

The Inverse Problem

6. Theoretical Attributes of the Inverse Problem

The direct problem in our setting is to calculate sediment temperatures from given thermal properties, a bottom water temperature function and the geothermal gradient. The inverse problem is to calculate the bottom water temperature function (or its parameters) and the geothermal gradient from given measurements of the sediment temperature. As outlined in Section 1, the direct problem is well examined in both mathematics and geosciences and the solutions and methods are established. The inverse problem on the other hand is only examined for settings that differ from our model in the time-scale, the linearity, or the parameterization.

The theory for our special forward operator and the resulting inverse problem is not yet widely explored. In this section, we will therefore give a short introduction to inverse problems and general approaches to get theoretical statements on solvability and uniqueness. A more thorough introduction can be found in (Engl et al., 2000) or (Rieder, 2003). We will then try to prove uniqueness for our special inverse problem.

6.1. A Short Introduction to the Theory of Inverse Problems

For the theoretical discussion of the inverse problem, we consider a continuous measurement $\Theta(x) \in H^1([0, x_E])$ at the end of the time interval $[0, T]$. The (continuous) inverse problem aims to reconstruct the water temperature over the whole time interval.

Definition 6.1.1. *The inverse problem.*
For given $g > 0$, $u_0 \in H^1([0, x_E])$, $\kappa(x)$ and continuous measurements $\Theta(x) \in H^1([0, x_E])$ the inverse problem *is to determine the bottom water temperature function $W[f](t) \in L^2([0, T])$ or its parameters f and the geothermal gradient $g > 0$ such that*

$$S(W, g)(x, T) = \Theta(x).$$

Note that S is the solution operator from Definition 2.3.1.

Inverse problems are called *ill-posed*, if there exists data for which there is no or no unique solution to the inverse problem or if the solution does not depend continuously on the data (Engl et al., 2000). When dealing with real measured data, this data is most definitely noisy and thus, we cannot expect a unique solution. However, we can expect to find a best-approximate solution that hopefully converges to the unique solution for decreasing noise level (Engl et al., 2000). This best-approximate solution is calculated with the (Moore-Penrose) generalized inverse for linear forward operators and is unique. Thus, the well-posedness is a question of the continuity of the generalized inverse. Unfortunately, this is not the case for most forward operators from physical applications and regularization schemes come into play to construct approximations that depend continuously on the data (Engl et al., 2000). For linear forward operators, the theory is very well developed Engl et al. (2000). For non-linear operators on the other hand, the theory is more complicated as every non-linear problem is non-linear in its own way. A minimum-norm solution can be defined but does not need to exist or to be unique. A strategy to solve non-linear inverse problems can be to linearize the problem and solve the linear problem with existing theory, or to build new theory for the special problem at hand.
To see if our forward operator is ill-posed, we need to examine if a solution to the inverse problem is unique and if it depends continuously on the given data, without having a detailed look at the inversion process. Note, that the solution operator S from Definition (2.3.3) is linear, while the forward operator F from Definition 2.3.1 is not.
For linear operators it can be shown, that all compact operators lead to ill-posed problems (Engl et al., 2000), for non-linear problems there is no such easy criteria and the ill-posedness needs to be examined for each problem individually.

For two water temperature functions $W_1(t), W_2(t) \in L^2([0, T])$ and two geothermal gradients $g_1, g_2 > 0$ we get two unique solutions of the forward model $u_1(x, t)$ and $u_2(x, t)$. Now, we assume that the two temperature models are the same at the end of the time interval, i.e. $u_1(x, T) = \Theta(x) = u_2(x, T)$.
Thus the difference $v(x, t) := u_1(x, t) - u_2(x, t)$ is a solution to the weak formulation (2.8)

$$\tilde{g}\,\psi(x_E) + \tilde{h}\tilde{W}(t)\psi(0) = \partial_t \int_0^{x_E} \mathbf{v}^t(x)\psi(x)\,dx + \tilde{h}\mathbf{u}^t(0) \tag{6.1}$$

$$+ \int_0^{x_E} \kappa(x)\partial_x \mathbf{u}^t(x)\partial_x\psi(x)\,dx. \tag{6.2}$$

for all $\psi \in H^1([0, x_E])$ and $t > 0$, with $\tilde{g} = g_1 - g_2$ and $\tilde{W} = W_1 - W_2$ and the additional condition that $u(x, T) = 0$.
Now, if the above integral equation and end-value condition only have a unique solution if $\tilde{W}(t) \equiv 0$, then the inverse problem has a unique solution for every measurement $\Theta(x)$. And if this unique solution additionally depends continuously on the data the inverse problem would be well-posed.

6.2. Eigenvalues of the Forward Operator

In this chapter, we want to establish assumptions under which the forward model for piecewise constant thermal diffusivity has a unique solution.
For $u : [0, T] \times [0, x_E] \to \mathbb{R}$, being the solution to the system (2.1) – (2.3), we recall the definition in (2.6)

$$\mathbf{u}^t(x) = u(x, t) \in H^1([0, x_E]). \tag{6.3}$$

In the following, we want to derive a formulation of $\mathbf{u}^t$ in terms of the eigenvalues of the forward operator. Therefore, we cite the following result from McLean (2000).

Theorem 6.2.1. *(from (McLean, 2000, Thm. 2.36))*
Let H act as a pivot space for V, and assume that H is infinite-dimensional and that the inclusion map $V \to H$ is compact. If the bounded linear operator $A : V \to V'$ is self-adjoint and coercive, then there exists a sequence of vectors $\phi_1, \phi_2, \ldots$ in V and a sequence of real numbers $\lambda_1, \lambda_2, \ldots$ having the following properties:

(i) For each $j \geq 1$, ϕ_j is an eigenvector of A with eigenvalue λ_j.

(ii) The eigenvectors $\phi_1, \phi_2, \ldots$ form a complete orthonormal system in H.

(iii) The eigenvalues satisfy $\lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \ldots$ with $\lambda_j \to \infty$ as $j \to \infty$.

(iv) For each $u \in V$,

$$Au = \sum_{j=1}^{\infty} \lambda_j \langle \phi_j, u \rangle \phi_j \qquad \text{(convergence in } V'\text{)}.$$

To make use of this theorem, we identify $L^2([0, x_E])$ with H and $H^1([0, x_E])$ with V. That makes $V' = H^1([0, x_E])'$.

We now identify the operator $A\colon H^1([0, x_E]) \to H^1([0, x_e])'$ with the bilinear form (2.7), such that $A : u \mapsto a[u, \cdot]$. In the proof to Theorem 2.2.3 we already showed that the biliear form a is coercive. To show that A is also self-adjoint, we note that $H^1([0, x_E])'' \equiv H^1([0, x_E])$. With the symmetry of the bilinear form it follows, that

$$\begin{aligned} \langle Au, \phi \rangle_{H^{1\prime} \times H^1} &= A[u](\phi) = a(u, \phi) \\ &= a(\phi, u) = A[\phi](u) \\ &= \langle u, A^* \phi \rangle_{H^1 \times H^{1\prime}}. \end{aligned}$$

Thus, we meet the assumptions of Theorem 6.2.1 and therefore get a set of eigenfunctions $\phi_j \in H^1([0, x_e])$ and corresponding eigenvalues λ_j, such that

$$A[u](\phi_j) = a[u, \phi_j] = \lambda_j \langle \phi_j, u \rangle \qquad \forall\, u \in H^1([0, x_E]).$$

Furthermore, from Theorem 6.2.1 (iv) we get that

$$a[u, \psi] = A[u](\psi) = \sum_{j=1}^{\infty} \lambda_j \langle \phi_j, u \rangle \int_0^{x_E} \phi_j \psi \, dx. \tag{6.4}$$

Now, using any of the eigenfunctions $\phi_l \in H^1([0, x_e])$ as testfunction in the weak formulation

(2.8) and substituting the bilinearform with the representation (6.4), we get

$$\begin{aligned} g\,\phi_l(x_E) + \kappa(0)hW(t)\phi_l(0) &= \partial_t\langle \mathbf{u}^t, \phi_l\rangle + a[\mathbf{u}^t, \phi_l] \\ &= \partial_t\langle \mathbf{u}^t, \phi_l\rangle + \sum_{j=1}^{\infty} \lambda_j \langle \phi_j, \mathbf{u}^t\rangle\langle \phi_j, \phi_l\rangle. \end{aligned}$$

The eigenfunctions ϕ_j are orthonormal, which leaves us with

$$g\,\phi_l(x_E) + \kappa(0)hW(t)\phi_l(0) = \partial_t \mathbf{u}_l(t) + \lambda_l \mathbf{u}_l(t) \tag{6.5}$$

for all $l = 1, 2, \ldots$ and for fixed $t > 0$, where $\mathbf{u}_l(t) = \langle \mathbf{u}^t, \phi_l\rangle$. This ordinary differential equation is then solvable for every l and has the solution

$$\mathbf{u}_l(t) = \mathbf{u}_l(0)\mathrm{e}^{-\lambda_l t} + \mathrm{e}^{-\lambda_l t}\int_0^t \left(g\,\phi_l(x_E) + \kappa(0)h\phi_l(0)W(\tau)\right)\mathrm{e}^{\lambda_l \tau}\,d\tau. \tag{6.6}$$

This lets us formulate the following uniqueness result.

Theorem 6.2.2.
If we can show, that the eigenfunctions of the bilinear form have different boundary values, then that implies that the inverse problem 6.1.1 has a unique solution for bottom water temperature functions $W(t)$ of the form of a shifted cosine (5.1)*, and continous measurements $\Theta(x)$ at time $T > 0$.*

Proof. As established before, we only show that $\Theta(x) = 0$ yields $W(t) = 0$ and $g = 0$. Because of the periodicity of $W(t)$ and therefore of the solution $u(x, t)$ of the forward model 2.1.1, we can shift the time variable t, such that T equals this period and thus $u(x, 0) = \Theta(T) = 0$. With Theorem 6.2.1 we find that for fixed $t > 0$, the solution solves (6.6) for all pairs of eigenfunctions and eigenvalues (ϕ_l, λ_l). With the measurements $\Theta(x) = 0$ at $t = T$ and the periodicity, this yields

$$0 = \Theta(x)\mathbf{u}_l(T) = \mathrm{e}^{-\lambda_l T}\int_0^T \left(g\,\phi_l(x_E) + \kappa(0)h\phi_l(0)W(\tau)\right)\mathrm{e}^{\lambda_l \tau}\,d\tau.$$

This equality holds if and only if the integral is zero. For the bottom water temperature function as in (5.1), the integral is analytically solvable, which brings us

$$
\begin{aligned}
-\int_0^T g\,\phi_l(x_E)\mathrm{e}^{\lambda_l\tau}\,d\tau &= \kappa(0)h\phi_l(0)\int_0^T W(\tau)e^{\lambda_l\tau}\,d\tau \\
\Leftrightarrow\ -g\left\{\frac{\phi_l(x_E)(\mathrm{e}^{\lambda_l T}-1)}{\lambda_l}\right\} &= \kappa(0)h\phi_l(0)\int_0^T (A+B\cos(\omega\tau+\varphi(d)))\,\mathrm{e}^{\lambda_l\tau}\,d\tau \\
&= A\left\{\frac{\kappa(0)h\phi_l(0)(\mathrm{e}^{\lambda_l T}-1)}{\lambda_l}\right\} \\
&\quad + B\left\{\frac{\kappa(0)h\phi_l(0)}{\lambda(1+\omega\lambda_l)}\,(-\cos(\varphi(d))-\omega\sin(\varphi(d)) \right. \\
&\quad \left. +(\cos(\omega T+\varphi(d))+\omega\sin(\omega T+\varphi(d))\mathrm{e}^{\lambda_l T})\right\}
\end{aligned}
$$

for all eigenpairs (λ_l, ϕ_l).
If now $\phi_l(0) \neq 0 \neq \phi_l(x_E)$ for at least three eigenfunctions, it follows directly that $A = B = g = 0$, which yields $g = 0 = W(t)$ and therefore uniqueness of the inverse problem. □

Remark 6.2.3. Be aware, that we neither proofed ill- nor well-posedness of our inverse problem, but only explored possible ways to do so. More analytical investigations in this direction would be interesting but exceed the scope of this work. We just note, that the inverse problem is at least *not hopelessly ill-posed* and are satisfied with the fact that the inversion schemes yield reasonable results.

7. Newton-like Inversion Methods

In this chapter, we will introduce different approaches for the inversion of the forward operator. For a general introduction to inverse problems and regularization schemes, we refer to Engl et al. (2000).
The used methods can be categorized into three different classes, which are

- Newton-like methods,
- Linear fitting, and
- Markov Chain Monte Carlo methods.

We first tackle the inverse problem by Newton-like methods: a simple Newton method and a statistical approach to the classic Newton method were implemented and analyzed, of which the first was already discussed and compared to the regularized inexact Newton method (REGINN) from Rieder (1999) by Miesner et al. (2015).
For the following chapters, we consider given data $\Theta^\epsilon \in \mathbb{R}^s$ with a relative noise level, such that $\|\Theta^\epsilon - \Theta\|/\|\Theta\| \leq \epsilon$, where the exact data Θ is unknown. The presented methods approximate the exact parameters $p^+ \in \mathbb{R}^{k+1}$, for which $F(p^+) = \Theta$. The methods are then rated for the relative reconstruction error $\|p - p^+\|/\|p^+\|$ and the relative discrepancy $\|F(p) - F(p^+)\|/\|F(p^+)\|$.

Newton-like Methods and the Discrepancy Principle

The idea of the Newton-like methods is rather simple. We rewrite the model equation $F(p) = \Theta$, such that

$$F(p) - \Theta = 0. \tag{7.1}$$

In order to find the parameter set p^+, we now have to find the root of $F(p) - \Theta$. For the basic Newton algorithm this is done in an iterative process. We consider the ongoing iteration, where the next iterate is constructed as follows:

$$p_{n+1} = p_n + s_n. \tag{7.2}$$

The ideal choice now would be $s_n^+ = p^+ - p_n$. As we do not know p^+, we can only aim for a good approximation of s_n^+.
We already know that the forward operator is Fréchet-differentiable and thus

$$F'(p_n)s_n^+ = \Theta - F(p_n) - \mathcal{E}(p^+, p_n) := b_n$$

holds, where $\mathcal{E}$ is the linearization error. Again, we do not know b_n exactly as we do not know the linearization error or the exact data Θ. We therefore denote the disturbed data as well as the variables calculated from disturbed data with a superscript ϵ. In the disturbed equation

$$F'(p_n^\epsilon)s_n^+ = \Theta^\epsilon - F(p_n^\epsilon) = b_n + \mathcal{E}(p^+, p_n^\epsilon),$$

we now know all variables on the left-hand side, and can approximate the error on the right-hand side by b_n^δ with $\|b_n - b_n^\delta\| \leq \delta \leq \epsilon + \mathcal{E}(p^+, p_n)$.
Finally, we calculate the Newton-step from

$$F'(p_n^\epsilon)s_n^\epsilon = b_n^\delta. \tag{7.3}$$

The iteration is stopped by the discrepancy principle.

Definition 7.0.1. *Discrepancy principle of Morozov (Engl et al., 2000).*
For a given noise level $\epsilon > 0$, choose the stopping index n^* such that

$$\frac{\|F(p_{n^*}) - \Theta^\epsilon\|}{\|\Theta^\epsilon\|} \leq \tau\epsilon < \frac{\|F(p_n) - \Theta^\epsilon\|}{\|\Theta^\epsilon\|}$$

for all $n < n^*$ and $\tau > 1$.

The discrepancy principle stops the iteration as soon as the relative discrepancy $\|F(p_n) - \Theta^\epsilon\|/\|\Theta^\epsilon\|$ between the data and the current approximation is in the same magnitude as the

(assumed) noise level of the data ϵ. This ensures, that we do not try to approximate the data *better* than it can be measured.
In the following sections, we will see four different approaches to solve the linear equation for the Newton-step (7.3). First, we will show and discuss the results from Miesner et al. (2015) on the simple Newton algorithm. After that we expand the Newton algorithm and split the derivative into two parts in order to be able to additionally reconstruct the geothermal gradient. Finally, we will introduce a statistical scheme to determine the *most probable* Newton step.

7.1. Basic Newton Algorithm

This section is essentially a summary of what has been already discussed by Miesner et al. (2015) for the basic Newton algorithm and the REGINN method. As the characteristics of the results does noes not differ much between the two methods (Miesner et al., 2015), we will only show the basic Newton algorithm here.
In the simple Newton algorithm (Algorithm 1) the Newton step s, as in Equation (7.3), is calculated via the MATLAB inbuilt backslash-function. In Miesner et al. (2015) the Newton algorithm is only used to reconstruct the parameters of the bottom water temperature function $f = (A, B, d)^T$ and so is the formulation of the algorithm here. However, the algorithm can be formulated to also reconstruct the geothermal gradient. The results for this extended algorithm are later shown alongside the results of the basic Newton algorithm.

Algorithm 1 The Newton Algorithm

Input: Θ^ϵ, ϵ, g, τ, $i_{\max}$
Initialize: f
  $b = \|F_g(f) - \Theta^\epsilon\| / \|\Theta^\epsilon\|$
  $i = 1$
  while $b > \epsilon\tau$ **and** $i \leq i_{\max}$ **do**
    $i = i + 1$
    $s = F_g'(f)\ b$
    $f = f - s$
    $b = \|F_g(f) - \Theta^\epsilon\| / \|\Theta^\epsilon\|$
  end while
  return f

The simple Newton algorithm has no regularizing properties. For the simple forward operator with the bottom water temperature function as in (5.1) the Fréchet derivatives are linear independent and thus, the simple Newton algorithm is sufficient.

Results for Noisy Artificial Data

The Newton algorithm, the REGINN method and their results on artificial data are already discussed in Miesner et al. (2015). There, exact parameters in the range of Example 1 were used. It was shown that data with 1 % additive noise (see Section 5.3) yielded a relative reconstruction error of about 1.1 % for the mean value of 40 reconstructions with both algorithms.

For this study, we tested the Newton algorithm and the extended Newton algorithm for different artificial data sets of all three examples. In the Appendix, the results are shown in Tables 6 – 9) are shown. For each setting, the algorithms were executed 100 times. In the tables we see the mean values of the 100 repetitions and the standard deviation (SD), as a measure for the stability of the results, as well as the relative reconstruction error. As the algorithms are stopped with the discrepancy principle from Definition 7.0.1, the achieved discrepancy is not given in the tables.

The results from Miesner et al. (2015) could be achieved with the example regions 1 and 3, where the relative difference between the mean temperature A and the amplitude B is $B = 0.73 \cdot A$ and $B = 0.16 \cdot A$ respectively. For the example region 2, this relative difference is $B = 0.05 \cdot A$ and the results are significantly worse, especially in the reconstruction of the amplitude B. As shown by Miesner et al. (2015), it is difficult to reconstruct the day of the annual minimum d accurately within the range of two weeks, since the first derivative of the cosine is very small in a neighborhood around the extrema. In general the results show, that the uncertainties in the reconstruction of the day of the annual minimum increase with decreasing relative difference between mean temperature and amplitude of the bottom water temperature. This effect can be shown for all examples. In Fig. 19, the relative reconstruction error (upper panel), the relative reconstruction errors only in the amplitude B (middle panel) and only in the day of the annual minimum d (lower panel) are plotted against the relation of the amplitude B to the mean temperature A. The data used for this test have 1 % white noise added.

We see, that the overall reconstruction error decreases significantly for $B \geq 0.04 \cdot A$ and tends to zero for $B \approx 0.06 \cdot A$ for Example 1, $B \approx 0.07 \cdot A$ for Example 2, and $B \approx 0.09 \cdot A$ for Example 3. The relative difference in the reconstruction of B and d decreases until it reaches about 0.5 % for $B \approx 0.9 \cdot A$.

For the extended Newton algorithm (which also reconstructs the geothermal gradient) we see slightly worse reconstruction results, but are still in the same error bounds as with the basic Newton algorithm.

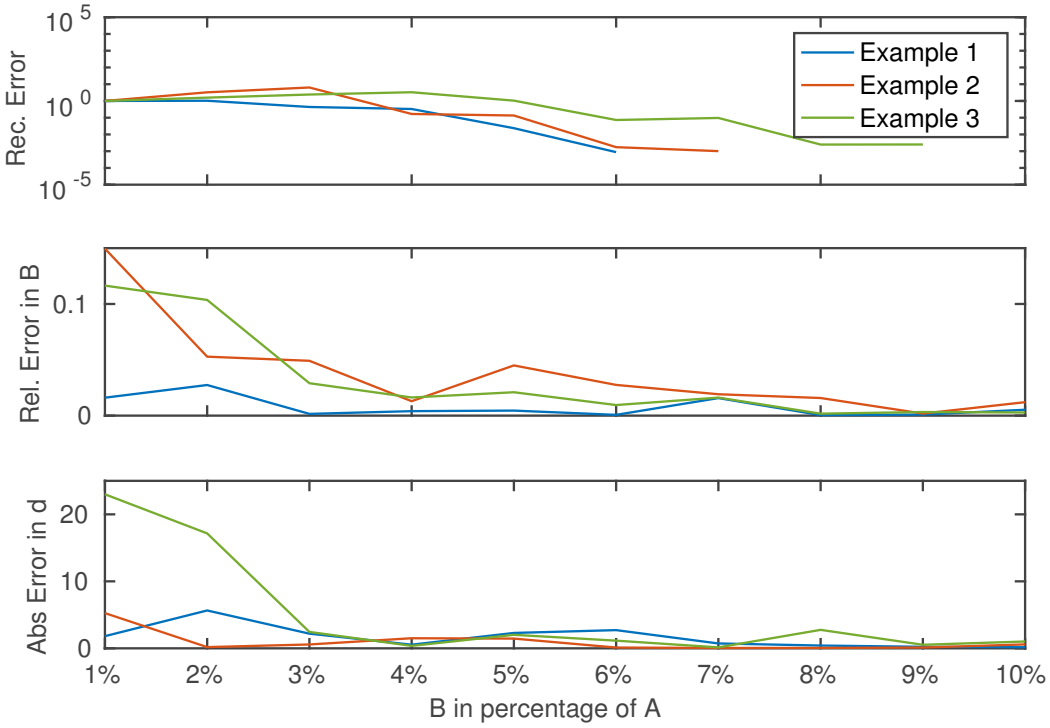

Figure 19.: Relative reconstruction error (upper panel) against the relation of B to the mean temperature A. Additionally the relative differences in the reconstruction of the amplitude B (middle panel) and the day of the annual minimum d (lower panel) are shown.

In conclusion, the Newton algorithm yields results with errors in the same magnitude as the initial noise level on the data for the mean temperature A for noise levels smaller than 1 %. This error bound also holds for the amplitude B, if it is not smaller than about $0.07 \cdot A$. The day of the annual minimum can then be reconstructed within a range of 10 days.

7.2. A Statistical Newton Method

In this section, we follow Kaipio and Somersalo (2004) to derive a method for linear statistical inverse problems. We will then apply this method to the Newton step (7.3) to solve our non-linear problem. For the notation of random variables, we will use capital letters for the random variable and lowercase letters for the realization. Note, that for a Gaussian random variable Z, with a variance σ^2 around an expected value z_e, i.e. $Z \sim \mathcal{N}(z_e, \sigma^2)$, a realization z has the probability of $\pi(z) = 1/\sqrt{\sigma^2 2\pi}\exp\{-1/2((z - z_e)/\sigma)^2\}$.

Statistical Inversion Theory for Linear Problems

We consider the linear problem $AX = Y$ with random variables X and Y. For the measured data $y^+ \in \mathbb{R}^k$ of the problem, we assume, that an input $x^+ \in \mathbb{R}^s$ and some error e exist, such that $y^+ = Ax^+ + e$. Any additional a-priori information on X is given via the prior density function $\pi_{\text{pr}}(x)$.
Now, the goal of statistical inversion theory is to find the *most probable* input x to have caused the observation y_{obs}. Thus, the *posterior probability density* π_{post} needs to be analyzed.

Theorem 7.2.1. *Bayes' Theorem (see e.g. Kaipio and Somersalo (2004)).*
Assume a random variable X with realizations $x \in \mathbb{R}^k$ has known prior probability density $\pi_{pr}(x)$ and the given data is the observed value $y_{obs} \in \mathbb{R}^s$ of an observable value Y, s.t. $\pi(y_{obs}) > 0$. Then the posterior probability density of X, given the data y_{obs} is

$$\pi_{post} = \pi(x|y_{obs}) = \frac{\pi_{pr}(x)\pi(y_{obs}|x)}{\pi(y_{obs})},$$

where $\pi(y_{obs}|x)$ is the likelihood function of y_{obs}, given x.

Note, that Bayes' Theorem is also valid for non-linear problems.
The *most probable* input x can then be approximated with different estimators: maximum a posteriori value, conditional mean, maximum likelihood, conditional covariance. Following (Kaipio and Somersalo, 2004), we will use the conditional mean $x_{CM} = \int_{\mathbb{R}^k} x\pi_{post}(x)\,dx$ in the following. Thus, in order to calculate this integral, we need to obtain an analytic formulation for the posterior probability density.

For $x \in \mathbb{R}^s$, the variance of the random variable is denoted with $\Gamma \in \mathbb{R}^{s\times s}$. Assuming $X \sim \mathcal{N}(x_e, \Gamma_{\text{pr}})$ and $E \sim \mathcal{N}(0, \Gamma_{\text{noise}})$, i.e. we can write

$$\pi_{\text{pr}}(x) \propto \exp\left(-1/2(x-x_e)^T\Gamma_{\text{pr}}^{-1}(x-x_e)\right),$$
$$\pi_{\text{noise}}(e) \propto \exp\left(-1/2(e)^T\Gamma_{\text{noise}}^{-1}(e)\right).$$

We further assume, that X and E are mutually independent, which means, that the probability density of E does not change for different given parameters x (see Kaipio and Somersalo (2004)). In consequence, the likelihood function is given by

$$\pi(y_{\text{obs}}|x) = \pi_{\text{noise}}(y - Ax).$$

This yields

$$\begin{aligned}\pi_{\text{post}}(x) &\propto \pi_{\text{pr}}(x)\pi_{\text{noise}}(y-Ax)\\ &= \exp\left(-1/2(x-x_e)^T\Gamma_{\text{pr}}^{-1}(x-x_e) - 1/2(y-Ax)^T\Gamma_{\text{noise}}^{-1}(y-Ax)\right)\end{aligned}$$

for the posterior probability density.
With some matrix calculation we end with explicit formulas for the posterior density and the conditional mean

$$\Gamma_{\text{post}} = \left(\Gamma_{\text{pr}}^{-1} + A^T\Gamma_{\text{noise}}^{-1}A\right)^{-1}, \tag{7.4}$$
$$x_{CM} = \Gamma_{\text{post}}\left(A^T\Gamma_{\text{noise}}^{-1}b + \Gamma_{\text{pr}}^{-1}x_e\right). \tag{7.5}$$

Choosing Priors for Our Model

We recall the Newton-step (7.3),

$$F'(p_n^\epsilon)s_n^\epsilon = b_n^\delta.$$

Here, $F'(p_n^\epsilon)$ will play the role of the linear operator A, the Newton step s_n^ϵ is the input x and the right-hand side $b_n^\delta = b_n + e_n$ is the observed value y_{obs}. Thus, we need to fix a-priori densities for the step s_n^ϵ and additive noise on this step e_n.
We therefore need to determine the a-priori densities for the parameter vector p and the

data Θ^ϵ. For the data the uncertainties are given from the measurements, i.e. the post-processing, as $E \sim \mathcal{N}(0, \Gamma_{\text{noise}})$, where Γ_{noise} is a diagonal matrix such that $\Theta^\epsilon = \Theta + e$ with $\|\Theta^\epsilon - \Theta\|/\|\Theta\| \leq \epsilon$. For the parameter vector p we will use a similar approach and assume $P \sim \mathcal{N}(p_e, \Gamma_{\text{pr}})$. The entries of the diagonal matrix Γ_{pr} will be a (guessed) input to the inversion scheme.
Now, the Newton-step s_n^ϵ is the approximation of the ideal step $s_n^+ = p^+ - p_n$. With the given probability density of p^+, we can assume that

$$S_n^\epsilon \sim \mathcal{N}(p_e - p_n, \Gamma_{\text{pr}}).$$

Further, the disturbed right-hand side is defined as $\Theta^\epsilon - F(p_n^\epsilon) = b_n^\delta + \mathcal{E}(p^+, p_n^\epsilon)$, where $\mathcal{E}$ is the linearization error. Assuming that the linearization error behaves similar to the white noise on the data, we can write

$$E_n^\delta \sim \mathcal{N}(0, \Gamma_{\text{noise}}).$$

Applying the Theory to the Newton-Step

Applying the results from the last two pages to the Newton algorithm 1, we can formulate the *statistical Newton algorithm*.

Algorithm 2 A Statistical Newton Algorithm.

Input: Θ^ϵ, ϵ, τ, $i_{\max}$, Γ_{pr}, Γ_{pr}
Initialize: $p = (f, g)$
  $j = 0$
  $b = \|F(p) - \Theta^\epsilon|/\|\Theta^\epsilon\|$
  while $b > \epsilon\,\tau$ **do**
    $A = F'(p)$
    $b = \Theta^\epsilon - F(p)$
    $s_e = p_e - p$
    $\Gamma_{\text{post}} = (\Gamma_{\text{pr}}^{-1} + A^T\,\Gamma_{\text{noise}}^{-1}\,A)^{-1}$
    $y = A^T\,\Gamma_{\text{noise}}^{-1}\,b + \Gamma_{\text{pr}}^{-1}\,s_e$
    $s = \Gamma_{\text{post}} y$
    $p = p + s$
    $b = \|F(p) - \Theta^\epsilon\|/\|\Theta^\epsilon\|$
  end while
  return p

Numerical Results for Noisy Artificial Data

The reconstruction results of the statistical Newton algorithm depend on the chosen prior covariance matrix Γ_{pr}. With fine-tuning stunningly accurate results can be achieved. However, for real measurements the a priori information is rarely very good. The presented results in the Tables 11 – 14 are obtained from broad prior probability densities. Here we see the same characteristics as for the simple and extended Newton algorithms: The mean temperature A is best reconstructed, the amplitude B is also quite good reconstructed, except for Example 2, the day of the annual minimum d is reconstructed within 15 days and the geothermal gradient is the most difficult to reconstruct. For Example 1 we see an underestimation for the mean temperature A, while all other parameters slightly overestimated. For Example 2 the reconstructed geothermal gradient is negative and four magnitudes overestimated, while the other parameters are reconstructed in reasonable error bounds. Only for Example 3 the reconstruction is usable with a mean error of 15.7 %.
The standard deviation (SD) is zero for 100 executions, as we used the same prior probability density but varying initial guesses. It is unclear how to model the uncertainties of prior probability densities for real measurements and thus, the investigation of the influence of varying prior probability densities is definitely necessary but also time consuming.

8. Linear Fitting

In Section 7.1 we saw that the simple and extended Newton algorithms 1 yield good results for artificial data. As seen in Section 7.2 we saw that the statistical Newton algorithm 2 did not add more accuracy to the reconstruction when we have poor a priori knowledge.
This is the reason why another approach is needed. The idea of the linear fitting is to generate a *basis of measurements* with the forward operator and basis vectors of the parameter space and then solve the matrix equation with Tikhonov regularization.

8.1. Linear Fitting with Tikhonov Regularization

Generating the Basis of Measurements

The forward operator from Definition 2.3.1 depends linearly on the bottom water temperature function and the geothermal gradient g. Using the cut-off Fourier series from (5.2), the bottom water temperature function itself depends linearly on the mean temperature A and the phase amplitudes B_i. In conclusion, we can find a linear basis of the measurement space if we consider the day of the annual minimum d, which appears non-linearly in the temperature function, to be known.

Definition 8.1.1. *Forward Operator for known d.*
Let F be the measurement operator from Definition 5.3.1. When the annual minimum $d \in [0, 365[$ is known, we split the parameter vector such that $f = (c, d)$ and define

$$F_d(c, g) := F((c, d), g). \tag{8.1}$$

Note that the measurement operator for known d is linear.

With this notation we can now generate a basis of the measurement space.

Definition 8.1.2. *Basis of Measurements.*
Let $d \in [0, 365[$ and $W[f](t)$ be the bottom water temperature function as in Equation (5.1), i.e $f \in \mathbb{R}^3$, and $g_0 > 0$ a guess for the geothermal gradient. We call

$$\begin{aligned}
&\hat{A}_d \in \mathbb{R}^{s\times 3}\\
&\hat{A}_d(\cdot, 1) = F_d((1,0),0),\\
&\hat{A}_d(\cdot, 2) = F_d((0,1),0),\\
&\hat{A}_d(\cdot, 3) = F_d((0,0),g_0).
\end{aligned}$$

a basis of measurements.

Note, that this definition can be easily adapted to the cut-off Fourier series (5.2).
Figure 20 shows the basis of measurements generated for the bottom water temperature function from Equation (5.1) for Example 1.

Theorem 8.1.3.
Let $d \in [0, 365[$ and $W[f](t)$ be the bottom water temperature function, $g_0 > 0$ a guess for the geothermal gradient and $\hat{A}_d \in \mathbb{R}^{s\times 3}$ the basis of measurements *as in Def. 8.1.2.*
For $(f, g) = p := ((p_1, p_2, d), p_3) \in \mathbb{R}^4$,

$$F(p) = \hat{A}_d\,(p_1, p_2, \tfrac{p_3}{g_0}) \tag{8.2}$$

holds.

Proof. Definition 8.1.2 of the basis measurements yields

$$\hat{A}_d\,(p_1, p_2, \tfrac{p_3}{g_0}) = p_1\,F_d((1,0),0) + p_2\,F_d((0,1),0) + \frac{p_3}{g_0}\,F_d((0,0),g_0).$$

The measurement operator F is the concatenation of the discrete forward operator $\hat{\Phi}$ and the linear interpolation $\mathcal{P}$ onto the measurement mesh $(\hat{x}, t_m)$, see Definition 5.3.1.
Recall the variational formulation (2.8), whose solution defines the continuous forward operator Φ in Definition 2.3.1 and insert the shifted cosine (5.1) for the bottom water temperature

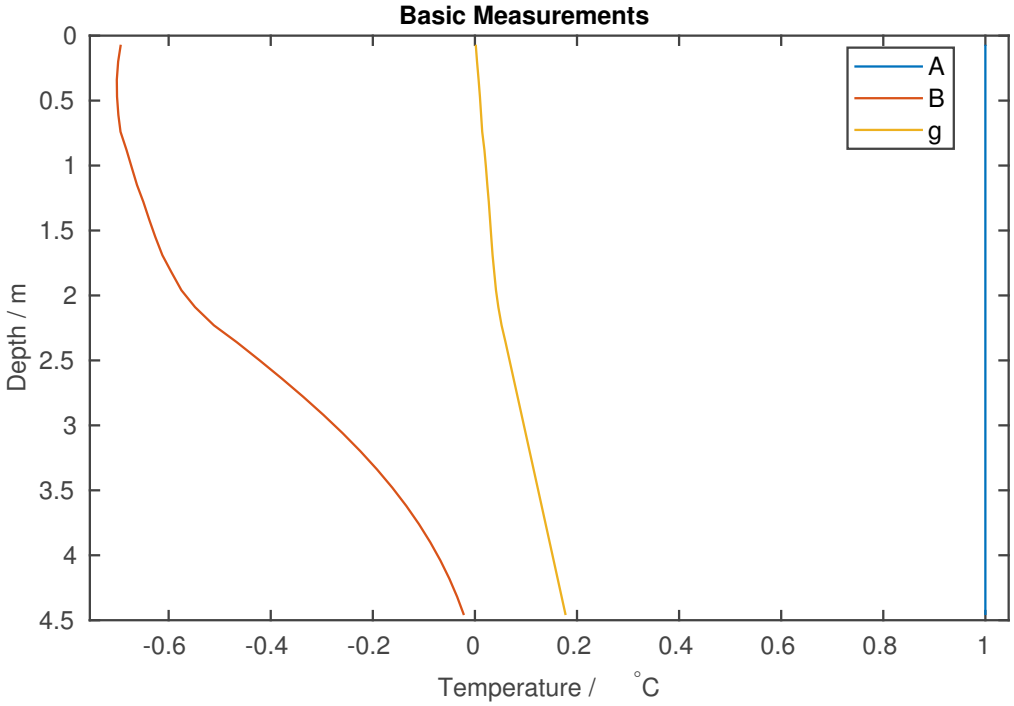

Figure 20.: Basis of measurements for Example 1 generated from the shifted cosine and the geothermal gradient. The plotted temperature-depth profiles are the columns of the basis of measurements Matrix $\hat{A}_d$.

function, to see

$$\partial_t \langle \mathbf{u}^t, \psi \rangle + a[\mathbf{u}^t, \psi] = g\psi(x_E) + \tilde{h}\ (A + B\cos(\omega t + \varphi(d)))\,\psi(0).$$

For fixed d we see easily that

$$\Phi((A, B, d), g) = A\,\Phi((1, 0, d), 0) + B\,\Phi((0, 1, d), 0) + g\,\Phi((0, 0, d), 1).$$

For the discrete forward operator $\hat{\Phi}$ from Definition 4.1.2, we see from the discrete variational formulation (4.1) that the same equality holds.
Now, as the interpolation operator $\mathcal{P}[\hat{x}, t_m]$ is linear the equality also holds for F, which leaves us with

$$\begin{aligned} F(p) &= p_1\,F_d((1,0),0) + p_2\,F_d((0,1),0) + \frac{p_3}{g_0}\,F_d((0,0),g_0) \\ &= \hat{A}_d\,(p_1, p_2, \tfrac{p_3}{g_0}). \end{aligned}$$

□

Now, the matrix equation

$$\hat{A}_d c = \Theta^\epsilon \tag{8.3}$$

can be solved in the sense of least-squares, e.g. with the Tikhonov regularization.
The entries of c correspond to parameters of the water function and the geothermal gradient relative to the first guess g_0. In consequence of Theorem 8.1.3 the solution to the inverse problem is given by $p^* = ((c_1, c_2, d), c_3\,g_0)^T$.
Consider the linear equation $Ax = y$ for a matrix $A \in \mathbb{R}^{m\times n}$, where $m >> n$, i.e. the data $y \in \mathbb{R}^m$ is of a higher dimension as the parameter $x \in \mathbb{R}^n$. This equation has not necessarily a unique solution. Instead one minimizes the discrepancy $\|Ax - y\|^2$ to obtain the least-squares solution (Engl et al., 2000). In Tikhonov regularization a penalty term $\alpha\,\|x\|^2$ is added to keep the parameter vector small.
A more thorough introduction to Tikhonov regularization can be found in Engl et al. (2000), we will here just cite the following theorem.

Theorem 8.1.4. *Tikhonov Minimization, see (Engl et al., 2000, Thm 5.1).*
Let $A \in \mathbb{R}^{m \times n}$, where $m >> n$, $y \in \mathbb{R}^m$ and $\alpha > 0$.
If $x \in \mathbb{R}^n$ solves

$$(A^*A + \alpha I_n)\, x = A^* y. \tag{8.4}$$

then it is the unique minimizer to the Tikhonov functional

$$J(x) = \|Ax - y\|^2 + \alpha \|x\|^2. \tag{8.5}$$

The implementation is shown in Algorithm 3. We used the Tikhonov regularization scheme and stopped it again with the discrepancy principle 7.0.1.

Algorithm 3 Linear Fitting with Tikhonov Regularization

Input: Θ^ϵ, ϵ, d g_0, τ, k, $\alpha = \alpha_1, \dots, \alpha_n$ with $\alpha_1 < \cdots < \alpha_n$
  $\hat{A}_d(:, j) = F([e_j, d], 0)$ for $j = 1, \dots, k$
  $\hat{A}_d(:, k+1) = F([0, \dots, d], g_0)$
  $b = 1$
  $i = 0$
  while $b > \tau\, \epsilon$ **and** $i \leq n$ **do**
    $i = i + 1$
    $c_i = (\hat{A}_d^T\, \hat{A}_d + \alpha_i^2\, I_{k+1})^{-1} \hat{A}_d^T\, \Theta^\epsilon$
    $b = \|\hat{A}_d\, c - \Theta^\epsilon\| / \|\Theta^\epsilon\|$
  end while
  return $f = (c_1, \dots, c_k)$, $g = c_{k+1}\, g_0$

Results for Noisy Artificial Data

The implementation of Algorithm 2 yielded good reconstructions of the parameters A, B, and g from artificial data with a maximum depth of 10 m and a basis generated from the shifted cosine (5.1). However, the day of the annual minimum needed to be known beforehand to generate the basis of measurements. When dealing with real measured temperature-depth-profiles, both assumptions are not easily satisfied. Most measurement devices have sensors to a maximum depth of 4 to 6 meters. As seen in Miesner et al. (2015) and in the results of the Newton-like inversion schemes in Section 7, the day of the annual minimum is hard

to reconstruct better than ±10 days. In Figures 21 and 22, the influence of these two assumptions are shown.

In Fig. 21, the parameter error is plotted for different depths of the measurement. We see that the reconstruction error $\|[c, d, c' g_0]^T - [f, g]^T\| / \|[f, g]^T\|$ decreases significantly when the lowermost sensor is deeper than 7 m, except for Example 2. Observe that also the difference between the reconstructions for different depths does not vary much between 7.5 m to 10 m (see Fig. 21).

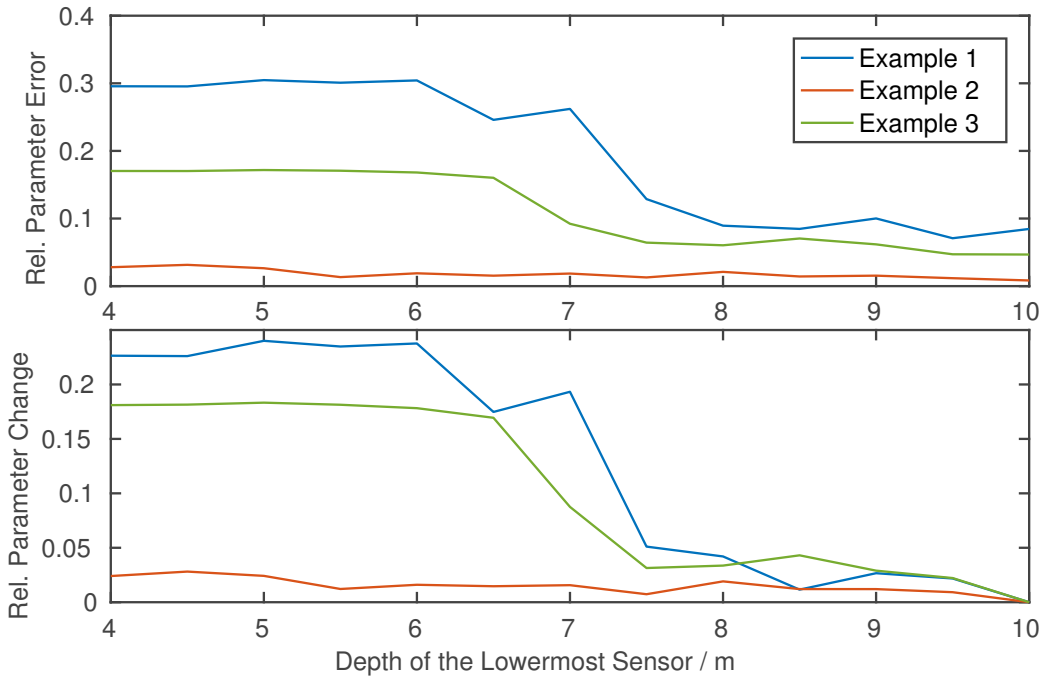

Figure 21.: Error of the reconstructed parameters against the depth of the lower most sensor. The upper panel shows the relative reconstruction error, while the lower panel shows the relative difference of the reconstructed parameters with the reconstruction for the highest depth as basis.

In Fig. 22, the reconstruction error is plotted against the guessed day of the measurement. The minimum reconstruction error does not correlate with the correct day of the annual minimum. However, for the exact day of the annual minimum the achieved reconstruction error is still relatively low. The changes for different guessed days of the measurements are highest for Example 1. In an interval of ± 10 days around the exact day of the measurement,

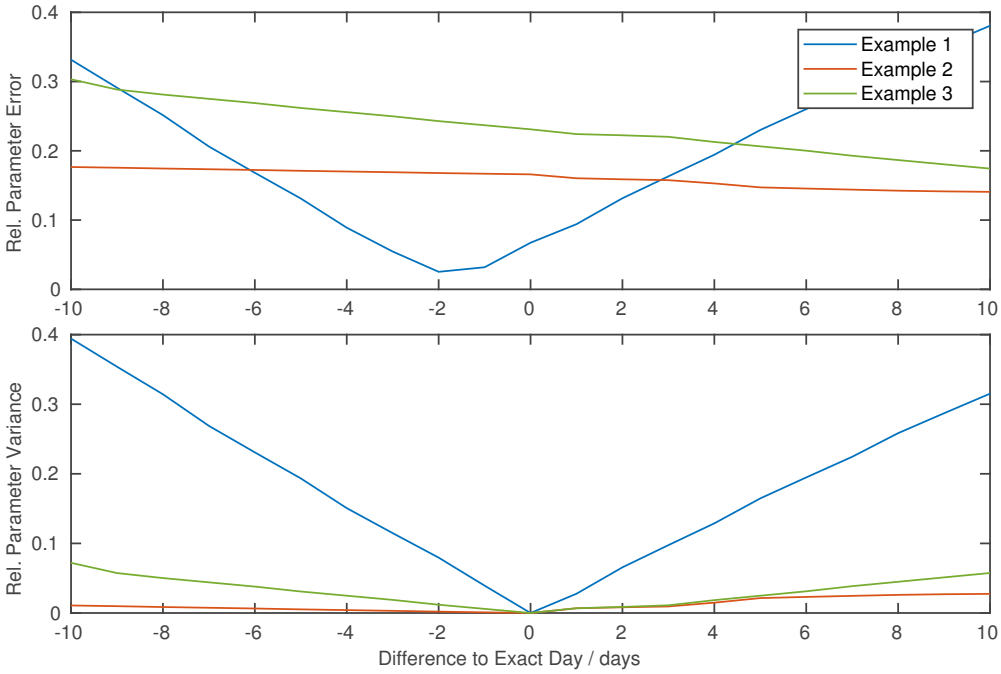

Figure 22.: Error of the reconstructed parameters against the guessed day of the measurement. The upper panel shows the relative reconstruction error, while the lower panel shows the relative variation of the reconstructed parameters with the result for the exact day of the measurement as basis.

the reconstructed parameters change about 40 % for Example 1, but only 8 % for Example 3 and less than 2 % for Example 2.

8.2. Heuristic Linear Fitting with Day Reconstruction

As the method yielded good results, especially for measurements deeper than 7.5 m, we wanted to find a way to reconstruct the day of the annual minimum d as well. Therefore, we plotted the achieved discrepancies for different guessed days of the annual minimum d for each realization of the regularization parameter α. We observed that the discrepancy does decrease around the real day of the measurements (see Fig. 23).

Determining the Day of the Measurement from the Reconstructions

Thus, we implemented a heuristic algorithm to guess the day of the annual minimum from the discrepancies. For each regularization parameter α, the day with the smallest discrepancy is calculated and the resulting mean value is used as the *probable* day of the measurement. For this *probable* day, the reconstruction result is then obtained by the discrepancy principle.

Algorithm 4 Heuristic Linear Fitting

Input: Θ^ϵ, ϵ, $\hat{d} = d_1, \dots, d_m$ g_0, τ, k, $\alpha = \alpha_1, \dots, \alpha_n$ with $\alpha_1 < \dots < \alpha_n$
  for $i = 1, \dots, m$ **and** $j = 1, \dots, n$ **do**
    $\hat{c}_{d_i}^{\alpha_j} = \min_{\hat{c}} \|\hat{A}_{d_i}\,\hat{c} - \Theta^\epsilon\|^2 + \alpha_j \|\hat{c}\|^2$
    $b_{d_i}^{\alpha_j} = \hat{A}_{d_i}\,\hat{c}_{d_i}^{\alpha_j} - \Theta^\epsilon$
  end for
  for $j = 1, \dots, m$ **do**
    $d_j^* := \min_{d_i} b_{d_i}^{\alpha_j}$
  end for
  $d^* := 1/n \sum_j d_j^*$
  for $i = 1, \dots, n$ **do**
    if $b_{d^*}^{\alpha_i} \leq \tau\,\epsilon$ **then**
      return $f = (c_{d^*1}^{\alpha_i}, \dots, c_{d^*k}^{\alpha_i})$, $g = c_{d^*(k+1)}^{\alpha_i}\, g_0$
    end if
  end for

Fig. 23 is based on Example 3 and the noise level on the artificial data is 0.5 %. Here, the *probable* day differs only three days from the exact day and the final parameter error was 6.5 %. For Example 2 the *probable* day matched the exact day, and yielded a parameter

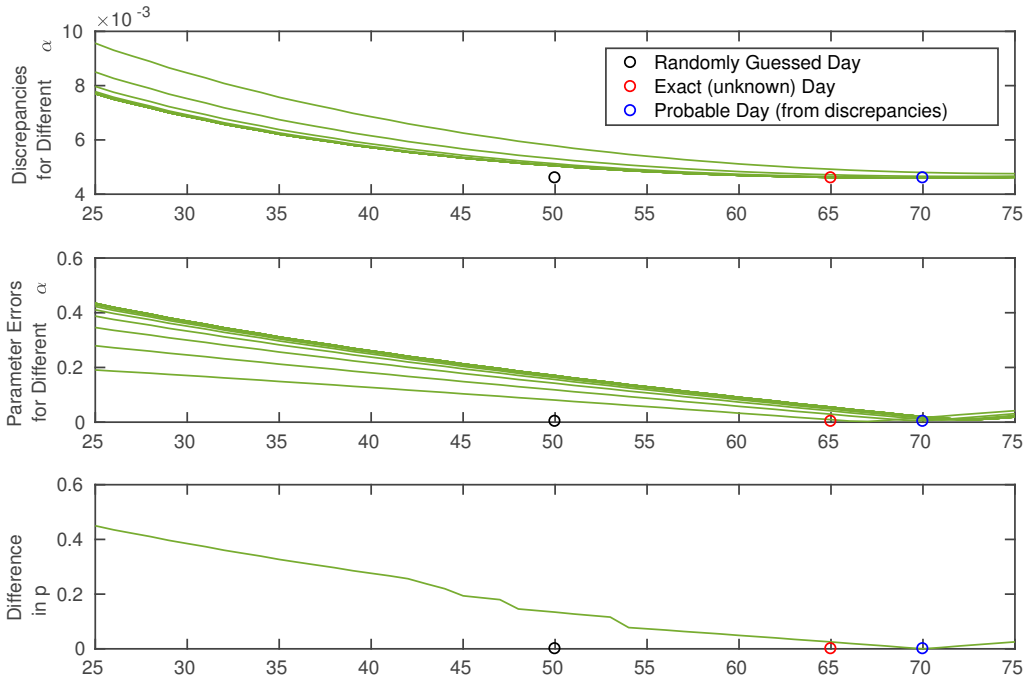

Figure 23.: Discrepancies and parameter errors for varying regularization parameter α. In the upper panel, the discrepancies for varying regularization parameter α are shown for a guessed day (black) $\pm$ 25 days. In the second panel the parameter error is shown in the same way. The exact day is indicated in red and the *probable* day in blue. In the lower panel, the difference in the reconstructed parameters p^* is plotted in relation to the results for the *probable* day (blue).

error of less than 2 %, while the *probable* day for Example 1 lay within five days and yielded a parameter error of 7.7 %.
The results are now of course depending on the choice of the regularization parameters α_i. To avoid influencing the accuracy of the results by choosing wisely, we used a randomization to choose 20 to 30 values in the range of 1×10^{-8} to 1×10^{1}

Results for Noisy Artificial Data

As with the algorithms before, the full results for the heuristic linear fitting are shown in Tables 16 – 19 in the Appendix. Additionally, in Fig. 24 the evolution of the relative error for Example 1 in dependence on the white noise on the data is shown.
The overall reconstructions errors is roughly 10 % for Example 2 with a decreases to about 5 % for noise under 0.1 %. For Example 3 we see decreasing reconstruction errors to about 1 % for noise levels under 0.2 % and for Example 1 the reconstruction error is lower than 0.5 % for noise levels under 1 %, see bottom panel in Fig. 24.
The reconstruction errors of the single parameters decrease for decreasing noise levels with a *saturation* for noise under 1 %. Remarkable is that the reconstruction error of the amplitude B is one order of magnitude higher for Example 2, than for the other two, see second panel in Fig. 24.
For the geothermal gradient g we can observe, that it is best reconstructed for Example 3 and worst for Example 1. This is probably due to the depth into which the annual signal propagates, see also Section 3.3.

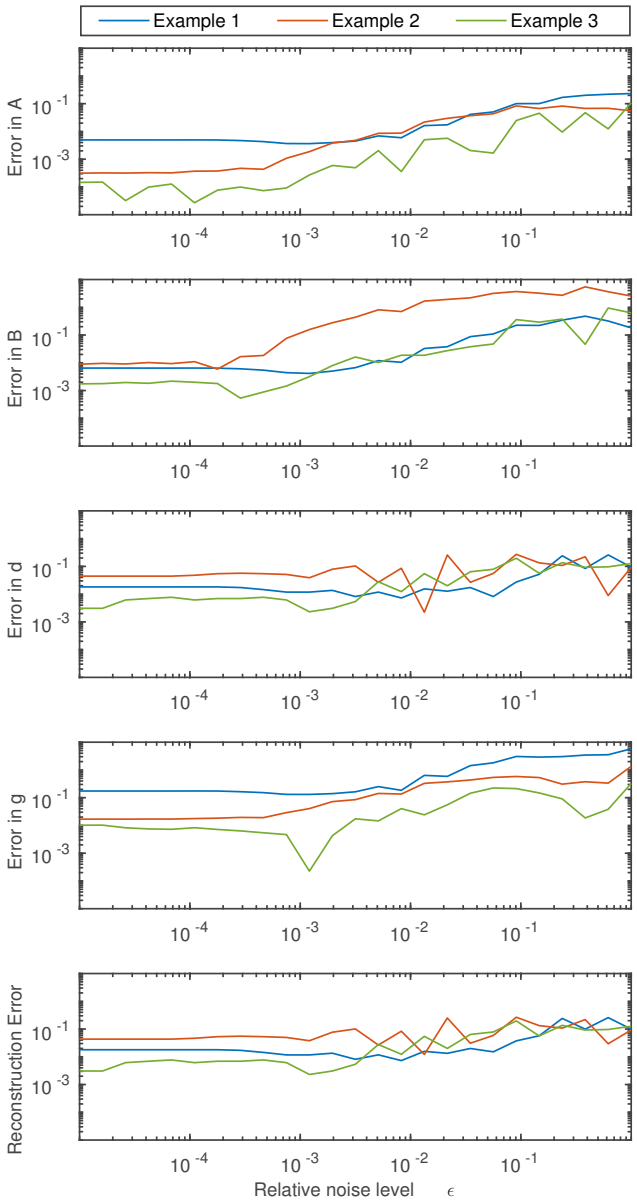

Figure 24.: Evolution of the relative reconstruction error in dependence on the white noise on the data. The error are shown for all three example data sets. The top four panels show the relative reconstruction error in the single parameters, while the bottom panel shows the overall reconstruction error.

9. Markov Chain Monte Carlo Methods

In this section, we introduce a purely statistical approach to find the *most probable* parameter set to explain the given data. In Section 7.2 we already defined the probability densities for data y^* and parameters x^* in a linear setting $Ax = y$. The *most probable* parameter set to cause the data can than be approximated from the posterior density.

In Section 7.2 we used the statistical approach to tackle the linearized Newton step. There, the posterior mean can be calculated explicitly. However, the model equation $F(p) = \Theta^\epsilon$ is nonlinear and in the nonlinear setting, we need to approximate the integral in the conditional mean with other techniques.

We again denote with capital letters the random variable and with lowercase letters their realization. We assume that the parameter and noise are normal distributed around the (unknown) expected value with (unknown) variance, i.e. $P \sim \mathcal{N}(p_e, \Gamma_{pr})$ and $E \sim \mathcal{N}(0, \Gamma_{noise})$ respectively. Further, the noise on the data shall be additive, i.e. $\Theta^\epsilon = \Theta + E$, and the error and data are mutually independent.

The inverse problem is now to find the *most probable* parameter p, such that $F(p) = \Theta_{obs}$ holds for given measured data.

Let us first recall Bayes' theorem of inverse problems as of Theorem 7.2.1:

$$\pi_{post} = \pi(p|\theta_{obs}) = \frac{\pi_{pr}(p)\pi(\theta_{obs}|p)}{\pi(\theta_{obs})}.$$

Further, for mutually independent data Θ and noise E, we can write the posterior probability as $\pi(\theta_{obs}|p) = \pi_{noise}(\theta_{obs} - F(p))$.

As already stated in Section 7.2, the goal is to determine the conditional mean

$$p_{CM} = \int_{\mathbb{R}^k} p\, \pi_{post}(p)\, dp. \tag{9.1}$$

While this integral can be solved explicitly for linear operators, such as in the Newton step in Section 7.2, other methods need to be obtained for nonlinear operators.

9.1. Monte Carlo Integration and Markov Chains

To calculate the integral (9.1), we will use Markov Chain Monte Carlo (MCMC) methods. We will follow the notation of Kaipio and Somersalo (2004).
Therefore, let μ be a probability measure over $\mathbb{R}^k$ and f a vector-valued measurable function, which is integrable over $\mathbb{R}^k$. In classic integration with quadrature methods, we define support points $x_j \in \mathbb{R}^k$ and weights w_j and then calculate the integral as the weighted sum

$$\int_{\mathbb{R}^k} f(x)\mu(dx) \approx \sum_{j=1}^{K} w_j f(x_j).$$

Now, in *Monte Carlo integration* the support points x_j are picked by their probability $\mu(x_j)$, such that the integral can be approximated by the *ergodic average*

$$\int_{\mathbb{R}^k} f(x)\mu(dx) \approx \frac{1}{K} \sum_{j=1}^{K} f(x_j).$$

This is done with Markov chains.

Definition 9.1.1. *(see Smith (2014))*
For a random variable X, a sequence of realizations $\{x_i, i \in \mathbb{N}\}$, with

$$\mu(x_{n+1}|x_0, \dots, x_n) = \mu(x_{n+1}|x_n)$$

is called a Markov chain.

The Markov property states that the next chain member depends only on the last chain

member and not on the whole chain.

Now, let x_k be the current element of the Markov chain. A candidate y for x_{k+1} is drawn from a *proposal distribution* $q(x_k, y)$. For now this is just any probability distribution and we will later define how it is chosen in the algorithm. The candidate y is than accepted with the probability

$$\alpha(x_k, y) = \min\left\{1, \frac{\mu(y)q(y, x_k)}{\mu(x_k)q(x_k, y)}\right\}. \tag{9.2}$$

If the probability of the candidate y is higher than the probability of the current element x_k, the candidate is accepted with probability 1. Otherwise, the probability of acceptance is not zero.

The following lemma provides criteria for the proposal distribution q such that the approximation of the integral via the ergodic average converges. Recall that a Markov chain $\{x_j\}$ is called *irreducible* if it visits every set $A \subset \mathbb{R}^k$ of positive measure $\mu(A) > 0$ with a probability greater than zero. It is called *periodic* if there exists a periodic loop in which a certain number of sets are visited over and over again.

Lemma 9.1.2. *(see (Kaipio and Somersalo, 2004, Proposition 3.11))*
Let μ be a probability measure in $\mathbb{R}^k$, $\{x_j\}$ a Markov chain with the proposal distribution q and $f \in L^1(\mu(dx))$. If $\{x_j\}$ is irreducible and aperiodic, then for all $x \in \mathbb{R}^k$ it holds that

$$\lim_{K\to\infty} \frac{1}{K}\sum_{j=1}^{K} f(x_j) = \int_{\mathbb{R}^k} f(x_j)\mu dx.$$

9.2. The Metropolis-Hastings Algorithm

Applying this theory to the inverse problem, we need to approximate the conditional mean (9.1) $p_{CM} = \int_{\mathbb{R}^k} p\pi_{post}(p)\, dp$. Therefore, $f(p) = p$ and $\mu(dp) = \pi_{post}dp$.

In the Metropolis-Hastings (MH) construction of q, we assume that the next element of the chain behaves like white noise around the current element, i.e. we construct the candidate $p_{can} \sim \mathcal{N}(p_k, V)$ for a chosen covariance matrix V. This leads to a proposal distribution q that is irreducible and aperiodic. Thus, Lemma 9.1.2 is applicable and the integral equation for the conditional mean converges. Furthermore, it is also symmetric, so the acceptance

probability α, see (9.2), becomes

$$\alpha(p_k, p_{can}) = \min\left\{1, \frac{\pi_{post}(p_{can}|\theta)}{\pi_{post}(p_k|\theta)}\right\}.$$

With the use of Bayes' theorem 7.2.1 and the representation for $\pi(\theta_{obs}|p)$, we end with

$$\alpha(p_k, p_{can}) = \min\left\{1, \frac{\pi_{pr}(p_{can})\pi_{noise}(\theta - F(p_{can}))}{\pi_{pr}(p_k)\pi_{noise}(\theta - F(p_k))}\right\}.$$

With the following theorem (which we cite without proof), we have a tool to compute p_{can} with the Cholesky decomposition of the covariance matrix V.

Theorem 9.2.1. *(from (Smith, 2014, Thm. 4.23))*
Let $Y = [Y_1, \dots, Y_n]^T$ be a normally distributed random vector, $Y \sim \mathcal{N}(\mu, V)$, where V is positive definite and $\mu \in \mathbb{R}^n$. Let $Z \sim \mathcal{N}(0, I_n)$, where I_n is the $n \times n$ identity. Then $Y = Z + \mu$, where $V = RR^T$ and R is a lower triangular matrix.

Now, we have every step to formulate the Metropolis-Hastings Algorithm.

Algorithm 5 The Metropolis-Hastings Algorithm, see also Smith (2014, Alg. 8.5)

Input: Θ^ϵ, ϵ, K and covariance matrix V for the proposal distribution
  $p_1 = p_e$
  calculate R as the Cholesky decomposition of V
  $R = 1/\|R\|R$
  for $j = 2, \dots, K$ **do**
    draw $z_j \sim \mathcal{N}(0, I_s)$
    $p_{can} = p_{j-1} + R\, z_j$
    draw $t \sim \mathcal{U}(0, 1)$
    $\alpha = \min\left\{1, \frac{\pi_{noise}(\Theta^\epsilon - F(p_{can}))\, \pi_{pr}(p_{can})}{\pi_{noise}(\Theta^\epsilon - F(p_{j-1}))\, \pi_{pr}(p_{j-1})}\right\}$
    if $t < \alpha$ **then**
      $p_j = p_{can}$
    else
      $p_j = p_{j-1}$
    end if
  end for
  return $p = \overline{p} = 1/K \sum_{j=1}^{K} p_j$

9.3. Delayed Rejection and Adaptive Metropolis

The proposal distribution q used in the Metropolis-Hastings algorithm 5 only produces a *good* Markov Chain if the covariance Matrix V is well chosen. A too narrow covariance will result in a low acceptance rate and thus (in the worst case) in a stationary Markov Chain. A too broad covariance on the other hand will not stay in areas of high probability long enough to explore it thoroughly.

There are some strategies to cope with these two mechanisms. We will introduce the *delayed rejection adaptive Metropolis* (DRAM) algorithm, introduced by Haario et al. (2006).

The *delayed rejection* algorithm from Mira (2001) proposes a second stage candidate p_{can}^2 if the first candidate was rejected. The second stage proposal distribution q_2 now may also depend on the first proposed candidate p_{can}. The second stage candidate is then accepted with a ratio α_2. As long as the proposal distributions q_i are irreducible, aperiodic and symmetric, just as the initial proposal distribution q, the obtained chain is still markovian and thus the ergodic property holds (Mira, 2001). This ensures that the chain does not stay stationary, even for a narrow covariance matrix.

The *adaptive Metropolis* from Haario et al. (2001) on the other hand, adapts a given covariance matrix V_0 of the proposal distribution to the earlier chain members

$$V_i = \begin{cases} V_0, & i \leq k_0 \\ s_k \operatorname{Cov}(p_0, \dots, p_{i-1}) + s_k \epsilon I_k, & i > k_0. \end{cases} \tag{9.3}$$

Here, s_k is a parameter that depends only on the dimension of the parameter space and $\epsilon > 0$ is a small constant to ensure that the covariance matrix does not become singular.

The obtained chain is then of course no longer Markovian, but the ergodic properties can be proven nonetheless (Haario et al., 2006). In practice it was found that the adaption should only be done every k_0 steps and that $s_k = 1.4^2/d$ is an optimal scaling parameter (see (Haario et al., 2006)).

For the implementation it is advantageous that the covariance as well as the chain mean $\overline{p}$ follow a recursion formula (Haario et al., 2006).

Putting all this together, the DRAM algorithm reads as follows:

Algorithm 6 The DRAM Algorithm, see also Haario et al. (2006)

Input: Θ^{ϵ}, ϵ, K, and design parameters ε, k_0, s_k

$j = 1$

$p_j = p_e$

$\overline{p} = p_e$

$\sigma^2 = \|F(p_j) - \Theta^{\epsilon}\|^2/(s-k)$

$V = \sigma^2 \left(F'(p_j)^T F'(p_j)\right)^{-1}$

calculate R as the Cholesky decomposition of V

$R = 1/\|R\| R$

for $j = 2, \ldots, K$ **do**

  draw $z_j \sim \mathcal{N}(0, I_s)$

  $p_{can} = p_{j-1} + R\, z_j$

  draw $t \sim \mathcal{U}(0,1)$

  $\alpha_1 = \min\left\{1, \frac{\pi_{noise}(\Theta^{\epsilon} - F(p_{can}))\, \pi_{pr}(p_{can})}{\pi_{noise}(\Theta^{\epsilon} - F(p_{j-1}))\, \pi_{pr}(p_{j-1})}\right\}$

  if $t < \alpha_1$ **then**

    $p_j = p_{can}$

  else

    delayed rejection:

    draw p'_{can} from proposal distribution $q_2(p_j, p_can)$

    accept $p_j = p'_{can}$ with second stage probability

    $\alpha_2 = \min\left\{1, \frac{\pi_{post}(p'_{can})\, q_1(p'_{can}, p_{can})\, q_2(p'_{can}, p_{can}, p_{j-1})\, [1-\alpha_1(p'_{can})]}{\pi_{post}(p_{j-1})\, q_1(p_{j-1}, p_{can})\, q_2(p_{j-1}, p_{can}, p'_{can})\, [1-\alpha_1(p_{j-1})]}\right\}$

  end if

  update $\overline{p} = j/(j+1)\overline{p} + 1/(j+1)\, p_j$

  update $V = ((j-1)/j)\, V + s_k/j\, (j\overline{p}\,\overline{p}^T - (j+1)\overline{p}\,\overline{p}^T + p_j\, p_j^T + \varepsilon I_s)$

  update R every k_0 steps

end for

return $p = \overline{p}$

Results for Artificial Data

Again, the results are shown in the Appendix in Tables 21 – 24. The algorithms were executed for a total number of 70000 ensemble members. The standard deviation is calculated from the generated ensemble rather than from multiple executions.

The MH algorithm yields good reconstruction results for added noise of 0.5 %, with the exception of the amplitude B for all examples and the geothermal gradient g for Example 3. For artificial data with 1 % noise the reconstruction results worse, especially for the geothermal gradient and the whole parameter set of Example 3.

For the DRAM algorithm, we can observe good results with relative reconstruction errors of under 8 % for Examples 1 and 2 and still quite good results with reconstruction errors around 25 % for Example 3. The reconstruction results do not worsen much for higher percentage of added noise.

However, both algorithm have proven convergence, which leads to the conclusion that the error bounds could be lowered, i.e. with a higher number of ensemble members.

10. Inversion Results

10.1. Results for Artificial Data

For a comparison of the used methods, the inversion results of all methods for realistic artificial data are shown in the following tables. Additionally for each noise level the resulting temperature-depth profiles for one example are presented (Figs. 25 and 26). We use data sets generated over 22 sensors at a maximum depth of 5.95 m with 0.5 % (Table 3) and 1 % (Table 4) added white noise to realistically approximate real data. The error bounds of the real data sets (see also Section 3.2) are in the range of 0.3 to 2 %, and the measurement device has 22 sensors (see Section 3.1), so these assumptions seem quite fitting. However, in all examples some sensors could not be used due to less penetration depth or technical issues, and also Example 1 was retrieved with a different device with a maximum penetration depth of 3 m. In consequence, *realistic artificial data* are just that: artificial data which tries to resemble real data sets.

We additionally implemented and tested the REGINN algorithm (as discussed by Miesner et al. (2015)) and a two-step Newton algorithm, where the Newton step is alternately applied to the bottom water temperature parameters f only with fixed geothermal gradient g and vice versa. Both algorithms showed similar characteristics as the other Newton-like algorithms with slower convergence. They added no further insights into the inverse problem and are thus not part of this thesis.

The tables show the mean results of 100 executions of all methods, respectively 70000 chain members for the MCMC methods, the discrepancy of said mean result, and the mean calculation time. Some of the methods did not converge for all realistic data sets. It is possible that (method specific) parameter settings exist for which those methods do converge, that we did not find. We did not put too much effort into finding those parameter

settings for two reasons: First, finding the perfect settings for 9 methods exceeds the scope of this work, and second it does not help with real data sets, as fine tuning mostly only works for a small scope of data sets. The methods that did not converge are therefore marked with dashes in the tables and are not included in the figures. However, the discrepancy and calculation time are given in the tables. The mean result and standard deviation of the tables are shown in Figs. 25 and 26 in the upper right panel.

All calculations were done with MATLAB R2017a on a Debian 8 machine with 32GB RAM and an Intel® Core© i7-3770 processor with 3.40GHz and 8 cores.

With the exception of the Metropolis-Hastings Algorithm, all methods that did converge yielded good results. The achieved accuracy resembles what we already discussed for general artificial data for the different methods. The mean temperature A is the best reconstructed value in all methods and example data sets. The accuracy of the reconstructed amplitude B is also very good, with the exception of Example 2. We already discussed that this is due to the relative high difference between the values of A and B (see Section 7.1, especially Fig. 19). The day of the annual minimum d is reconstructed mostly within 3 weeks, which is not ideal but as discussed by Miesner (2013) not surprising due to the small derivative of the cosine around the extrema. Lastly, the geothermal gradient was also reconstructed very well for Examples 2 and 3 and not very well for 1. Again, this is due to the combination of high thermal diffusivity values and amplitude and the resulting thermal length (see Section 3.3).

The example results in Figs. 25 and 26 show resulting bottom water temperature function in the upper left panel and the resulting temperature-depth profiles at the given day of measurement. The temperature-depth profiles are given with a higher depth than the given measurements, to show how the temperature field can vary for varying parameter sets while the discrepancies are relatively close together.

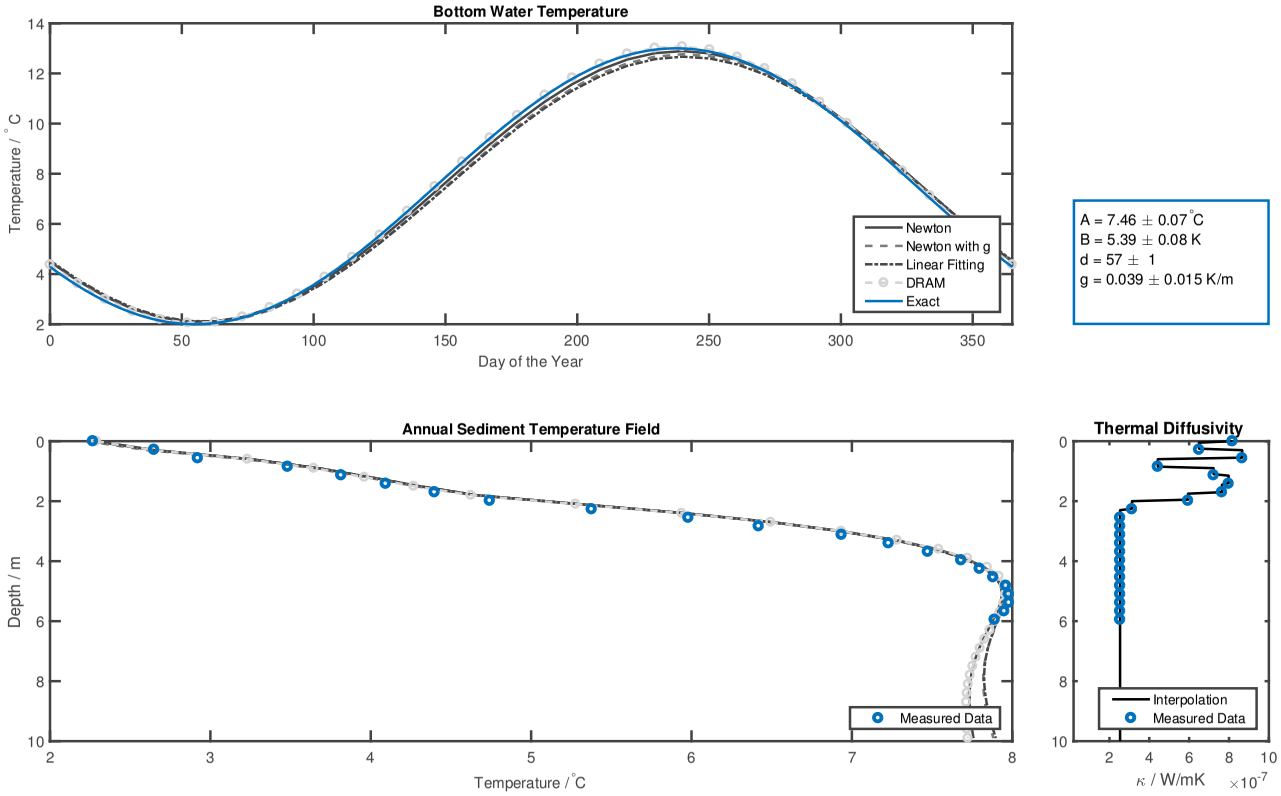

Figure 25.: Mean results of the different methods for artificial data with thermal properties from Example area 1 and 0.5 % added white noise.
The measured temperatures and thermal diffusivity values are plotted as circles in the lower left and lower right panel, respectively. In the upper left panel the bottom water temperature functions are shown for the reconstructed parameters for each method and in the upper right panel the mean results and standard deviation of all methods are shown. In the lower left panel the calculated temperature-depth-profiles for the reconstructed parameters for each method are shown alongside the measured data.

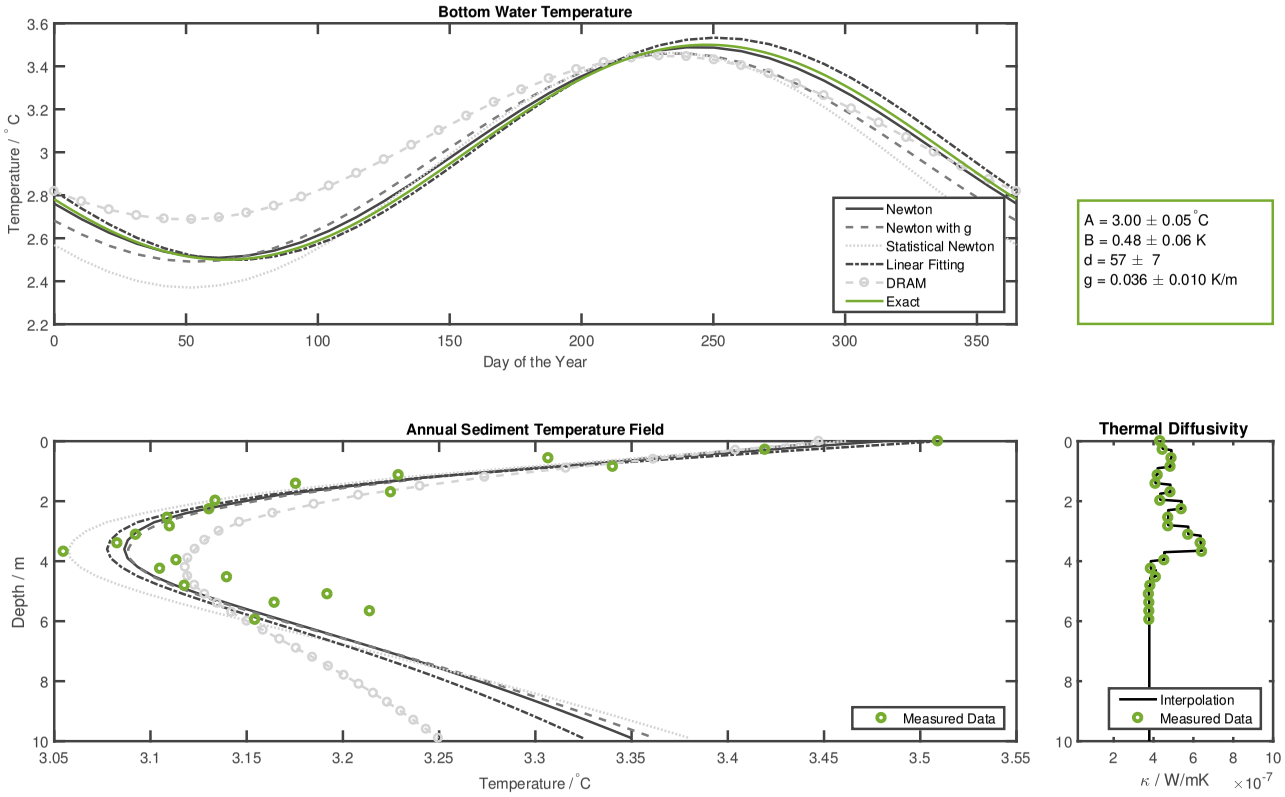

Figure 26.: Mean results of the different methods for artificial data with thermal properties from Example area 3 and 1 % added white noise.
The measured temperatures and thermal diffusivity values are plotted as circles in the lower left and lower right panel, respectively. In the upper left panel the bottom water temperature functions are shown for the reconstructed parameters for each method and in the upper right panel the mean results and standard deviation of all methods are shown. In the lower left panel the calculated temperature-depth-profiles for the reconstructed parameters for each method are shown alongside the measured data.

Table 3.: Inversion results for artificial data generated over 22 sensors at a depth of up to 5.95 m with 0.5 % white noise. For each column, the *best* result is marked in blue. Be aware that the simple Newton algorithm does not reconstruct the geothermal gradient, but uses the given value. It is therefore only shown for comparison (in light blue).
Results, for which the discrepancy is greater than 5.4 times the initial noise level are not included.

	A	B	d	g	disc	t
Example 1	°C	K	days	K/m		h
Exact Parameter	7.50	5.50	55	0.030		
Newton	7.49	5.40	57	0.030	0.026	20 h
Newton with g	7.41	5.36	58	0.052	0.024	11 d
Statistical Newton	–	–	–	–	2.446	4 d
Linear Fitting	7.38	5.29	58	0.056	0.025	10 h
Metropolis Hastings	–	–	–	–	0.053	3 d
DRAM	7.56	5.51	55	0.019	0.023	9 d
Example 2	A	B	d	g	disc	t
Exact Parameter	10.50	0.30	45	0.084		
Newton	10.50	0.29	48	0.084	0.005	20 h
Newton with g	10.59	0.48	52	0.066	0.006	11 d
Statistical Newton	–	–	–	–	0.037	4 d
Linear Fitting	10.46	0.22	48	0.090	0.006	10 h
Metropolis Hastings	–	–	–	–	0.029	3 d
DRAM	10.56	0.32	49	0.058	0.006	9 d
Example 3	A	B	d	g	disc	t
Exact Parameter	3.00	0.50	65	0.038		
Newton	3.00	0.49	65	0.038	0.005	20 h
Newton with g	3.01	0.48	64	0.033	0.006	11 d
Statistical Newton	2.88	0.56	51	0.052	0.015	4 d
Linear Fitting	3.01	0.50	68	0.034	0.006	10 h
Metropolis Hastings	–	–	–	–	0.473	3 d
DRAM	3.03	0.44	50	0.021	0.007	9 d

Table 4.: Inversion results for artificial data generated over 22 sensors at a depth of up to 5.95 m with 1 % white noise. For each column, the *best* result is marked in blue. Be aware that the simple Newton algorithm does not reconstruct the geothermal gradient, but uses the given value. It is therefore only shown for comparison (in light blue).
Results, for which the discrepancy is greater than 5.4 times the initial noise level are not included.

	A	B	d	g	disc	t
Example 1	°C	K	days	K/m		h
Exact Parameter	7.50	5.50	55	0.030		
Newton	7.49	5.39	57	0.030	0.027	10 h
Newton with g	7.38	5.29	58	0.053	0.020	11 d
Statistical Newton	–	–	–	–	0.108	5 d
Linear Fitting	7.37	5.28	58	0.056	0.027	10 h
Metropolis Hastings	7.53	5.52	52	0.036	0.017	3 d
DRAM	7.56	5.52	54	0.023	0.026	9 d
Example 2	A	B	d	g	disc	t
Exact Parameter	10.50	0.30	45	0.084		
Newton	10.50	0.32	59	0.084	0.011	10 h
Newton with g	10.57	0.49	65	0.067	0.012	11 d
Statistical Newton	–	–	–	–	0.029	5 d
Linear Fitting	10.38	0.06	56	0.104	0.011	10 h
Metropolis Hastings	10.72	0.92	52	0.060	0.011	75.94
DRAM	10.50	0.25	48	0.061	0.010	9 d
Example 3	A	B	d	g	disc	t
Exact Parameter	3.00	0.50	65	0.038		
Newton	3.00	0.49	62	0.038	0.010	10 h
Newton with g	2.98	0.49	53	0.041	0.011	11 d
Statistical Newton	2.92	0.55	51	0.050	0.012	5 d
Linear Fitting	3.02	0.52	69	0.033	0.010	10 h
Metropolis Hastings	–	–	–	–	0.248	3 d
DRAM	3.07	0.38	50	0.019	0.014	9 d

10.2. Results for the Basic Regions

In this section we show the inversion results for the three measured data sets introduced in Section 3.2. In Table 5 the mean inversion results for 20 repeated executions, respectively 70000 ensemble members for the MCMC methods, the achieved discrepancy, and the mean calculation time are shown. In Figures 27 – 29 the resulting bottom water temperature functions and temperature-depth profiles are shown in relation to the measured data. Additionally the mean and standard deviation of the results in Table 5 are shown in the upper right panel.

For the Newton-like methods the discrepancy principle as in Definition 7.0.1 with $\tau = 10.1$ was used to stop the iteration. Thus, in order to compare all methods only results for which $\|F(p_{\text{mean}}) - \Theta^\epsilon\|/\|\Theta^\epsilon\| \leq 5.1\,\epsilon$ are included in the tables and figures.

For the temperature-depth profiles a plot-depth of 10 m was chosen to show the influence of the geothermal gradient, especially for Example 1. While the reconstruction results of all methods fit the data (which has a maximum depth of 2.5 m) quite well, the temperature differences are increasing below 4 m depth. The negative geothermal gradient g, as reconstructed by the heuristic linear fitting (Algorithm 4) and the statistical Newton (Algorithm 2), are rather unlikely from a physical point of view. The standard deviation of the reconstructed parameters are all too big to be satisfying but still okay.

For Example 2 we can observe the aforementioned troubles that result from the relatively large differences between the values for the mean temperature and the amplitude. The curve in the upper 0.5 m in the data is not mapped in all the reconstruction results. Some fit the data at a depth of 0.5 m to 2.5 m and some in 2.5 m to 6 m. The sign of the amplitude is negative in the linear fitting reconstruction, which results in a phase-shift of nearly 180 days. Still, all shown reconstructions have a discrepancy below $5.1\,\epsilon$, and also the mean temperature A, the absolute value of the amplitude $|B|$, and the geothermal gradient g are reconstructed with a small standard deviation throughout the different methods.

Lastly, for Example 3 we have very small standard deviations for the mean temperature A, the amplitude B, and the geothermal gradient and a standard deviation of ± 13 days for the day of the annual minimum d, which is an overall very good and reliable result. Again, we can observe how the temperature-depth profiles differ more below a depth of 6 m. Note that the differences between the resulting temperature-depth profiles seem way higher for

Example 2 and 3 than for Example 1 as the scales on the x-axis are not the same.

Table 5.: Inversion results for the real data sets. As the exact parameter are not known, only for the discrepancy and the computation time the *best* result is marked in blue. Be aware that the simple Newton algorithm does not reconstruct the geothermal gradient, but uses the given value. It is therefore only shown for comparison (in light blue).
Only results for which the discrepancy is smaller than 10.1 times the assumed noise level of the data are included. A discrepancy of -1 denotes, that the method did not converge.

	A	B	d	g	disc	t
Data Set 1	°C	K	days	K/m		h
Newton	10.36	5.65	69	0.030	0.027	10 h
Newton with g	–	–	–	–	-1.000	–
Statistical Newton	8.71	9.90	124	-0.312	0.106	3 d
Linear Fitting	12.54	8.31	61	-0.910	0.035	10 h
Metropolis Hastings	–	–	–	–	0.454	3 d
DRAM	9.72	4.96	53	0.026	0.081	9 d
Data Set 2	A	B	d	g	disc	t
Newton	13.68	0.09	44	0.084	0.002	10 h
Newton with g	13.93	0.29	68	0.044	0.006	4 d
Statistical Newton	–	–	–	–	0.240	3 d
Linear Fitting	13.68	-0.06	48	0.072	0.002	10 h
Metropolis Hastings	–	–	–	–	0.338	3 d
DRAM	–	–	–	–	0.043	9 d
Data Set 3	A	B	d	g	disc	t
Newton	3.20	0.12	63	0.038	0.004	10 h
Newton with g	–	–	–	–	0.055	4 d
Statistical Newton	–	–	–	–	0.085	3 d
Linear Fitting	3.22	0.14	101	0.029	0.003	10 h
Metropolis Hastings	–	–	–	–	0.202	3 d
DRAM	3.25	0.08	50	0.020	0.007	9 d

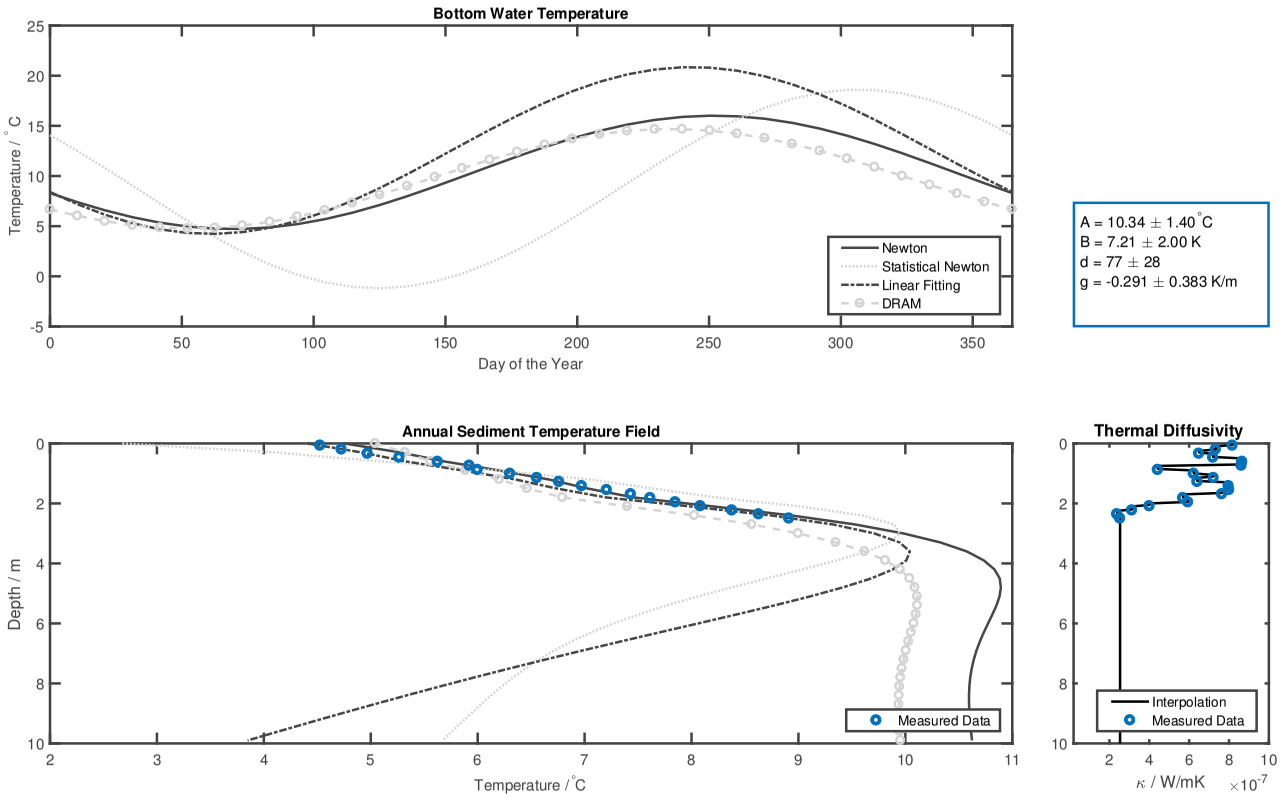

Figure 27.: Reconstruction results for the data set from Example 1.
The measured temperatures and thermal diffusivity values are plotted as circles in the lower left and lower right panel, respectively. In the upper left panel the bottom water temperature functions are shown for the reconstructed parameters for each method and in the upper right panel the mean results and standard deviation of all methods are shown. In the lower left panel the calculated temperature-depth-profiles for the reconstructed parameter for each method are shown alongside the measured data.

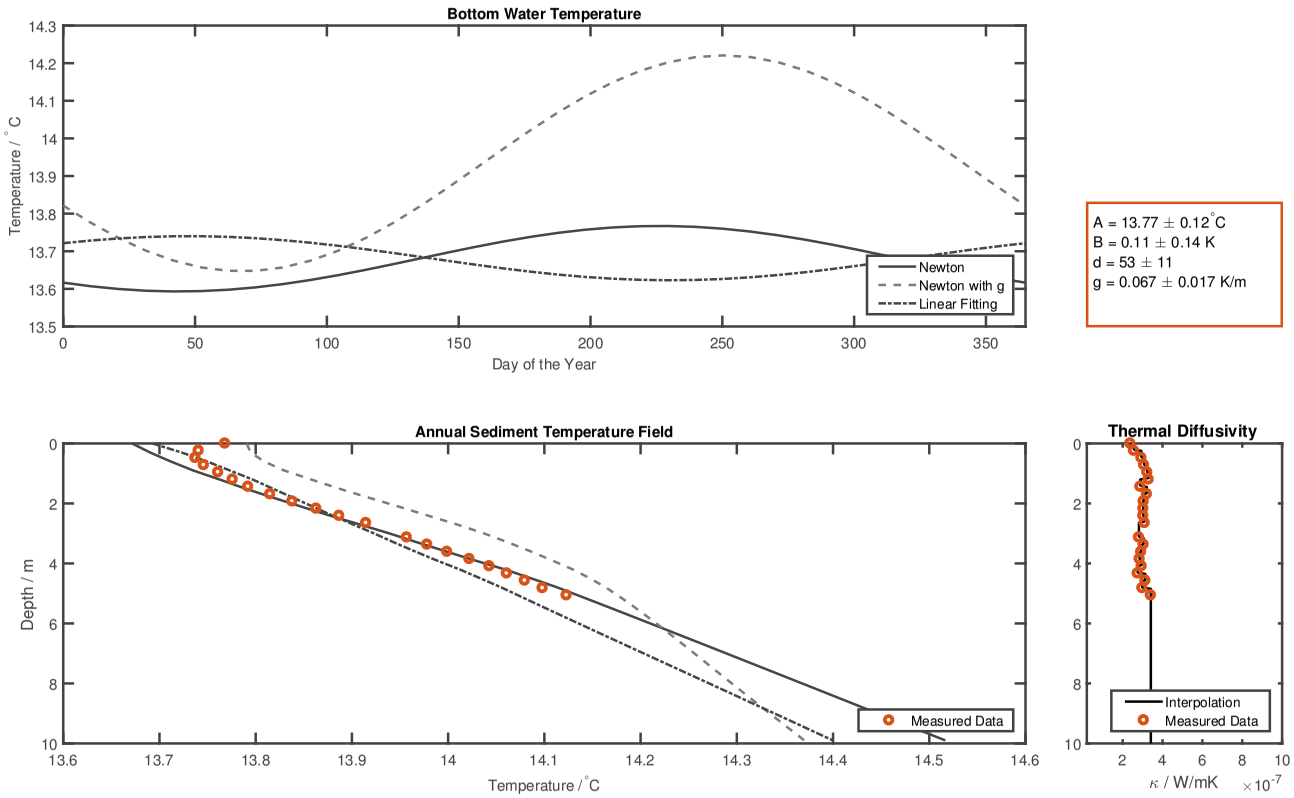

Figure 28.: Reconstruction results for the data set from Example 2.
The measured temperatures and thermal diffusivity values are plotted as circles in the lower left and lower right panel, respectively. In the upper left panel the bottom water temperature functions are shown for the reconstructed parameters for each method and in the upper right panel the mean results and standard deviation of all methods are shown. In the lower left panel the calculated temperature-depth-profiles for the reconstructed parameter for each method are shown alongside the measured data.

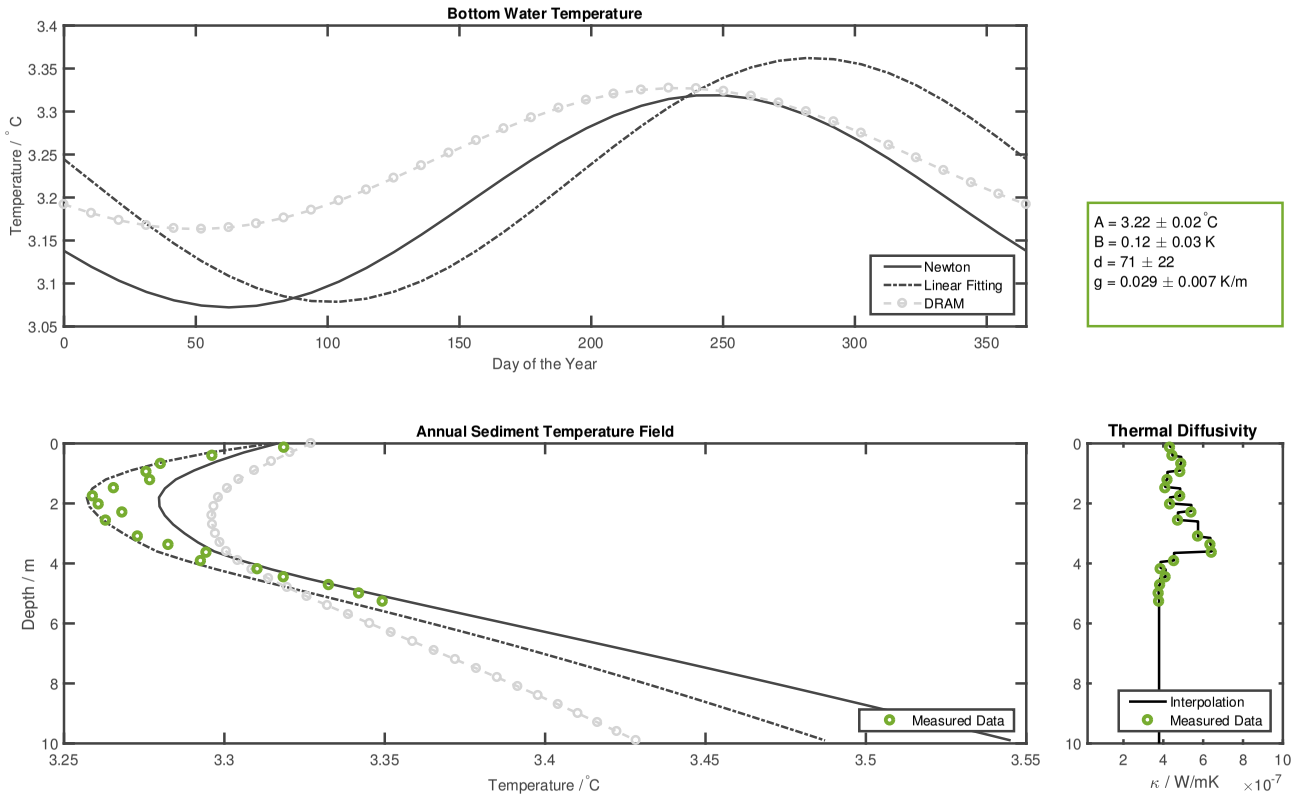

Figure 29.: Reconstruction results for the data set from Example 3.
The measured temperatures and thermal diffusivity values are plotted as circles in the lower left and lower right panel, respectively. In the upper left panel the bottom water temperature functions are shown for the reconstructed parameters for each method and in the upper right panel the mean results and standard deviation of all methods are shown. In the lower left panel the calculated temperature-depth-profiles for the reconstructed parameter for each method are shown alongside the measured data.

Part III.

Discussion

11. Generalization of the Results and Applicability of the Methods

In this work, we introduced a forward model to calculate temperature-depth profiles at a specific day from measurements of the thermal diffusivity, parameters of a bottom water temperature function, and the geothermal gradient (Chapter 2). We then introduced six inversion methods to reconstruct the parameters of the bottom water temperature function (5.1) and the geothermal gradient from measurements of the sediment's temperature- and thermal diffusivity-depth profiles (Chapters 7 – 9).

We introduced three example areas (Section 3.2) with different thermal diffusivity profiles and bottom water temperature deviations. All methods – two Newton-like methods, a heuristic linear fitting method and two Markov Chain Monte Carlo methods – yield good results for artificial data with a depth of up to 10 m and satisfying results for realistic artificial data with a depth of up to 6 m (Section 10.1). We can observe that the accuracy of the day of the annual minimum d is limited by the small derivative of the cosine around the extrema. However, a reconstruction within three weeks is already satisfying. The mean temperature A was reconstructed very well for all methods and all example areas. The accuracy of the reconstruction of the amplitude B decreased with increasing relative difference to the mean temperature (see Fig. 19), while the accuracy of the reconstruction of the geothermal gradient increased with increasing depth of the data and decreasing absolute value of the amplitude B.

Under the assumption that the bottom water temperature function can be reasonably well approximated by a shifted cosine as in equation (5.1), we can conclude that a reconstruction

of the parameter of said shifted cosine and the geothermal gradient is possible. The quality of the reconstruction increases with increasing depth of the measured temperature-depth profile. While a maximum depth of 10 m would be ideal, 6 m should be sufficient in most areas. The need for deeper measurements increases with increasing amplitude of the bottom water temperature deviation (see Section 3.3). There were no significant differences between the different basic regions, when it comes to deciding on the *best* algorithm.

The heuristic linear fitting (Algorithm 4) has the best performance, i.e. a good accuracy to time ratio. The achieved accuracy for artificial data is good, but due to the heuristic approach not theoretically backed. The DRAM algorithm (Algorithm 6) on the other hand is rather expensive, but the convergence is proven. However, as seen in the last sections we did not achieve convergence for the realistic artificial data sets with *just* 70000 ensemble members.

The statistical Newton (Algorithm 2) also yields satisfying results for artificial data but did not always converge for the realistic artificial data. It is possible that parameter settings can be found to improve the convergences on the precise data sets but this would not yield any advantage for real data sets. Likewise, the Metropolis Hastings Algorithm (Algorithm 5) did not converge for some artificial data sets.

The simple Newton method (Algorithm 1) yields good results if the geothermal gradient was treated as an input variable rather than a reconstruction goal.

In conclusion, we recommend the use of measurement devices with sensors at maximum depth of at least 6 m, preferably 7.5 m and more, and to use the simple Newton method or the heuristic linear fitting for the inversion to achieve good results in reasonable time and the DRAM algorithm with more than 70000 ensemble members if time cost is not an issue.

Appendix

Inversion Results

In the following section the results of the inversion methods on artificial data are presented. The data is generated from measured thermal properties and geothermal gradient and guessed exact water parameters as presented in Section 3.2. The data was generated over 72 sensors at up to 10 m depth to show the general performance of the method, as well as over 22 sensors at up to 5.95 m depth to model the performance on realistic data. For each setup white noise of 0.5 and 1 % was added and the methods were executed 100 times to calculate mean results and standard deviations. The Markov chain Monte Carlo methods give this information by itself and therefore were only executed once.
All the following tables show the results for all three example data sets for the same setting. The exact parameter set is written in the first row, followed by the mean reconstruction results and the standard deviation and finally the relative reconstruction error of the mean result.

Newton algorithm

Table 6.: Inversion results for the simple Newton algorithm for artificial data with 72 sensors at a depth of up to 10 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Newton Mean	7.50	5.59	52	0.030
Newton SD	0.005	0.016	0.5	0.0000
Newton Rel. Error	0.06 %	1.56 %	4.81 %	0.00 %
Newton with g Mean	7.54	5.63	52	0.025
Newton with g SD	0.026	0.029	0.4	0.0032
Newton with g Rel. Error	0.50 %	2.34 %	5.86 %	17.15 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Newton Mean	10.50	0.31	49	0.084
Newton SD	0.011	0.090	14.1	0.0000
Newton Rel. Error	0.04 %	2.85 %	9.87 %	0.00 %
Newton with g Mean	10.50	0.30	58	0.085
Newton with g SD	0.073	0.129	50.0	0.0075
Newton with g Rel. Error	0.01 %	1.45 %	28.94 %	0.82 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Newton Mean	3.00	0.49	64	0.038
Newton SD	0.004	0.016	4.3	0.0000
Newton Rel. Error	0.08 %	2.52 %	0.87 %	0.00 %
Newton with g Mean	2.99	0.49	60	0.039
Newton with g SD	0.027	0.021	6.2	0.0029
Newton with g Rel. Error	0.26 %	2.36 %	7.14 %	4.27 %

Table 7.: Inversion results for the simple Newton algorithm for artificial data with 72 sensors at a depth of up to 10 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Newton Mean	7.50	5.57	52	0.030
Newton SD	0.010	0.039	1.1	0.0000
Newton Rel. Error	0.05 %	1.22 %	5.10 %	0.00 %
Newton with g Mean	7.54	5.64	52	0.024
Newton with g SD	0.059	0.058	1.1	0.0069
Newton with g Rel. Error	0.47 %	2.47 %	5.72 %	18.55 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Newton Mean	10.50	0.29	54	0.084
Newton SD	0.020	0.140	34.5	0.0000
Newton Rel. Error	0.01 %	1.96 %	19.14 %	0.00 %
Newton with g Mean	10.51	0.34	81	0.084
Newton with g SD	0.138	0.212	80.3	0.0151
Newton with g Rel. Error	0.11 %	11.74 %	80.28 %	0.00 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Newton Mean	3.00	0.48	60	0.038
Newton SD	0.007	0.027	7.4	0.0000
Newton Rel. Error	0.08 %	3.67 %	7.15 %	0.00 %
Newton with g Mean	2.99	0.48	58	0.039
Newton with g SD	0.028	0.037	5.8	0.0036
Newton with g Rel. Error	0.26 %	3.54 %	10.23 %	3.59 %

Table 8.: Inversion results for the simple Newton algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Newton Mean	7.49	5.40	57	0.030
Newton SD	0.012	0.022	0.6	0.0000
Newton Rel. Error	0.15 %	1.77 %	3.82 %	0.00 %
Newton with g Mean	7.41	5.36	58	0.052
Newton with g SD	0.000	0.000	0.0	0.0000
Newton with g Rel. Error	1.24 %	2.57 %	5.37 %	71.92 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Newton Mean	10.50	0.29	48	0.084
Newton SD	0.017	0.084	8.2	0.0000
Newton Rel. Error	0.00 %	3.08 %	6.50 %	0.00 %
Newton with g Mean	10.59	0.48	52	0.066
Newton with g SD	0.085	0.314	2.5	0.0208
Newton with g Rel. Error	0.85 %	58.65 %	15.61 %	21.73 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Newton Mean	3.00	0.49	65	0.038
Newton SD	0.007	0.014	5.3	0.0000
Newton Rel. Error	0.02 %	1.20 %	0.21 %	0.00 %
Newton with g Mean	3.01	0.48	64	0.033
Newton with g SD	0.005	0.001	1.8	0.0019
Newton with g Rel. Error	0.33 %	3.67 %	1.94 %	12.97 %

Table 9.: Inversion results for the simple Newton algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Newton Mean	7.49	5.39	57	0.030
Newton SD	0.021	0.053	1.3	0.0000
Newton Rel. Error	0.19 %	2.08 %	3.80 %	0.00 %
Newton with g Mean	7.38	5.29	58	0.053
Newton with g SD	0.066	0.092	1.0	0.0173
Newton with g Rel. Error	1.58 %	3.80 %	5.37 %	76.49 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Newton Mean	10.50	0.32	59	0.084
Newton SD	0.038	0.175	47.0	0.0000
Newton Rel. Error	0.00 %	6.92 %	30.17 %	0.00 %
Newton with g Mean	10.57	0.49	65	0.067
Newton with g SD	0.238	0.238	58.7	0.0421
Newton with g Rel. Error	0.62 %	62.58 %	44.99 %	20.42 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Newton Mean	3.00	0.49	62	0.038
Newton SD	0.013	0.023	8.2	0.0000
Newton Rel. Error	0.03 %	1.93 %	4.97 %	0.00 %
Newton with g Mean	2.98	0.49	53	0.041
Newton with g SD	0.096	0.094	14.0	0.0229
Newton with g Rel. Error	0.81 %	2.96 %	17.84 %	10.23 %

Table 10.: Inversion results for the simple Newton algorithm for measured data.

Example 1	A °C	B K	d days	g K/m
Newton Mean	10.36	5.65	69	0.030
Newton SD	0.001	0.301	9.5	0.0000
Newton with g Mean	–	–	–	–
Newton with g SD	–	–	–	–
Example 2				
Newton Mean	13.68	0.09	44	0.084
Newton SD	0.001	0.008	9.7	0.0000
Newton with g Mean	13.93	0.29	68	0.044
Newton with g SD	0.689	0.048	52.1	0.0642
Example 3				
Newton Mean	3.20	0.12	63	0.038
Newton SD	0.001	0.013	20.6	0.0000
Newton with g Mean	3.42	0.12	69	0.015
Newton with g SD	0.905	0.081	54.7	0.0755

Statistical Newton

Table 11.: Inversion results for the statistical Newton algorithm for artificial data with 72 sensors at a depth of up to 10 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Statistical Newton Mean	6.71	5.81	76	0.038
Statistical Newton SD	3.487	1.257	63.0	0.0246
Statistical Newton Rel. Error	10.59 %	5.64 %	38.08 %	25.11 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Statistical Newton Mean	10.40	0.26	47	0.083
Statistical Newton SD	0.222	0.144	12.9	0.0180
Statistical Newton Rel. Error	0.96 %	11.70 %	5.07 %	0.81 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Statistical Newton Mean	2.95	0.51	51	0.042
Statistical Newton SD	0.072	0.028	14.2	0.0060
Statistical Newton Rel. Error	1.81 %	2.34 %	21.98 %	12.31 %

Table 12.: Inversion results for the statistical Newton algorithm for artificial data with 72 sensors at a depth of up to 10 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Statistical Newton Mean	9.47	9.23	56	0.025
Statistical Newton SD	12.500	31.566	33.9	0.0402
Statistical Newton Rel. Error	26.27 %	67.89 %	1.11 %	15.24 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Statistical Newton Mean	11.67	0.45	88	0.089
Statistical Newton SD	12.120	1.527	82.0	0.0187
Statistical Newton Rel. Error	11.11 %	49.89 %	96.21 %	5.93 %
Example 3				
Exact Parameter	10.50	0.30	45	0.084
Statistical Newton Mean	11.67	0.45	88	0.089
Statistical Newton SD	12.120	1.527	82.0	0.0187
Statistical Newton Rel. Error	11.11 %	49.89 %	96.21 %	5.93 %

Table 13.: Inversion results for the statistical Newton algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Statistical Newton Mean	22.32	6.43	83	0.081
Statistical Newton SD	90.249	4.014	69.2	0.1810
Statistical Newton Rel. Error	197.62 %	16.92 %	51.06 %	170.97 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Statistical Newton Mean	10.89	0.93	118	0.092
Statistical Newton SD	37.094	4.309	97.8	0.0434
Statistical Newton Rel. Error	3.72 %	208.89 %	161.67 %	9.43 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Statistical Newton Mean	2.88	0.56	51	0.052
Statistical Newton SD	0.210	0.091	14.0	0.0186
Statistical Newton Rel. Error	4.00 %	11.85 %	20.87 %	39.52 %

Table 14.: Inversion results for the statistical Newton algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Statistical Newton Mean	7.04	5.98	55	-0.034
Statistical Newton SD	8.416	1.406	27.0	0.1575
Statistical Newton Rel. Error	6.11 %	8.70 %	0.79 %	212.20 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Statistical Newton Mean	10.28	0.73	147	0.112
Statistical Newton SD	0.593	0.687	96.1	0.0929
Statistical Newton Rel. Error	2.06 %	142.62 %	227.34 %	33.43 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Statistical Newton Mean	2.92	0.55	51	0.050
Statistical Newton SD	0.191	0.092	12.9	0.0194
Statistical Newton Rel. Error	2.76 %	9.41 %	21.28 %	32.29 %

Table 15.: Inversion results for the statistical Newton algorithm for measured data.

Example 1	A °C	B K	d days	g K/m
Statistical Newton Mean	8.71	9.90	124	-0.312
Statistical Newton SD	5.335	7.265	112.0	0.6973
Example 2				
Statistical Newton Mean	17.34	1.44	128	0.064
Statistical Newton SD	2.215	1.526	111.0	0.0453
Example 3				
Statistical Newton Mean	2.96	1.31	105	0.067
Statistical Newton SD	1.400	1.243	70.9	0.0440

Linear Fitting with Day Reconstruction Heuristic

Table 16.: Inversion results for the heuristic linear fitting algorithm for artificial data with 72 sensors at a depth of up to 10 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Linear Fitting Mean	7.54	5.64	52	0.025
Linear Fitting SD	0.023	0.031	0.6	0.0033
Linear Fitting Rel. Error	0.53 %	2.47 %	6.00 %	16.66 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Linear Fitting Mean	10.47	0.22	46	0.089
Linear Fitting SD	0.050	0.144	15.0	0.0063
Linear Fitting Rel. Error	0.29 %	28.23 %	2.69 %	5.79 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Linear Fitting Mean	3.00	0.50	66	0.038
Linear Fitting SD	0.015	0.015	4.9	0.0023
Linear Fitting Rel. Error	0.02 %	0.18 %	1.85 %	0.16 %

Table 17.: Inversion results for the heuristic linear fitting algorithm for artificial data with 72 sensors at a depth of up to 10 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Linear Fitting Mean	7.53	5.61	52	0.027
Linear Fitting SD	0.047	0.071	1.0	0.0070
Linear Fitting Rel. Error	0.34 %	2.08 %	6.00 %	9.78 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Linear Fitting Mean	10.45	0.19	48	0.090
Linear Fitting SD	0.103	0.287	22.8	0.0136
Linear Fitting Rel. Error	0.43 %	36.82 %	7.04 %	7.44 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Linear Fitting Mean	3.00	0.50	65	0.038
Linear Fitting SD	0.023	0.025	7.7	0.0035
Linear Fitting Rel. Error	0.07 %	0.99 %	0.54 %	0.42 %

Table 18.: Inversion results for the heuristic linear fitting algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Linear Fitting Mean	7.38	5.29	58	0.056
Linear Fitting SD	0.084	0.099	1.2	0.0197
Linear Fitting Rel. Error	1.55 %	3.91 %	5.76 %	85.35 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Linear Fitting Mean	10.46	0.22	48	0.090
Linear Fitting SD	0.112	0.244	24.7	0.0205
Linear Fitting Rel. Error	0.35 %	28.18 %	6.42 %	7.44 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Linear Fitting Mean	3.01	0.50	68	0.034
Linear Fitting SD	0.049	0.030	10.1	0.0108
Linear Fitting Rel. Error	0.42 %	0.53 %	3.86 %	9.37 %

Table 19.: Inversion results for the heuristic linear fitting algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
Linear Fitting Mean	7.37	5.28	58	0.056
Linear Fitting SD	0.149	0.175	2.1	0.0348
Linear Fitting Rel. Error	1.67 %	4.05 %	5.80 %	86.71 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
Linear Fitting Mean	10.38	0.06	56	0.104
Linear Fitting SD	0.182	0.393	30.4	0.0337
Linear Fitting Rel. Error	1.14 %	80.42 %	23.96 %	23.84 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
Linear Fitting Mean	3.02	0.52	69	0.033
Linear Fitting SD	0.082	0.053	17.6	0.0173
Linear Fitting Rel. Error	0.53 %	3.47 %	5.74 %	12.08 %

Table 20.: Inversion results for the heuristic linear fitting algorithm for measured data.

	A	B	d	g
Example 1	°C	K	days	K/m
Linear Fitting Mean	12.54	8.31	61	-0.910
Linear Fitting SD	0.996	1.062	3.2	0.4180
Example 2				
Linear Fitting Mean	13.68	-0.06	48	0.072
Linear Fitting SD	0.200	0.519	3.4	0.0250
Example 3				
Linear Fitting Mean	3.22	0.14	101	0.029
Linear Fitting SD	0.005	0.031	0.4	0.0051

Table 21.: Inversion results for the MH and DRAM algorithm for artificial data with 72 sensors at a depth of up to 10 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
MH Mean	7.32	5.13	51	0.061
MH SD	0.304	0.567	0.2	0.0524
MH Rel. Error	2.44 %	6.72 %	8.11 %	103.25 %
DRAM Mean	7.53	5.68	51	0.028
DRAM SD	0.351	0.544	8.4	0.0541
DRAM Rel. Error	0.37 %	3.34 %	6.85 %	8.19 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
MH Mean	10.00	-0.50	48	0.041
MH SD	0.172	0.826	0.7	0.0352
MH Rel. Error	4.77 %	267.19 %	7.21 %	51.08 %
DRAM Mean	10.57	0.49	49	0.070
DRAM SD	0.651	2.163	13.3	0.0839
DRAM Rel. Error	0.65 %	64.18 %	9.08 %	16.41 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
MH Mean	3.54	-0.12	49	-0.014
MH SD	0.348	0.397	0.1	0.0748
MH Rel. Error	18.12 %	124.65 %	25.19 %	138.35 %
DRAM Mean	3.01	0.46	50	0.035
DRAM SD	0.315	0.591	14.2	0.0526
DRAM Rel. Error	0.31 %	7.64 %	22.44 %	6.80 %

Table 22.: Inversion results for the MH and DRAM algorithm for artificial data with 72 sensors at a depth of up to 10 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
MH Mean	7.50	5.46	52	0.023
MH SD	0.176	0.290	0.1	0.0282
MH Rel. Error	0.07 %	0.77 %	6.21 %	24.34 %
DRAM Mean	7.51	5.63	50	0.027
DRAM SD	0.214	0.319	5.1	0.0327
DRAM Rel. Error	0.20 %	2.41 %	8.48 %	9.25 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
MH Mean	10.93	3.41	53	0.025
MH SD	0.691	1.540	1.1	0.0603
MH Rel. Error	4.06 %	1037.66 %	18.72 %	69.84 %
DRAM Mean	10.59	0.49	48	0.065
DRAM SD	0.626	2.180	13.5	0.0828
DRAM Rel. Error	0.84 %	64.44 %	6.91 %	23.08 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
MH Mean	3.32	-0.36	50	-0.003
MH SD	0.235	0.384	0.1	0.0284
MH Rel. Error	10.57 %	172.84 %	23.81 %	109.05 %
DRAM Mean	3.00	0.47	50	0.035
DRAM SD	0.245	0.445	14.1	0.0416
DRAM Rel. Error	0.01 %	6.50 %	22.74 %	6.75 %

Table 23.: Inversion results for the MH and DRAM algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 0.5 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
MH Mean	8.00	6.14	51	0.000
MH SD	0.495	0.733	0.1	0.0887
MH Rel. Error	6.72 %	11.67 %	7.69 %	98.82 %
DRAM Mean	7.56	5.51	55	0.019
DRAM SD	0.382	0.520	7.5	0.0887
DRAM Rel. Error	0.80 %	0.27 %	0.83 %	36.95 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
MH Mean	11.30	3.34	48	-0.007
MH SD	0.436	0.215	0.0	0.0961
MH Rel. Error	7.62 %	1012.96 %	6.71 %	108.30 %
DRAM Mean	10.56	0.32	49	0.058
DRAM SD	0.642	2.161	13.5	0.0930
DRAM Rel. Error	0.56 %	6.73 %	8.41 %	30.78 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
MH Mean	4.46	0.57	51	0.039
MH SD	0.000	0.000	0.0	0.0000
MH Rel. Error	48.55 %	14.68 %	21.29 %	4.07 %
DRAM Mean	3.03	0.44	50	0.021
DRAM SD	0.361	0.603	14.4	0.0902
DRAM Rel. Error	1.15 %	12.77 %	23.33 %	43.60 %

Table 24.: Inversion results for the MH and DRAM algorithm for artificial data with 22 sensors at a depth of up to 5.95 m and 1 % added white noise.

	A	B	d	g
Example 1	°C	K	days	K/m
Exact Parameter	7.50	5.50	55	0.030
MH Mean	7.53	5.52	52	0.036
MH SD	0.251	0.440	0.1	0.0928
MH Rel. Error	0.36 %	0.43 %	5.77 %	19.83 %
DRAM Mean	7.56	5.52	54	0.023
DRAM SD	0.388	0.523	7.3	0.0899
DRAM Rel. Error	0.85 %	0.33 %	1.34 %	24.34 %
Example 2				
Exact Parameter	10.50	0.30	45	0.084
MH Mean	10.72	0.92	52	0.060
MH SD	0.546	0.301	0.0	0.0076
MH Rel. Error	2.12 %	207.69 %	16.30 %	28.91 %
DRAM Mean	10.50	0.25	48	0.061
DRAM SD	0.602	1.998	13.1	0.0921
DRAM Rel. Error	0.04 %	17.98 %	7.72 %	26.98 %
Example 3				
Exact Parameter	3.00	0.50	65	0.038
MH Mean	3.87	0.61	50	-0.020
MH SD	0.213	0.486	0.1	0.0707
MH Rel. Error	29.00 %	22.02 %	22.98 %	152.28 %
DRAM Mean	3.07	0.38	50	0.019
DRAM SD	0.442	0.861	14.1	0.0952
DRAM Rel. Error	2.26 %	23.83 %	23.59 %	48.14 %

Table 25.: Inversion results for the MH and DRAM algorithm for measured data.

	A	B	d	g
Example 1	°C	K	days	K/m
MH Mean	7.19	5.26	49	0.030
MH SD	0.000	0.000	0.0	0.0000
DRAM Mean	9.22	4.33	51	0.026
DRAM SD	1.790	2.639	13.7	0.0997
Example 2				
MH Mean	9.20	0.36	51	0.028
MH SD	0.000	0.000	0.0	0.0000
DRAM Mean	13.04	-0.65	49	0.065
DRAM SD	1.242	3.384	13.8	0.0978
Example 3				
MH Mean	3.83	0.40	53	0.013
MH SD	0.000	0.000	0.0	0.0000
DRAM Mean	3.25	0.08	50	0.019
DRAM SD	0.511	1.114	14.3	0.0982

Bibliography

G. Allaire. *Numerical Analysis and Optimization - An Introduction to Mathematical Modelling and Numerical Simulation*. Oxford University Press Inc., New York, 2007.

G. R. Beardsmore and J. P. Cull. *Crustal Heat Flow: a Guide to Measurement and Modelling*. Cambridge University Press, New York, 2001.

M. Berzins, P. M. Dew, and R. M. Furzeland. Developing software for time-dependent problems using the method of lines and differential- and algebraic integrators. *Applied Numerical Mathematics*, 5:375–397, 1989.

D. Bonté, J. D. van Wees, and J. M. Verweij. Subsurface temperature of the onshore Netherlands: new temperature dataset and modelling. *Netherlands Journal of Geosciences – Geologie en Mijnbouw*, 91–4:491–515, 2012.

H. Brakelmann and J. Stammen. Thermal Analysis of Submarine Cable Routes: LSM or FEM? In *IEEE-conference PECon Putra Jaya, Malaysia*, pages 560–565, 2006.

Bundesamt für Seeschifffahrt und Hydrographie BSH. MARNET-Messnetz, 2014. URL `http://www.bsh.de/de/Meeresdaten/Beobachtungen/MARNET-Messnetz/Stationen/debu.jsp,lastaccess:5June2014`.

E. C. Bullard. Heat Flow in South Africa. *Proc. R. Soc. Lond. A*, 173:474–502, 1939.

H. S. Carslaw and J. C. Jaeger. *Conduction of Heat in Solids*. Oxford at the Clarendon Press, 1959.

G. D. Clow. The extend of temporal smearing in surface-temperature histories derived from borehole temperature measurements. *Paleogeogr., Paleoclimatol., Paleoecol. (Global Planet. Chnage Sect.)*, 98:81 – 86, 1992.

D. Crespo, T. F. Grilo, J. Baptista, J. P. Coelho, A.I Lillebø, F. Cássio, I. Fernandes, C. Pascoal, M. A. Pardal, and M. Dolbeth. New climatic targets against global warming: will the maximum 2°C temperature rise affect estuarine benthic communities? *Scientific Reports*, 7:3918, 2017.

J. H. Davies. Global map of solid earth surface heat flow. *Geochem. Geophys. Geosyst.*, 00: 5, 2013.

M. Dillon, C. Müller, and R. Usbeck. Acquiring Thermal Conductivity Data From Shear-Resistant Sediments. *Sea Technology*, 53:57–61, 2012.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 2000.

L. C. Evans. *Partial Differential Equations*. American Mathematical Society, Providence, Rhode Island, 2010.

J. S. Gray and H. Christie. Predicting long-term changes in marine benthic communities. *Mar. Ecol. Prog. Ser.*, 13:87–94, 1983.

H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7: 223–242, 2001.

H. Haario, M. Laine, and A. Mira. DRAM: Effective adaptive MCMC. *Stat. Comput.*, 16: 339–354, 2006.

K. Hancke, B. K. Sorell, L. C. Lund-Hansen, M. Larsen, T. Hancke, and R. N. Glud. Effects of temperature and irradiance on a benthic microalgal community: A combined two-dimensional oxygen and flourescence imaging approach. *Limnol. Oceanogr.*, 59(5):1599–1611, 2014.

M. Hanke-Bourgeois. *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens*. Vieweg+Teubner, Wiesbaden, 2009.

A. Hartmann and H. Villinger. Inversion of marine heat flow measurements by expansion of the temperature decay function. *Geophys. J. Int.*, 148(3):628–636, 2002.

R. D. Hyndman, E. E. Davis, and J. A. Wright. The measurement of marine geothermal heat flow by a multipenetration probe with digital acoustic telemetry and insitu thermal conductivity. *Mar. Geophys. Res.*, 4:181–205, 1979.

International Heat Flow Commission (IHFC). The International Heat Flow Commission: Global Heat Flow Database, 2014. URL `http://www.heatflow.und.edu/index2.html`. (last access: June 2017).

E. A. Irene. *Electronic Materials Science*, chapter Diffusion in Solids. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2005.

J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer Science+Business Media, New York, 2004.

M. F. Lavìn and S. G. Marinone. *Nonlinear Processes in Geophysical Fluid Dynamics*, chapter An Overview of the Physical Oceanography of the Gulf of California, pages 173–204. Kluwer Academic Publishers, Dordrecht, 2003.

W. Lowrie. *Fundamentals of Geophysics*. Cambridge University Press, Cambridge, 2007.

L. Madsen. Approximate geothermal gradients in denmark and the danish north sea sector. *Danm. geol. Unders.*, pages 5–16, 1974.

O. I. Mamayev. Chapter 1 Introduction. In *Temperature - Salinity Analysis of World Ocean Waters*, volume 11 of *Elsevier Oceanography Series*, pages 1–7. Elsevier, Amsterdam, 1975.

MathWorks. MATLAB Documentation as of Version R2006a, 2017. URL `http://mathworks.com/help/`.

W. McLean. *Strongly Elliptic Systems and Boundary Integral Equations*. Cambridge University Press, Cambridge, 2000.

T. Merck and R. Wasserthal. *Assessment of the environmental impacts of cables*. OSPAR Commission, London, 2009.

F. Miesner. Reconstruction of Bottom Water Temperatures from Instantaneous Measurements of Temperature and Thermal Diffusivity in Marine Sediments. Master's thesis, University of Bremen, Department of Mathematics, 2013.

F. Miesner, A. Lechleiter, and C. Müller. Reconstructing bottom water temperatures from measurements of temperature and thermal diffusivity in marine sediments. *Ocean Sci.*, pages 559–571, 2015.

A. Mira. On Metropolis-Hastings algorithms with delayed rejection. *Metron.*, LIX (3–4): 231–241, 2001.

C. Müller, F. Miesner, R. Usbeck, and T. Schmitz. 2K-criterion: measuring and modelling temperatures and thermal conductivities/diffusivities in shallow marine sediments. In *Conference on Maritime Energy 2013, TUHH, Hamburg*, pages 475–490, 2013.

H. N. Pollack, P. Y. Shen, and S. Huang. Inference of ground surface temperature history from subsurface temperature data: Interpreting ensembles of borehole logs. *Pageoph*, 147: 537–550, 1996.

J. Rhode, P. Tett, and F. Wulff. *The Seas, Volume 14*, chapter 26. The Baltic and North Seas: A regional review of some important physical-chemical-biological interaction processes, pages 1029–1071. Harvard University Press, Cambridge, 2007.

M. H. Ribergaard. *Oceanographic Investigations off West Greenland 2011*. Danish Meteorological Institute Centre for Ocean and Ice, Kopenhagen, 2011.

A. Rieder. On the regularization of nonlinear ill-posed problems via inexact Newton iterations. *Inverse Problems*, 15:309–327, 1999.

A. Rieder. *Keine Probleme mit Inversen Problemen: Eine Einführung in stabile Lösungen*. Vieweg, Wiesbaden, 2003.

J. M. Robles and S. G. Marinone. Seasonal and interannual thermohaline variability in the Guaymas Basin of the Gulf of California. *Continental Shelf Research*, 7:715–733, 1987.

J. G. Sclater, D. Hasterok, J. Hillier B. Goutorbe, and R. Negrete. *Encyclopedia of Marine Geosciences*, chapter Marine Heat Flow. Springer Science+Business Media, Dordrecht, 2014.

P. Y. Shen and A. E. Beck. Least Squares Inversion of Borehole Temperature Measurements in Functional Space. *J. Geophys. Res.*, 96:19 965–19 979, 1991.

R. D. Skeel and M. Berzins. A Method for the Spatial Discretization of Parabolic Equations in one Space Variable. *J. Sci. Stat. Comput.*, 11:1–32, 1990.

J. E. Smerdon, H. N. Pollack, V. Cermak, J. W. Enz, M. Kresl, J. Safanda, and J. F. Wehmiller. Daily, seasonal, and annual relationships between air and subsurface temperatures. *Journal of Geophysical Research: Atmospheres*, 111(D7), 2006.

R. C. Smith. *Uncertainty Quantification: Theory, Implementation and Applications*. Society for Industrial and Applied Mathematics, Philadephia, 2014.

C. A. Stein. Heat Flow of the Earth. *Global Earth Physics*, pages 144 – 158, 1995.

C. Stranne and M. O'Regan. Conductive heat flow and nonlinear geothermal gradients in marine sediments – observations from ocean drilling program boreholes. *Geo-Mar Lett*, 36:25–33, 2016.

E. Sundvor and A. M. Myhre. *Component parts of the World Heat Flow Data Collection*. PANGAEA, Bremerhaven, 1987.

C. Zucco, W. Wende, T. Merck, I. Köchling, and J. Köppel. *Ecological Researchon Offshore Wind Farms: International Exchange of Ecperiences*, volume 186. BfN-Skripte, Berlin, 2006.