ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

CORSO DI LAUREA MAGISTRALE IN
INGEGNERIA ENERGETICA

DEVELOPING A DATA PROCESSING FRAMEWORK FOR ANALYZING BOREHOLE TEMPERATURE PROFILES IN GEOTHERMAL HEAT PUMP SYSTEMS WITH DTS TECHNOLOGY

Relatore

Prof. Claudia Naldi

Correlatore

Prof. Alberto Lazzarotto

Presentata da

Matilde Mascellani

Sessione Marzo 2025

Anno Accademico 2023/2024

Table of contents

Chapter 1

Introduction

1.1 Background and Motivation

In May 2023, Emilia-Romagna experienced in just 36 hours an unprecedented downpour, equivalent to the typical rainfall of seven months. The disaster resulted in 17 fatalities and an estimated €10 billion in damages. In the province of Ravenna alone, 27,775 people were evacuated, with infrastructure, farmland, roads, rail lines, and the broader economy suffering extensive damage still awaiting full assessment [1], [2].
One year later Bologna, capitol city of region, faces renewed floodings. The causes of the floods are several and of different nature, with climate change and global warming at the top of the list.

This scenario, along with other extreme weather events and climate catastrophes, is becoming increasingly common worldwide, and it stresses the urgency to address modern climate challenges. A viable approach is to increase the share of renewable sources in the total primary energy consumption, complemented by high-efficiency systems.
In line with this focus European countries signed and ratified the Paris Agreement in 2015, committing at reaching net-zero greenhouses gas emissions by 2050. The strategies to achieve climate neutrality were outlined in the European Green Deal, launched in 2019, with key actions including energy saving, diversification of energy supplies and clean energy production [3]. Specifically, the target for 2030 is to bring renewable energy sources to 42.5% of the European Union's final energy use (they represented an estimated 24.1% in 2023) [4].

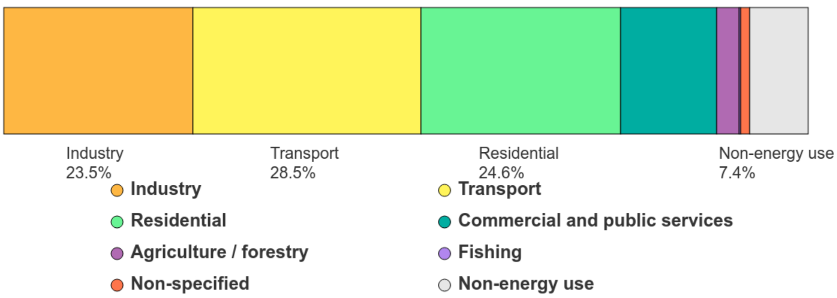


Figure 1: Total Final Energy Consumption in Europe for the year 2022 [4]

The residential sector accounted for 24.6% of the final energy use in 2022 [5]; around 80% of this share is consumed for space and water heating [6]. In light of this, renewable sources heat pumps, such as ground source ones, are playing an increasingly central role in the energy sustainability of the residential sector.

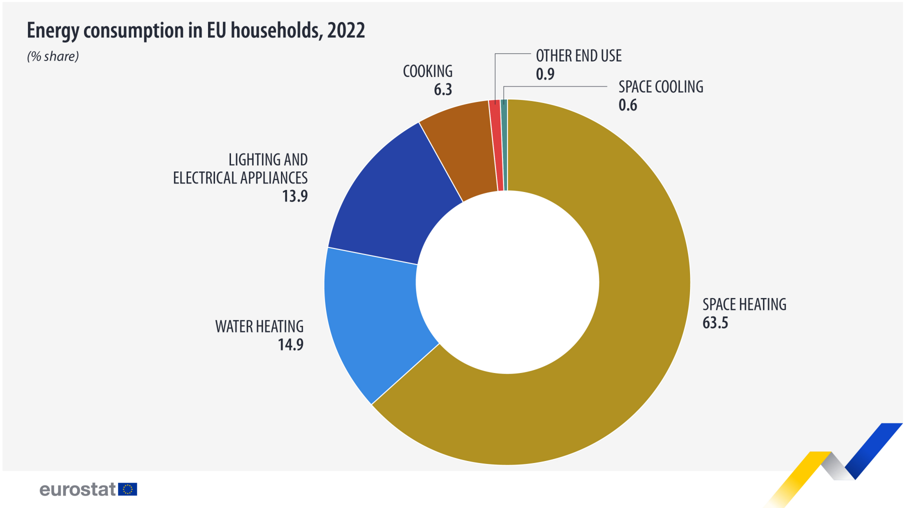


Figure 2: Energy Consumption in EU Households for the year 2022 [7]

Ground source heat pumps (GSHPs) allow extraction of geothermal energy stored under the Earth's surface, and continuously produced by the decay of particles in the core. This technology is widely implemented in northern European countries, where winters are particularly cold. In Sweden geothermal energy is dominated by low temperature systems, and more than 95% are GSHPs for space and domestic hot water heating for single-family buildings. On the 31st of December 2019 there were more than 590000 geothermal heat pumps installed, for a total capacity of 6680 MWt [8].

In Italy the total number of GSHPs registered in 2018 is about 15000, resulting in 800 MWt capacity installed. The data, provided by *GSE* and *EurObserv'ER(2018)*, is based on collected information from the main heat pumps producers and sellers, but it lacks accuracy, due to difficulties on obtaining updated and homogeneous data. The most widespread technologies within the Italian geothermal sector are district heating and ground source heat pumps, that have doubled their capacity since 2010. This trend has been enhanced during the last few years by several factors. Legislative requirements to increase renewable energy share of buildings together with the strong interest in protecting the air quality, especially in northern Italy, are promoting the conversion of fossil fuel-based heating (and cooling) systems to low-enthalpy geothermal systems. The combination of geothermal systems with other energy renewable sources, such as already installed photovoltaic systems, largely diffused in the territory, is perceived as an easy solution to be promoted in this regard. Furthermore, the development of innovative and various solutions in geothermal field is creating new contexts where the use of GSHPs is possible.
In conclusion, Italian ground source heat pumps market is not as developed as in other european countries, but conditions are favorable for its expansion [9], [10].

1.2 Significance of Data Processing in Geothermal Research

Knowledge of local geology and borehole properties is essential for accurate system dimensioning and performance optimization.
Prior to GSHP system installation and configuration choices, thermal response tests (TRTs) are conducted to investigate the ground properties. Conducting a thermal response test is both time-consuming and expensive, similar to other methods aiming at characterizing geothermal

sources. Unlike solar, hydro, or other renewable resources, the potential of geothermal energy sources is not easy to assess. On the other hand, once the data is collected, it can be applied to multiple projects over long periods, providing valuable insights for future developments [11].
Data collection is essential for the creation of data-driven models that can replace more complex mathematical models. These models enable dynamic analysis, performance evaluation and the prediction of various potential scenarios. Physical models require the formulation of equations governing the system, a challenging task due to difficulties in estimating necessary parameters. Accordingly, data-driven models are gaining more and more popularity, dragged by the advancements in Big Data processing and storage.
The main limitation of data-driven models lies in the quality and quantity of available data. For data to be truly useful and accessible, it must be well-structured, easily understandable, securely stored and preserved. The lack of well-organized and comprehensive data prevents data-driven models from expressing their full potential [12]. Moreover, a weak data management framework can increase costs and limit opportunities within the geothermal sector. On the contrary, reliable, accurate and standardize data enable researchers to set realistic energy targets, attracting investments, supporting and enhancing the sector's ability to compete with other renewable energy sources [13].
In conclusion, effective management and proper data handling are critical for the sustainable development of geothermal resources at economically viable costs [14].

This project aims to create a strong, reliable process for collecting ground temperature measurements produced in the last five years. The measurements were taken along the lengths of two boreholes, located within the same borehole field.
The goal is to organize and process this data, making it clear, structured, and ready for use.
Secondly, the focus will be on analyzing the available data to estimate the interference between the two close boreholes and evaluate the temperature profiles over time. By doing so, a method to post-process the data, extracting the desired information, will be provided.

1.3 Objectives of the Thesis

A distributed temperature sensing system (DTS) has been installed in 2016 in two different boreholes. It has been detecting the ground temperature of the two close boreholes, at different depths, at each time step. One of the two boreholes has been actively used along this period of time, while the second one was only used for experimental porposes, rarely subjected to heat injection or extraction.
Building upon the significance of accurate data management in geothermal research, the objectives of this thesis are:

1. To develop a structured and reproducible methodology for processing borehole temperature data.

2. To analyze temperature patterns and the historical interaction between two boreholes of the same field.

3. To lay the groundwork for future testing and system optimization.

Chapter 2

Background on GSHP Systems and DTS Technology

2.1 Ground-Source Heat Pumps

Heat pumps operate by transferring heat from a source to a sink, which is at higher temperature than the source. This is obtained thanks to the work of the compressor (W in Figure 3), allowing the working fluid, a refrigerant, to reach a state of high temperature and pressure (point *2* in Figure 3). The refrigerant is directed into the condenser, where it releases heat to the sink (Q_2 in Figure 3). Eventually the fluid is fully condensed and in the liquid phase (point *3* in Figure 3). The liquid refrigerant, still under high pressure, then flows through an expansion valve or throttling device. This component reduces the refrigerant's pressure and temperature, returning it to a biphasic stage of low pressure and temperature (point *4* in Figure 3). In a second heat exchanger, the evaporator, the remaining liquid part is completely vaporized thanks to the heat provided by the source (Q_1 in Figure 3). At this point, the refrigerant is back to its original condition (point *1* in Figure 3). and can enter the compressor restarting the cycle. The exchanged energies are linked: $Q_1 + W = Q_2$.
These technologies could eventually be used for cooling purposes, working as refrigerators. In the latter case, the source would be in place of the sink, while the sink is the environment that provides heat for the evaporation phase.

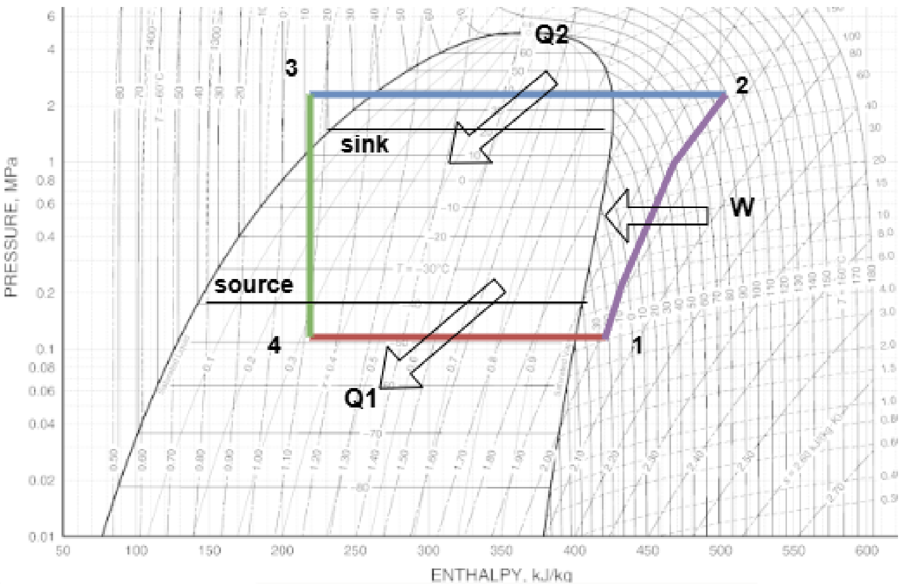


Figure 3: Refrigeration cycle Diagram

The efficiency of the heat pump is measured by the ratio of the desired output (heat) to the work (energy) required to produce it and, in accordance with Figure 3, it's called Coefficient of Performance and expressed as follows: $COP = \frac{Q_2}{W} = \frac{Q_2}{Q_1 - Q_2}$.
The generated heat is larger than the energy used by the compressor, leading to COP values greater than 1 and on average around 3.5 [15].
HPs can be classified according to the heat source and the sink, the most common types being

air-source HPs (air-to-water and air-to-air), water source HPs (water-to-air and water-to-water) and ground (brine) source HPs (ground-to water and ground-to-air). In most of the worldwide installed heat pumps, the heat source is ambient air. In this setup the external unit in contact with the source, is generally small and easy to install even in pre-existing buildings, making air source heat pumps a practical solution, especially in dense urban areas. Ground source heat pumps use the heat stored underground as a source, to be extracted through copper or plastic tubes buried in the ground. Therefore, GSHPs require a sufficient large available land, as well as a high initial investment. The field size depends on the type of soil, characterized by its thermal conductivity, and the installation setup. Open-loop systems directly extract water from natural sources such as aquifers, lakes or rivers, which typically offer a stable and moderate water temperature (5–10°C), enhancing the heat pump's performance. However, resource availability is limited and often restricted by environmental regulations, and the system's components are easily subjected to corrosion. The most common setups are closed-loop systems, where horizontal or vertical collectors are filled with water and anti-freeze mixture, which transfers heat from the borehole to the refrigerant loop in the heat pump. Multiple vertical wells containing heat exchange tubes, collectively forming a borehole field, can maximize the collectors' surface while requiring relatively little land. This makes them a popular choice, particularly in areas with limited space.

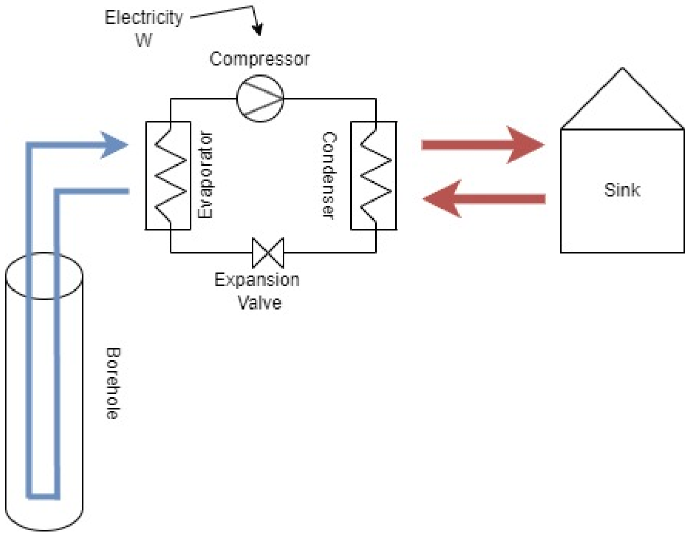


Figure 4: GroundSource Heat Pump connected to a vertical borehole

The heat pump performance is related to several factors. The temperature difference (ΔT) between the outlet temperature (to the sink) and the ground source temperature. Lower (ΔT) leads to higher COP.

Low-temperature heating systems and proper sizing of the ground collectors are critical, and their inadequate design can lead to efficiency losses. GSHPs with vertical boreholes tend to have better performance due to stable ground temperatures at deeper levels compared to shallow, horizontal configurations. These factors expose a key drawback of air source HPs compared to GSHPs.

The earth's underground temperature stability makes GSHPs effective year-round, regardless of external weather fluctuations [16]. Therefore, the underground temperature profiles at different times of the year have been modeled as shown in the following figure.

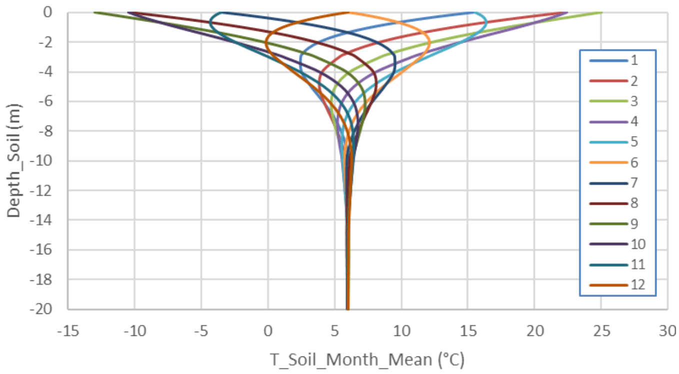


Figure 5: Underground Temperature Profiles [16]

As a result, GSHPs deliver higher COP and SCOP values over the year and typically achieve better values even under standard operating conditions, since air volumetric specific heat capacity is significantly lower than that of soil or water [15], [17].

2.2 Overview of Distributed Temperature Sensing (DTS) Technology

As already mentioned, thermal response tests (TRTs) are the most popular method to estimate the ground thermal properties, especially the ground thermal resistance and its thermal conductivity, which are essential design and optimization parameters of geothermal heat pump systems.

Response tests involve imposing a known thermal load on the borehole and analyzing the temperature response over time, allowing the determination of key factors such as ground thermal conductivity, thermal resistance, ground diffusivity, or the influence of groundwater flow on heat transfer efficiency. An essential metric revealed through TRTs is the undisturbed ground temperature, which reflects the baseline temperature of the ground and determines the temperature gradient. This drives heat transfer and influences overall performance, as efficiency depends on the difference between the ground temperature and the working fluid temperature [18].

The borehole thermal resistance per unit length is defined as $R_b = \frac{T_m - T_b}{q_l}$ where T_m is the mean temperature of the fluid inside the borehole (usually water or a water-antifreeze mixture), T_b is the mean temperature at the borehole's external surface and q_l is the average thermal power exchanged per unit length between the ground and the fluid. Thermal resistance can be determined through analytical approximations or numerical simulations. Analytical methods rely on the line source model, treating the pipes as infinitely long thermal sources placed at the pipe axes.
Typical TRT consists of two main phases: first, the measurement of the undisturbed ground temperature, and second, the heating phase, where heat is generated, mainly through electrical resistances, and injected, while the inlet and outlet temperatures of the circulating fluid are recorded. The second phase is approximately 36 to 48 hours long. After completing the test, the recorded data is analyzed and fitted to the temperature evolution equation $T_m - T_g = A\ln(t) + B$, allowing the constants A and B to be determined through linear interpolation. From these constants, the thermal conductivity of the ground is calculated as $k_g = \frac{q_l}{4\pi A}$ and the borehole thermal resistance is determined using $R_b = \frac{1}{q_l}(B - \frac{4\alpha_g}{k_g}\ln(a))$, where α_g is the thermal diffusivity of the ground [19], [20], [21].

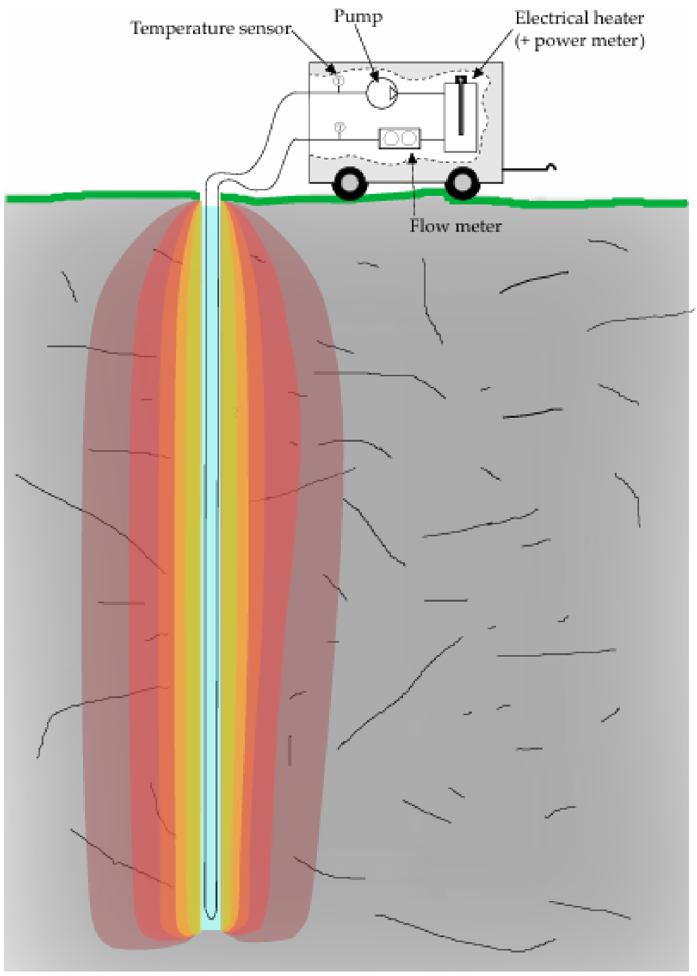


Figure 6: Setup for a Thermal Response Test [22]

Standard TRTs only measure the borehole inlet and the borehole outlet fluid temperatures, using the mean as averaging method. This approximation assumes uniform heat flux along the borehole, prioritizing horizontal conduction as the dominant heat transfer mechanism over vertical conduction and groundwater flow advection. The resulting effective borehole capacity and thermal conductivity estimates often lead to an overestimation of the borehole length.

Higher precision can be obtained with Distributed Thermal Response Tests (DTRTs).
DTRTs measure the entire vertical fluid temperature profiles using fiber optic cables placed along the boreholes, eliminating reliance on mean temperature approximations. This approach allows local estimation of thermal properties, directly utilizing the spatial and temporal distribution of temperatures. The spacial variations in subsurface heat transfer can be derived. Generally, DTRTs enable more accurate, detailed, and reliable evaluations [23], [24].

The optical fibers mentioned are part of Distributed Temperature Sensing (DTS) systems. DTS technology implements Raman optical time domain reflectometry, specifically leveraging the Raman scattering phenomenon. This refers to the frequency scattering of laser light pulses injected from an optical fiber, and subsequently reflected, re-emitted and detected. The frequency of the output signal has one part at lower frequencies (stokes) and one part at higher frequencies (anti-stokes) than the original laser lights [25]. The frequency difference is correlated to the energy gap between the scatter lights. The incident photon may turn into a phonon (a molecular quantum vibration), resulting in lower frequency, or it may capture a phonon energy, resulting in higher frequency [26].

The ratio between stokes and anti-stokes intensity is related to the Temperature in accordance with Bose-Einstein statistics:

$$\frac{P_S}{P_{AS}} \propto e^{\frac{\Delta E}{k_B T}} \tag{1}$$

Where k_B is the Boltzman constant, T is the temperature, P_S and P_{AS} are, respectively, the Stokes and anti Stokes intensities and $\Delta E = \hbar\Delta\omega$ is the energy difference between incident and scattered light [27].

The temperature may be explicitly inferred [28]:

$$T(z,t) = \frac{\gamma}{\ln(\frac{P_S(z,t)}{P_{AS}(z,t)}) + C(t) - \Delta R(z) - \Delta\alpha z} \tag{2}$$

Where z is the distance from the fiber, t is the time and $\gamma = \frac{\Delta E}{k_B}$. The equation accounts through the term C for the fraction of light back-scattered to the instrument and the instrument detector effciency. The potential step losses ΔR represent the concentrated losses that might be caused by different sections' connection, or local strains, fiber damages… Finally, the term $\Delta\alpha$ accounts for the difference in the power attenuation coefficients for the Stokes and anti-Stokes signals.
DTS instruments average temperature readings over continuous fiber sections. The measurements accuracy depends on the fiber's refractive index, the density of data captured by the instrument and the size of the measured section. Better accuracy is obtained when more photons are observed per unit time. Longer sections or distances reduce the photon density and therefore the precision. The user can set the time steps duration at which the measurements are performed, affecting the quality of output temperatures, which results from the trade-off between the precision in time and space. Longer integration times improve temperature resolution, but resolution decreases with distance due to signal attenuation. Calibration procedure ensures reliable results by correcting offsets and differential losses. This is often done by placing fibers in environments with known temperatures, such as ice baths. There are two types of setups: single-ended and double-ended. Most measurements are performed in single-ended mode, sending laser pulses in one direction, but double-ended mode, where pulses are sent from both ends, allows compensation for differential losses and attenuations. The uncertainty of the temperature values arises from a variety of factors and it's not easy to estimate. The calibration procedure, mainly the number and placement of bath, directly influences the accuracy changing ΔR and $\Delta\alpha$. Temperature fluctuations in the environment can induce fictitious drifts in the measurements, as well as variations in fiber materials and installation errors, leading to systematic deviations. Additionally, the flow regime plays a role: laminar flow introduces systematic uncertainty due to uneven temperature profiles near pipe walls [25], [29], [30], [31], [32].

Temperature is a significant element for different entities' observation and/or evaluation. The DTS systems' capability to monitor it in real time, makes the technology effective and useful in various applications, especially, but not exclusively, in the hydraulic and geothermal fields. Such applications include studies on surface hydraulic properties [33], soil moisture content [34], tree species influencing forest canopy temperatures [35] or boundary layer height [36].

Regarding geothermal investigations, DTRTs are just an example of practical DTS implementation. In a recent case study, a DTS system was employed to analyze the thermal performance of district-scale geothermal borefields and to understand subsurface heat transfer mechanisms. The data provided was critical for identifying variations in thermal properties across depths and lithological layers, which standard models often oversimplify [37]. Variations in streambed temperature, measured using a distributed temperature sensor (DTS) were the key to identify discrete groundwater discharge zones in a stream [38]. DTS was also used for groundwater flow characterization in an aquifer [39]. By detecting temperature variations of geothermal wells, temperatures by DTS enable to identify distinct flow zones, providing insights into flow contribution and flow assurance issues, enhancing production strategies [40].
Nevertheless, DTS is helpful in monitoring the general performance of a large-scale GSHP. It captures dynamic variations in temperature reflecting the subsurface heat transfer patterns, whose knowledge is a validation criterion of Borehole Heat Exchanger (BHE) numerical models. Additionally, the system's ability to detect thermal anomalies helps in identifying inefficiencies, such as uneven heat distribution or operational faults in the GSHP [41]. Assessing the local soil's thermal response, thanks to DTS technology, during the operation of a Dual-Source Heat Pump (DSHP) in ground-source mode, helped the estimation of the soil's recovery rate after heat extraction. It also validated the potential of undersized BHE configurations to meet heating demands across varying operational scenarios [42].

2.3 Data Processing Methods for GSHPs and DTS systems

Sensornet and Silixa are the main manufacturers for Distributed Temperature Sensing Technologies.
The raw data produced is typically stored in folders in the form of configuration files (CFG) and text files (DDF), which are automatically created at every new measurement. The volume of data can be very large and it is stored in a format that is not directly meaningful, making detailed analysis and evaluation challenging and difficult to manage.

In most case studies, the temperatures are used to train data-driven simulation models, always requiring steps to process and refine the data. Literature offers limited content focused on techniques for handling DTS data.
A Research on detecting pipeline leaks was carried out using DTS data collection. It involved several days of temperature monitoring along the pipeline, resulting in daily datasets containing more than 50000 samples. To manage the data effectively, the raw information was organized in different datasets, after a preprocessing procedure which involved transforming the 2D temperature-time data into 3D temperature-time-distance images. Machine learning models, including Long Short-Term Memory, Convolutional Neural Networks, and Autoencoders, were explored to analyze the DTS data [43]. A similar approach was performed to improve the accuracy of the model of an aquifer, monitored with DTS system, using Assisted History Matching method. Due to the high amount of data, a part of it could be used for history matching, while a different part was set aside for validating the updated models. In this case study data preprocessing involved identifying and excluding poor-quality data, as well as assessing and adjusting the observation error [44]. A different investigation combined the temperature data with Distributed Acoustic Sensing to estimate the flow profiles and assess the dynamics of a wellbore. Finite-element and multiphase flow simulators were used to interpret DTS data, later presented as temperature profiles or spatial maps [40].

DTS data is often combined with other sensors to provide a more comprehensive view of the study subject. It is the case for HP systems, usually equipped with temperature sensors, flow meters, pressure sensors, and power meters. As part of an analysis conducted on GSHP performance, DTS temperature measurements were taken at high frequency. A centralized Building Management System was used to log data from all sensors, including the corresponding uncertainties, while raw data was periodically backed up. Missing or inconsistent data was interpolated or imputed, and statistical and numerical techniques were used to analyze patterns, trends, and system performance [41].
Building Management Systems (BMS) is an integrated real-time management framework that monitors and controls smart buildings. It is one of the available tools to handle systems network, including GSHPs. A comprehensive data model for GSHP systems should include static details (e.g., equipment specifications) and dynamic time series data (e.g., flow, temperature, and power) [45]. Industry-standard tagging systems like Project Haystack can be adopted for this purpose. Project Haystack is an open source tool that provides standardized tags to help define the relations between different entities [46]. Haystack tags can be enhanced using Brick, a schema designed for representing metadata about buildings' subsystems, which are graphically represented in hierarchical classes. The semantic interoperability ensures extensibility and integration with BMS [47]. An alternative effective solution to store organized data is offered by online databases, such as PostgreSQL or InfluxDB, where data and metadata are stored in relational tables or time-series databases linked via unique identifiers.

Tools are available for managing data specifically produced by DTS devices [48].
The CTEMPs MATLAB DTS Toolbox provides multiple graphical user interfaces designed to calibrate DTS data and process it, in order to parse the files into MATLAB compatible datasets. Users need to specify the files formats and the directory where they are saved, additional information regarding fiber length limits, temperature reference and few other parameters are required for the calibration procedure [49]. Nonetheless, if files are spread across multiple folders, users must organize them beforehand. Overall, the manual interventions required can be time-intensive, especially with large amount of data, not to mention that incorrect selections can significantly degrade the data output quality.
An additional limitation is the significant increase in processing time as dataset size expands, together with the user's need to be familiar with MATLAB, being able to customize the script as desired. DTSGUI is public software that implements tools mainly from Python libraries to import, manage fiber-optic distributed temperature sensor data and visualize it in the form of heat maps of temperature versus distance and time. The loaded data must come from a DDF files folder, trimmed adjusting the minimum distance value and can therefore be automatically displayed, and eventually exported for further use [50]. Once again, if files are distributed in multiple subfolders they must be re-arranged, in spite of that, the software is intuitive and directly provides a variety of plots.
Finally, two Python libraries exist that are dedicated to DTS data calibration and processing. Python DTS Calibration Toolbox was developed by Delft University of Technology as a calibration tool. It expects raw data files as inputs and the fiber section used as calibration reference. The output consists of different types of graphs displaying the temperature uncertainties computed at each length value [51].
The University of Bayreuth Micrometeorology is responsible for the creation of Python DTS Data Processing Toolbox, designed for processing large and long-term DTS setups specifically for Silixa devices. It automates calibration, leveraging the dts calibration Python package described above, and mapping of data. It focuses on data conversion into netCDF format, where

the outputs are physically labeled data along with temperature and intensity metrics [52].

This thesis presents an alternative method, similar to those mentioned, for transforming DDF files content into more readable and accessible data, primarily using Julia programming language.

Chapter 3

System Description and Data Acquisition

3.1 Boreholes and Setup Description

3.1.1 Boreholes Description

The boreholes monitored with the DTS system are part of a larger borehole field located on the Royal Institute of Technology (KTH) main Campus, in Stockholm, in 2016. This field consists of 12 boreholes, 11 of which follow the common single U-tube setup and serve as ground source heat exchangers for a GSHP system. This system consists of three heat pump units, each containing two vapor compression cycle, and provides space heating and domestic hot water for three student accommodation buildings. Stockholm is characterized by a cold climate, making heating the predominant need, while cooling is not actively provided.

Unlike standard commercial systems, this borehole field incorporates unique design features. One of the boreholes, the 100-meter coaxial one, referred to as number 1, is dedicated exclusively to research. It is part of a project at Live-in Lab, a research platform at KTH, serving as testbed for smart and sustainable building technologies. The latter BHE is made of a central rigid tube in High Density PolyEthylene (HDPE), with a double wall filled with air, ensuring insulation between upward and downward flows, reducing the mutual heat exchange (the thermal shunt). The outer tube is a membrane, called Energy Capsule, which is kept in direct contact with the borehole walls by static pressure.
Moreover, the boreholes vary in length, ranging from 100 to 350 meters, and are not perfectly vertical but rather inclined and spread in different directions, as illustrated in Figure 7. This geometric distribution optimizes heat exchange with the ground while minimizing thermal interference between neighboring boreholes at deeper levels [22], [53].

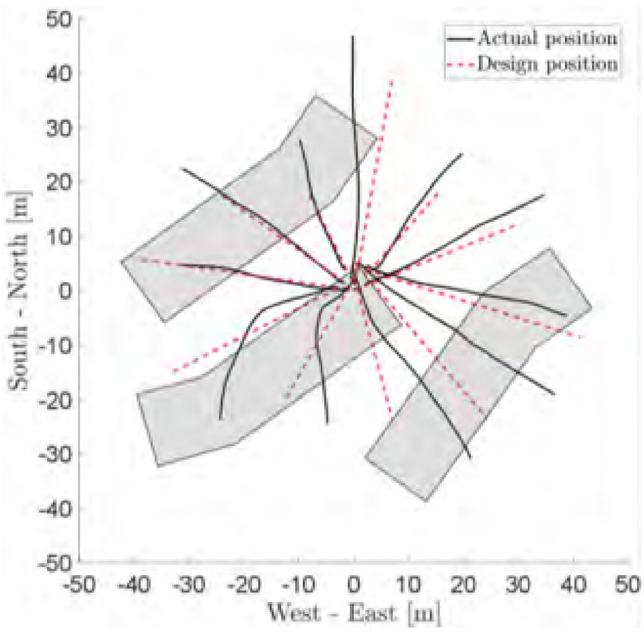


Figure 7: Top view of the borehole field [22]

Two boreholes are equipped with optical fibers part of a distributed temperature sensing system:

- **Borehole 1**
  - Length: 100 meters
  - Orientation: Straight
  - Configuration: Coaxial
  - Status: mainly inactive (briefly used for tests)
  - Fiber position: Inside the inner pipe (undetermined distance from the central axis)
  - Borehole Diameter: 115 mm
  - Inner Diameter: 35.2 mm
  - Outer Diameter:114 mm
  - Middle Diameter:50 mm
  - Pipe characteristics: Inner tube in HDPE, outer tube is an Energy Capsule
  - Borehole filling material: Groundwater and air
  - Secondary fluid type: Water-ethanol mixture (concentrations varied based on the tests performed during that period)
  - Design effective thermal resistance: 0.05 mK/W
- **Borehole 10**
  - Length: 350 meters
  - Orientation: Inclined
  - Configuration: Single U-tube
  - Status: Active
  - Fiber position: Between the pipe and the borehole wall (undetermined location within this space).
  - Borehole Diameter: 115 mm
  - Pipe characteristics: PEM DN40 PN8
  - Borehole filling material: Groundwater and grout
  - Secondary fluid type: Water-ethanol 28%-wt
  - Design effective thermal resistance: 0.05 mK/W

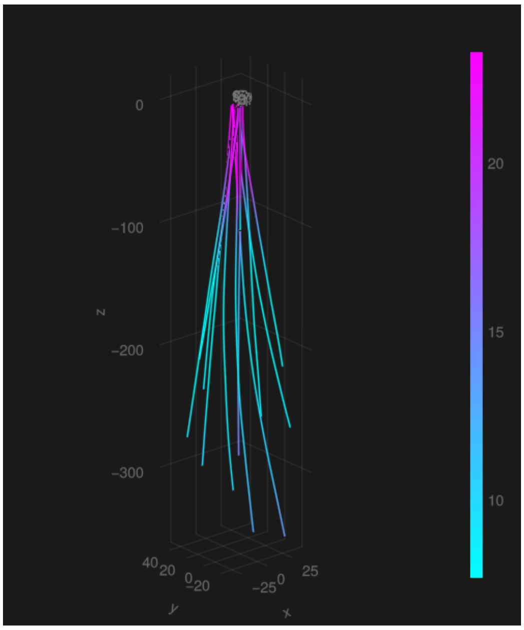
The field is situated in an area with relatively low groundwater level (43 meters below the surface) [54], due to its location on a 42-meter-high hill [55]. The deeper part of the boreholes was filled by groundwater, while, to ensure efficient heat transfer, the boreholes were grouted over the dry section, except for the research borehole, which uses a coaxial design to minimize thermal shunt.

Figure 8: 3D view of the buildings heated by GSHPs connected to the borehole field [55]

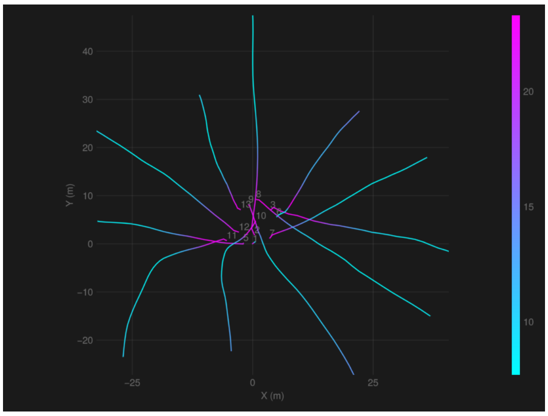
This coaxial borehole is also integrated into a test rig that includes a Thermia heat pump, piping system, pumps, valves, heat exchangers, a water tank, sensing devices, and a control system. The system allows for heat pump operation in both extraction and injection modes, with controlled power modulation. While the test rig is still being optimized, this borehole has only been used briefly in the past few years and is to be considered inactive. However, its distinct features make the setup ideal for further research, performance evaluations, and investigations [56].

For the characterization of the ground, the main parameters are reported below [22]:
- Loop type: Closed loop
- Ground composition: Metamorphic rocks
- Groundwater level below ground: 43 m
- Undisturbed ground temperature: 9.8 °C
- Design ground thermal conductivity: 3.4 W/(mK)
- Design volumetric heat capacity: 2.2 MJ/(kgK)

The borehole position was determined based on Figure 7 (a drawing from the design phase) and cross-checked with the design coordinates listed on the Geological Survey of Sweden website [54]. Based on available data, the recorded position aligns with reality. Additionally, the deviation measurements were conducted using a DevDrill ("Peewee" model) Devico machine, which has an integrated survey system that continuously records inclination and tool orientation in real time. The data are available in the form of CSV tables with North, East and Elevation data detected every 10 meters along each borehole [57]. This data allows for a complete reconstruction of the 3D geometric distribution of the underground boreholes. The analysis results are presented in the following figure.



(a) 3D view



(b) 2D view

Figure 9: 3D and 2D views of the Borehole field.

3.1.2 Data Collection Methods

At KTH campus two out of 12 of the boreholes are provided with a DTS measuring system: number 1 and 10, respectively referred as channel 1 and 2. The optical fibers were installed in 2016, simultaneously with the drilling of the boreholes. The optical fiber is in the inner pipe for the coaxial Borehole 1, while it lays outside of the U-tube pipes of Borehole 10. At that time, few tests were run, but the current final setup, from which continuous and consistent measurements have been produced, dates back to 2020. The data are progressively saved and stored by HaloDTS software in a computer powered by Windows 2000. This computer, along with the two connected fibers, was previously located in different areas, and possibly installed in a different system than a borehole. The Distributed Temperature Sensing system, ran by HaloDTS, is manufactured by Sensornet, a United Kingdom corporate office. It continuously stores the results from the measurements in files, organized across a hierarchy of folders. The main folder, called *SensornetDTS*, consists of 3 subfolders: *full data set*, *latest data* and *temperature only* (an empty folder). In the first one, data from channel 1 and 2 are divided in many additional subfolders that refer to different locations, projects, years and months of the year.

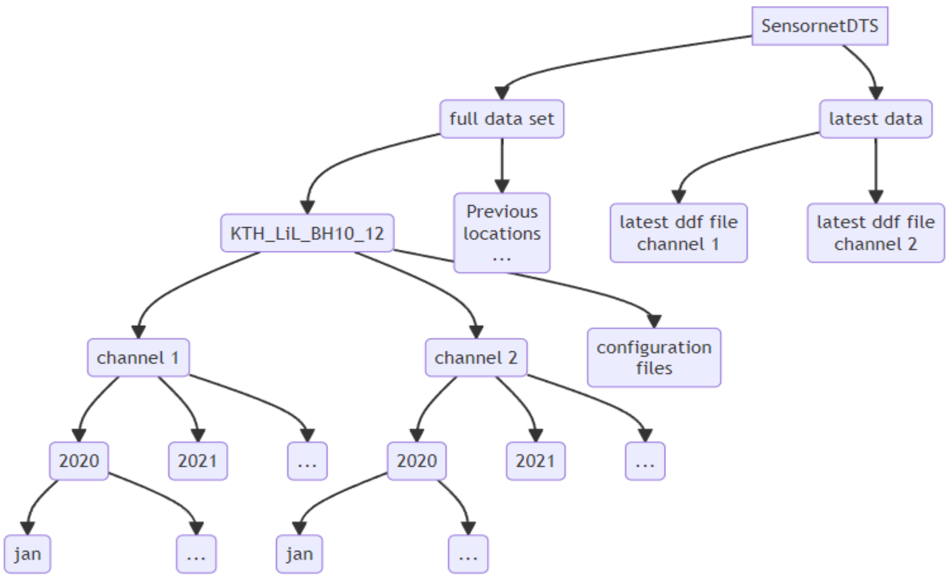


Figure 10: Diagram of the folders' structure

There's one folder, named *KTH LiL BH10_12*, for the ongoing project, where the data produced since 2020 is safely saved and updated every new measurement. In latest data only the very last created files are stored, allowing easier and faster access for current tests values.
The data is primarily represented in the form of DDF files, where the first lines contain metadata, data about the data, while the following section contains columns with length values, corresponding to a position along the fiber, temperature, stokes and anti-stokes values relative to that position. Other interesting produced files are the configuration ones, providing additional metadata regarding the current settings and calibration parameters, such as temperature reference or the reference section of the fiber.

3.2 Data Processing Pipeline

The routine to transform the raw temperature data into actionable insight consists of a reproducible sequence of steps.

All the metadata and data contained in every DDF file was organized in tables and stored in a PostgreSQL Database. Once the material was saved, it was possible to select specific information from the database, and export them into a local directory. An important step entailed determining the filtering criteria to identify the valuable part of the data. The next step involved smoothening of the data, reducing the noise effects, followed by an interpolation method to evaluate the missing values. The processed temperature data was prepared for visualization as heatmaps, 3D plots, or contour plots, depicting variations over time and distance across different time intervals.

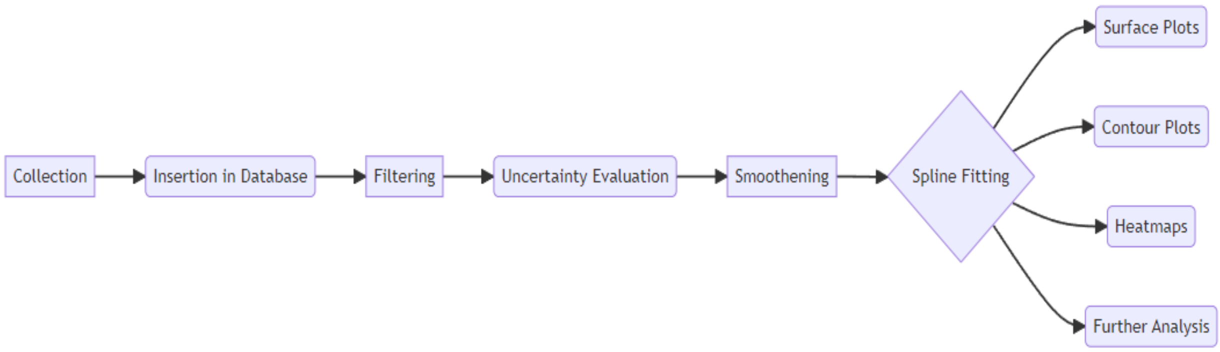


Figure 11: Diagram of the Data Processing Pipeline

3.2.1 Collection and Insertion in Database

In this work, a method is developed for data collecting and processing to facilitate the insertion of the measurements' results into a PostgreSQL database. The database was designed to store key information in an architecture that enables efficient storage, retrieval, and analysis of large volumes of data. The folders were copied from the desktop computer running Windows 2000 to a laptop by means of a USB drive. Using Julia programming language, a code was implemented to crawl through all folders, identify DDF files, extract their contents and structure it into data frames.
It extracts the metadata section and assigns a value for each line, including the following categories.
- **datetime**: Refers to the year-month-day and time of the day at which the file was created.
- **Installation**: Refers to the folder where the file will be located.
- **Differential loss correction**: Single-ended or double-ended setups.
- **Forward channel**: Refers to the tested channel, either channel 1 or 2.
- **Forward acquisition time**: Time interval for each forward measurement.
- **Gamma**: Calibration coefficient relating to the differential attenuation of the Stokes and anti-Stokes signals.
- **Default loss term (dB/km)**: Describes signal attenuation along the fiber due to material properties.
- **Fiber end**: Reference to the fiber's physical termination point.
- **T internal reference**: Used to calibrate the system, typically from stable temperature baths.

The rest of the data was included in a different table composed of 6 columns:
- **datetime**: Exported from the corresponding metadata.
- **channel**: Exported from the corresponding metadata (*Forward channel*).
- **length**: Position along the fiber where the measurement was performed.
- **temperature**: Temperature value measured.
- **stokes**: Stokes intensity value corrresponding to the measurement.
- **anti-stokes**: Anti-Stokes intensity value corrresponding to the measurement.

A third table was created to store the content of configuration files. The approach mirrored the previously described method: the same code was used to crawl through the folders, identify CFG files, and process each line. The first part of each line was converted into a DataFrame header, under which it was stored the second part of the line, transformed into a single string object. These files provided additional calibration parameter and general setting infromation,

such as:
- **spatial averaging**: The distance between two following measurement points along the fiber.
- **measurement time**: The time interval between two following measurements.
- **internal reference start**: Position along the fiber where the reference section starts.
- **internal reference end**: Position along the fiber where the referece section ends.
- **range**: Total length of the fiber section involved in measurements.
- **range in points**: Number of points in the fiber where the measurements are performed.

Each described table has either the datetime or the filename column, used to sort the data and trace the source of the measurement, ensuring the interconnectivity of the three tables.

The optical fibers are much longer than the boreholes, with significant portions of the cables lying outside the ground, mainly before the start of the boreholes, to connect to the computer located in a basement room, running under the building to accommodate excess length. Temperature values, on the other hand, are generated along the entire length of the fibers. The same fibers were previously used in different installations at various locations. However, they were all connected to the same computer and software, meaning the files in the folders refer to all these installations, not just the most recent one at KTH. As a result, the portion of the fibers actually along underground varied from one setup to another.
For this reason, the next step focuses on identifying the relevant section of the cable for evaluating the temperature within the boreholes.

HaloDTS software dispenses an interface that automatically displays a plot of the temperature over the length, enabling prompt visualization of the areas with high instability and inaccurate temperature readings. Once the dataset was saved and available in the personal device, it was possible to recreate it.

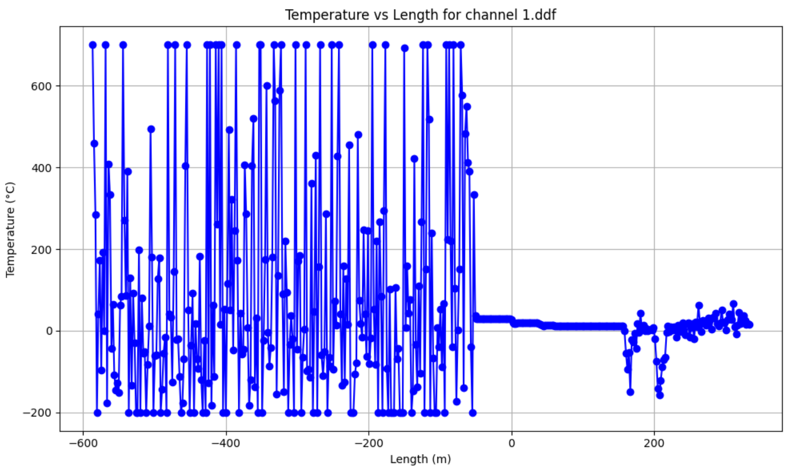


Figure 12: Temperature over Length Scattered Plot

The figure refers to a measurement of the current Live-In Lab installation, and clearly illustrates the unreliability of the temperature values for certain sections of length. Strong, rapid oscillations and out-of-scale absolute values were the criteria used to identify the relevant portion of the fibers. Specifically, the limits for the absolute values were set at -15.0 °C and 40.0 °C, a wide range considering that the temperature of the ground in Stockholm varies within a small range around 10 °C [58], [59]. Moreover, the segments where the temperature difference between two consecutive readings exceeded 3°C were excluded. The chosen criteria are relatively broad, especially considering that the analysis was applied to all files, including those from previous projects with a likely different placement of the optical fibers. This approach was taken to upload all folders from the computer into the database. The widest range was adopted as a filter to store data in the database. As a result, inaccurate values were still selected and saved, nonetheless some parts could be excluded, reducing memory usage. Considering that the data in the database can always be post processed and deeper filtered, it is preferable to store more than needed, avoiding the omission of important values.

To optimize memory usage, each file was analyzed, refined, and inserted into the database one at a time.
Julia offers a library, called *LibPQ*, that provides tools for connecting to PostgreSQL databases. With access credentials (username and password), users can perform various operations, such as querying tables to retrieve specific columns or modifying table structures, including adding columns or updating headers. For this work, queries were especially useful for selecting only the necessary columns from the tables. This approach enhances memory efficiency, allowing to work with smaller chunks of data at a time.
After selecting, inserting, and removing duplicates, the final size of the dataset (sum of the three tables) was approximately 8.5GB.

3.2.2 Filtering

This section describes the additional filtering strategy implemented to select the valuable part of Data, focusing only on the measurements produced in the current installation in KTH campus. Although the configuration and metadata tables provided information about the beginning and end point of the fiber, comparing the reported values with the actual measurement data revealed significant inaccuracies, rendering the information unreliable. The installation setup document [22] reported the lengths of each borehole, included the ones monitored. Those values matched the ones of the Geological Survey of Sweden (SGU) [54], the government authority responsible for the acquisition and collection of geological data of the national territory. The only reliable information at this point was the total length of the boreholes.
The data in the database still contained inaccurate temperature values that should be disregarded, as explained in the previous section. Given the lack of precise reference, the identification of the length filters was carried on through direct analysis of the data itself. A Julia script was developed to extract specific rows and columns from the database based on channel and datetime values, resulting in a dataframe whose section could be represented by Table 1a. The temperature, the datetime and the length columns were reshaped by unstacking them to form a matrix where each column represented temperature values at a specific length (in ascending order), and each row corresponded to a unique datetime (ordered chronologically).

Table 1.a: Example of data section extracted from the database
Table 1.b: Matrix obtained by unstacking Table 1

(a) Database Section

datetime	length	temperature
datetime1	length1	T_{11}
datetime1	length2	T_{12}
datetime1	length3	T_{13}
datetime2	length1	T_{21}
datetime2	length2	T_{22}
datetime2	length3	T_{23}
datetime3	length1	T_{31}
datetime3	length2	T_{32}
datetime3	length3	T_{33}

(b) Unstacked Matrix

datetime	length1	length2	length3
datetime1	T_{11}	T_{12}	T_{13}
datetime2	T_{21}	T_{22}	T_{23}
datetime3	T_{31}	T_{32}	T_{33}

Next, a differential matrix was created by calculating the differences between consecutive temperature values along the columns of the original matrix. Consequently, the differential matrix had the same number of rows but one less column compared to the original, either the first one or the last one depending on the differential method adopted. As shown in Figure 12, the first or last column contains discardable values, so no valuable information was lost by excluding them in this last matrix.

Table 2: Differential Matrix

datetime	length2	length3
datetime1	T_{12} - T_{11}	T_{13} - T_{12}
datetime2	T_{22} - T_{21}	T_{23} - T_{22}
datetime3	T_{32} - T_{32}	T_{33} - T_{32}

Subsequently, it generated a Boolean matrix, populating it with true values where the differential matrix values were less than or equal to a specific threshold, and the absolute temperature values fell within a range defined by the user. Otherwise, the matrix entries were marked as false. Finally, it produced an indexes vector by summing the values along each column of the Boolean matrix, knowing that true was treated as 1 and false as 0. This vector was then plotted against the length to identify length segments with high index values, indicating where all the described criteria were satisfied. The longest segment, where the indexes value was at least 90% of the maximum, was considered as the one corresponding to the actual fiber position.

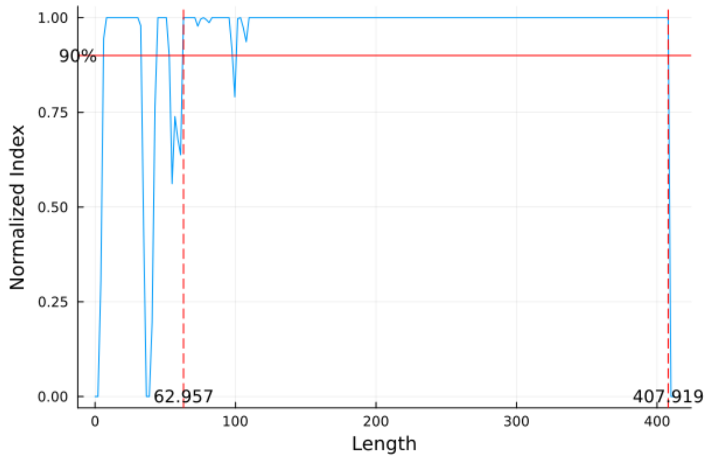


Figure 13: Length Filtering Criteria (Channel 2)

This code could be easily applied changing the filtering criteria as wanted. In the current case, the maximum temperature is set to $30°C$, the minimum temperature to $-5.0°C$, and the threshold for the difference between two consecutive temperatures is $0.8°C$. For Borehole 10, this resulted in minimum valid length being 62.957m, while maximum being 407.919m, for an overall segment of 344.962m. This type of approximation is acceptable in the context of purely qualitative analysis. For the shorter Borehole 1, the minimum valid length was set at 44.694m, while maximum at 154.271m, for an overall segment of 109.577m. While the end of the fiber is clearly identifiable in both cases (see Figure 13), determining the beginning is more challenging, especially considering that the oscillatory behavior near the surface may not be noise related but rather due to actual interactions with the external air, or between the boreholes. Knowing the total length fro be useful for deriving the starting point. In the case of Borehole 10, the total computed length is smaller than the length indicated in the official reports (350m), although only by a few meters. Therefore the initial value was taken as the final value minus the length reported in the documents $(407.919 - 350.0)$m $= 57.919$m. Similar approach was adopted for Borehole 1, whose lenght was longer than expected (100.0m), resulting in initial point at $(154.271 - 100.0)$m $= 54.271$m.

3.2.3 Uncertanty Evaluation

The procedure to estimate uncertainty is complex and not singular, as it relies on various factors, including the calibration method, the experimental setup, and the machine itself, as detailed in Chapter 2. This thesis evaluates uncertainty using the *dtscalibration* Python library, which calculates the variance of temperature at each spatial length measurement. These calculations are performed either through a linear approximation or using Monte Carlo methods. A description of the steps executed by this package is provided below. The temperature is derived from the logarithmic ratio of Stokes and anti-Stokes intensities,represented as $I(x,t)$, therefore, temperature uncertainty estimation requires the analysis of the variance in the Stokes and anti-Stokes measurements.

$$T(x,t) = \frac{\gamma}{C(t) + I(x,t) + \int_0^x \Delta\alpha(x')\, dx'} \tag{3}$$

$$I(x,t) = \ln\left(\frac{P_+(x,t)}{P_-(x,t)}\right) = \frac{\gamma}{T(x,t)} - C(t) - \int_0^x \Delta\alpha(x')\,dx' \tag{4}$$

Where γ depends on the sensitivity of scattering to temperature which is related to the fiber material, $\Delta\alpha$ is the differential attenuation, η is an additional correction parameter and accounts for detector sensitivity and attenuation between the detector and the fiber end connected to the DTS system. A lumped effect parameter, $C(t)$, addresses gain differences and scattering intensity dependencies on wavelength, which is constant along the fiber but must be evaluated at each time step. These parameters are derived during calibration using reference sections with known temperatures. In single-ended setups, as Stokes and anti-Stokes intensities are measured in a single direction from one fiber end, $\Delta\alpha$ is assumed to be constant, allowing further simplification of the temperature expression.

$$\int_0 x\Delta\alpha(x')\,dx' \approx \Delta\alpha \cdot x \tag{5}$$

$$T(x,t) = \frac{\gamma}{C(t) + I(x,t) + \Delta\alpha \cdot x} \tag{6}$$

The intensity ratio is expressed in a discrete form for each time n and fiber length m.

$$I_{m,n} = \frac{\gamma}{T_{m,n}} - \Delta\alpha x_m - C_n \tag{7}$$

This system is reformulated in matrix form for multiple locations and time steps.

$$y = Xa + \epsilon \tag{8}$$

$$y = \begin{bmatrix} I_{1,1} \\ I_{1,2} \\ \vdots \\ I_{M,N} \end{bmatrix} \quad X = \begin{bmatrix} \frac{1}{T_{1,1}} & -x_1 & -1 & 0 & 0 & \cdots & 0 \\ \frac{1}{T_{1,2}} & -x_1 & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T_{M,N}} & -x_M & 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \quad a = \begin{bmatrix} \gamma \\ \Delta\alpha \\ C_1 \\ C_2 \\ \vdots \end{bmatrix} \tag{9}$$

The parameters, collected in vector a, are determined by minimizing the sum of squared residuals using a Python library integrated into *dtscalibration*. Weights are assigned based on measurement noise to solve the system. The accuracy of variance estimation improves with the number of data samples collected from the DTS system.

$$\begin{aligned} \chi^2 &= (y - Xa)^T W (y - Xa) \\ a &= \left(X^T W X\right)^{-1} X^T W y \end{aligned} \tag{10}$$

The variance of stokes and anti-stokes measurements is evaluated using the distribution of their residuals. The latter are computed by fitting the equation to measured values, focusing particularly on the reference sections where the temperature remains approximately constant spatially, so that Stokes and anti-Stokes intensities are described as products of terms that depend respectively only on time or space.

$$\begin{aligned} P_+(x,t) &= G_+(t) \cdot H_+(x) \\ P_-(x,t) &= G_-(t) \cdot H_-(x) \end{aligned} \tag{11}$$

Measurements are collected over time, and the terms are determined through optimization techniques, such as least squares, to minimize the differences between measured and modeled intensities. The modeled intensities are calculated using the fitted parameters, while residuals are the differences between measured and modeled values.

$$\Delta P(x,t) = P^{\text{measured}}(x,t) - P^{\text{modeled}}(x,t) \tag{12}$$

$$\sigma_P^2 = (\Delta P(x,t))^2 \tag{13}$$

The noise variance is approximated by dividing the residual sum of squares by the degrees of freedom

$$I(x,t) = \ln\left(\frac{P_+}{P_-}\right) \quad \rightarrow \quad \sigma_I^2 = \frac{\sigma_{P_+}^2}{P_+^2} + \frac{\sigma_{P_-}^2}{P_-^2} \tag{14}$$

A normal distribution is assigned to both noise and calibration parameters, with the mean set as the measured value and the previously calculated variances defining the distribution's spread. The probability density functions are propagated through the model using either a linear approximation or Monte Carlo sampling, following JCGM guidelines. Monte Carlo sampling generates numerous temperature realizations by repeatedly sampling distributions and computing values for each location. These realizations approximate the probability density functions of the estimated temperature at each location and time. Standard uncertainties are derived from the standard deviations of the temperature realizations, while 95% confidence intervals are determined from the 2.5% and 97.5% percentiles of these realizations. In this work, the linear approximation approach was adopted, as it required less time and computational effort, with further refinement planned for a later stage [51].

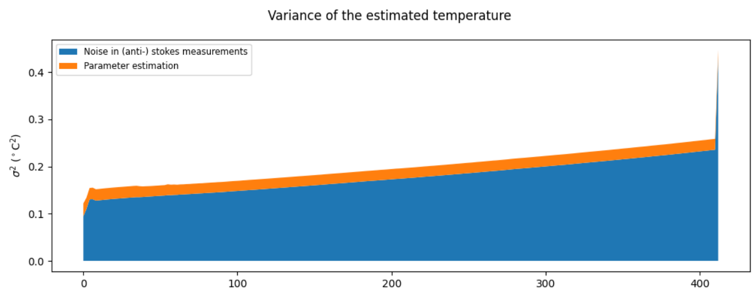


Figure 14: Temperature Variance for channel 2

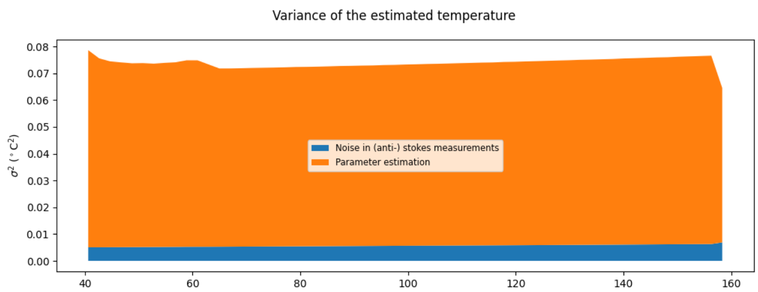


Figure 15: Temperature Variance for channel 1

The described procedure is applied when the functions from the python package are executed, providing the plots shown and a vector of the temperature variances values for each fiber length.

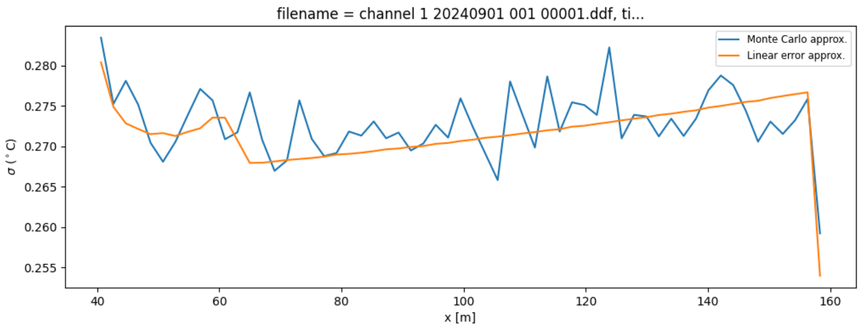


Figure 16: Linear and Monte carlo Approximation

3.2.4 Smoothening and Refinement

The procedure described in paragraph 3.2.2 is repeated to obtain the temperature matrix structured as Table 1b. After removing the first column containing the datetime values, the matrix contained all the temperature values from the sensors, measured during the time selected when extracting data from the database. The values are extremely noisy, and when directly plotted they would reproduce a strong oscillatory behavior, making it difficult to distinguish the actual dynamics and the random fluctuations.

A Kalman filter was applied to the matrix to reduce noise and provide a more accurate temperature estimate.
The Kalman filter is an algorithm widely used in control systems, robotics, finance, and signal processing to estimate the state of a system from noisy observations. It works by combining model-based predictions with noisy measurements, producing an optimal estimate of the system's state. The model equation describes how the system evolves over time, in this case, a simplified linear model was used:

$$x_t = A \cdot x_{t-1} + w_t \tag{15}$$

Here, x_t represents the system's state at time t, and A is the transition parameter reflecting the increasing ($A > 1$) or decreasing ($A < 1$) trend of the model. However, since temperature fluctuations were expected to be rapid and inconsistent, predicting a clear overall trend was challenging, A was set to 1.0, assuming the system follows a constant trend in the data unless corrected by measurements. The parameter w_t, the process noise, represents the uncertainty in the system's model, it allows the filter to deviate from the predictions, preventing overconfidence. The measurement model is described by the equation

$$z_t = x_t + v_t \tag{16}$$

It links the state x_t to the observed measurement z_t. Where v_t accounts for measurement errors introduced by the sensors, already charachterized throught the uncertainty evaluation, which provided variance values associated with each length.
Initially the algorithm predicts the state at time t x_t^{pred}, before considering any measurement,

as well as the error covariance P_t^{pred}, which forecasts the uncertainty of the next state based on the system model.

$$x_t^{\text{pred}} = A \cdot x_{t-1}^{\text{est}} \tag{17}$$

$$P_t^{\text{pred}} = A \cdot P_{t-1}^{\text{est}} \cdot A^T + Q \tag{18}$$

Secondly the Kalman gain is computed:

$$K_t = \frac{P_t^{\text{pred}}}{P_t^{\text{pred}} + R_t} \tag{19}$$

K_t determines the weight given to the measurement compared to the prediction: if R_t is small (indicating a reliable measurement), K_t approaches 1, giving more weight to the measurement, if R_t is large (unreliable measurement), K_t gets close to 0, favoring the prediction.
The state estimate is updated, considering the difference of the predictions from the measurement results:

$$x_t^{\text{est}} = x_t^{\text{pred}} + K_t \cdot (z_t - x_t^{\text{pred}}) \tag{20}$$

Consequently, the uncertainty is reduced and the covariance error updated as shown below [60], [61], [62]:

$$P_t^{\text{est}} = (1 - K_t) \cdot P_t^{\text{pred}} \tag{21}$$

The measurement system, although capable of capturing significantly more temperature values than standard non-distributed sensors, still evaluates temperature at discrete spatial points. To make the dataset x_i, y_i more meaningful and easier to visualize, it can be interpolated to obtain a continuous function of temperature over time and space. Various interpolation methods are available; in this case, splines have been adopted. Splines are piecewise-defined polynomial functions joined at their endpoints, called knots. The resulting function is continuous and has continuous derivatives up to a certain order at the knots. The polynomial order determines the type of functions used to represent the segments between consecutive data points. For cubic splines, the segment between x_i and x_{i+1} is expressed as:

$$S_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \tag{22}$$

The implementation of basis splines (B-splines) ensures computational efficiency and numerical stability, enabling the construction of splines S as a linear combination of these basis functions.

$$S(x) = \sum_{i=1}^{n} c_i B_i(x) \tag{23}$$

where $B_i(x)$ is the *i*-th basis spline function, and c_i are coefficients that determine the contribution of each basis function. Each basis function is nonzero only within a specific interval of the domain, by combining B-splines of different orders, one can construct splines of varying degrees (e.g., linear, quadratic, cubic). By default, splines are configured to pass through all given data points, providing a true interpolation. In this case, the smoothness coefficient s is 0, and the knot locations coincide with the data points. An approximation can be achieved by allowing the splines to deviate from some data points in a trade-off between accuracy (fidelity to the data) and smoothness, particularly suited for noisy datasets. Julia offers the Dierckx package, which facilitates spline approximation or interpolation over a given dataset. It supports multi-dimensional splines through Spline2D and SplineND. Users can dynamically adjust the level of accuracy by modifying the number of knots, the smoothness coefficient s, and the polynomial order. When $s > 0$, the number of knots is automatically reduced, resulting in fewer degrees of freedom for the spline and greater noise suppression, therefore reducing computational cost. Moreover, it is possible to concentrate knots in regions of rapid data variation (high curvature), but the knots selection and setting must be performed manually.

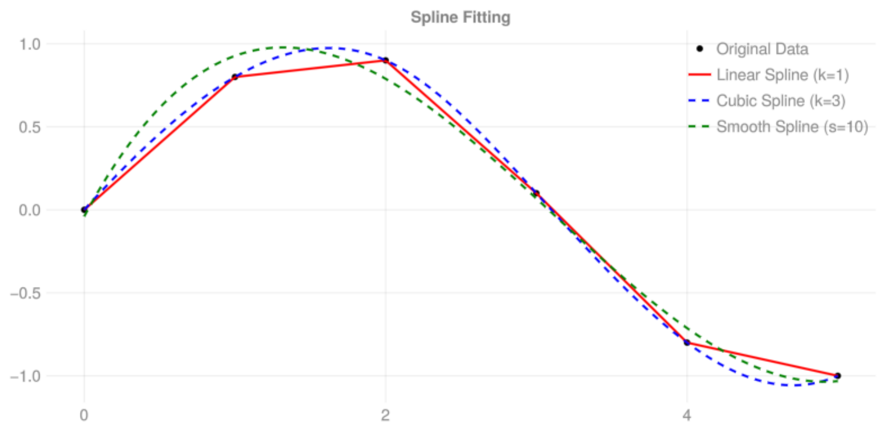


Figure 17: Example of Spline Interpolation and Approximation

The parameter s balances the residual error and the smoothness, determining how closely the splines fits the data and overall aiming at minimizing the penalized residual sum of squares P:

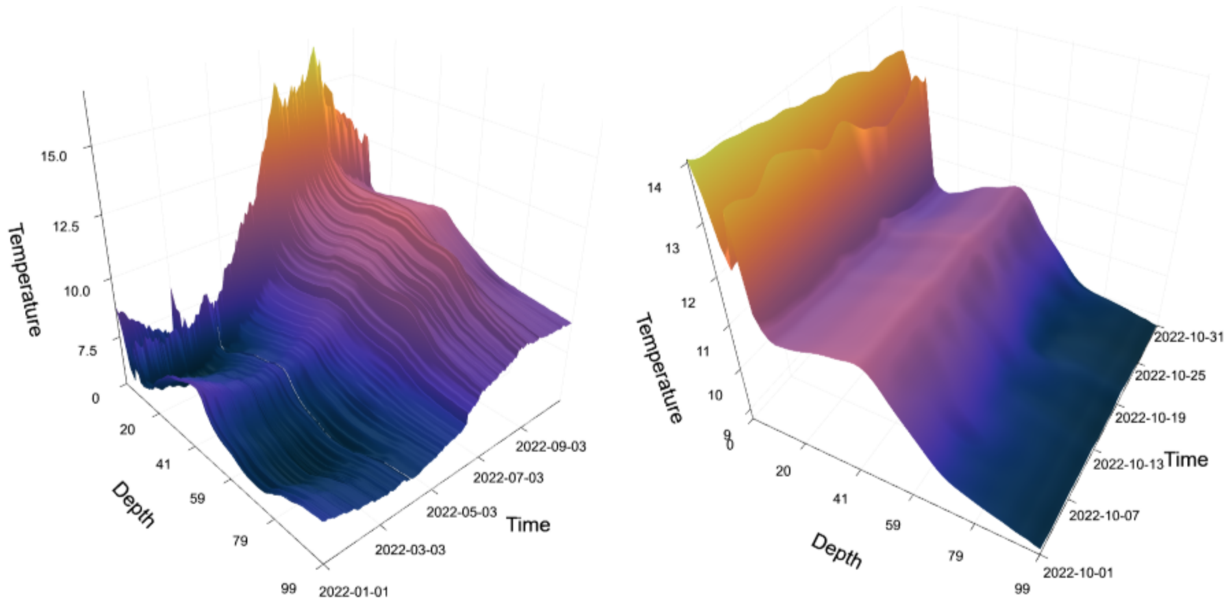
$$P = \sum_i w_i(y_i - S(x_i))^2 + s \int \left(S^{(k-1)}(x)\right)^2 dx \tag{24}$$

where $\sum_i w_i(y_i - S(x_i))^2$ represents the weighted sum of squared residuals fp, where w_i is the weight assigned to the *i*-th squared residual. The term $\int \left(S^{(k-1)}(x)\right)^2 dx$ acts as smoothness penalty, larger value indicates a more oscillatory spline, because it corresponds to greater changes in the *(k−1)*-th derivative of $S(x)$. As s increases, the smoothness term dominates, leading to a larger fp.
The difference between fp and s governs the fitting process: $\frac{|\text{fp}-s|}{s} <$ tol, where tol is the convergence tolerance. The optimal value of s depends on the dataset structure (its noise level) and the spline degree k. Choosing an appropriate s requires testing multiple values for each dataset segment. An excessively high s may result in low fidelity to the actual temperature data, erasing critical dynamics of borehole heat transfer. Conversely, a very small s may lead to high noise levels, causing the plotted data to misrepresent real dynamics and become difficult to interpret. Additionally, when s is too small, the spline-fitting algorithm may fail to converge within the maximum number of iterations set by the program. This happens because the weighted sum of squared residuals does not meet the convergence criterion, $\frac{|\text{fp}-s|}{s} <$ tol. A small s forces the spline to follow the data too closely, leading to numerical instability in the optimization process, especially for high-order splines [63], [64], [65].

Starting from the temperature matrix obtained after the application of the Kalman filter, Dierckx 2D splines were performed. The output object is a function that returns the approximated and interpolated temperature value for any possible combination of time and length. The error introduced by this approximation method was calculated as the difference between each element of the Kalman matrix and the corresponding element of the spline object. Both the mean and maximum absolute errors, along with the fp value, were used to guide the selection of the s parameter. Essentially, an iterative approach is employed to find the optimal balance between sufficient smoothness, to eliminate abrupt temperature fluctuations, and minimizing excessive errors.

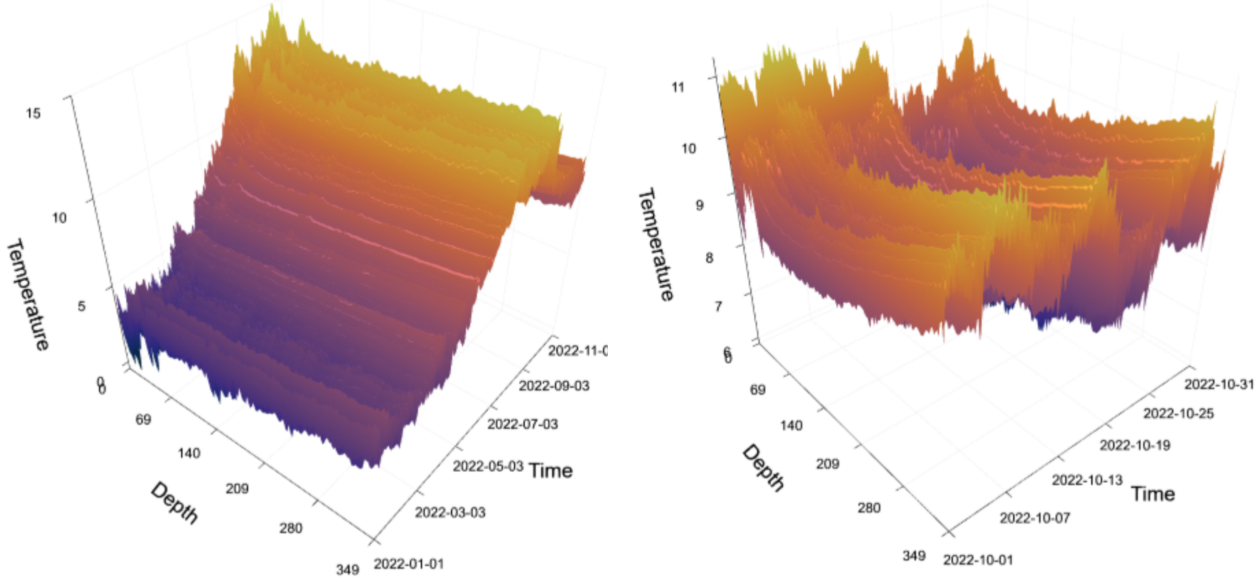
3.2.5 Plotting

The splines are the objects being plotted, with the advantage that they can be plotted for any chosen time or length value, even if they do not align with the original data points. For this reason, a time vector with more points than the original was created, to obtain a complete and uniform time span for plotting. A major challenge was that each year had several months with missing sensor data. To improve the interpretabilit of the plots, functions were implemented to identify the intervals where the sensors had correctly recorded data and return NaN for periods with missing data. This way, the resulting graphs avoid the display of spline approximations for months with missing data, replacing them with empty spaces, preventing confusion. The plots generated for each year and each month, for both channels, include surface plots, heatmaps, and contour plots. In the first case, temperature is represented in 3D as a function of space and time (with the x-axis for time, y-axis for space, and z-axis for temperature). Heatmaps work similarly, but instead of a 3D representation, temperature values are indicated using a colour gradient. Contour plots represent temperature using contour lines, where each line connects points of equal temperature, allowing for an intuitive visualization of temperature gradients over time and space. Below are some examples.



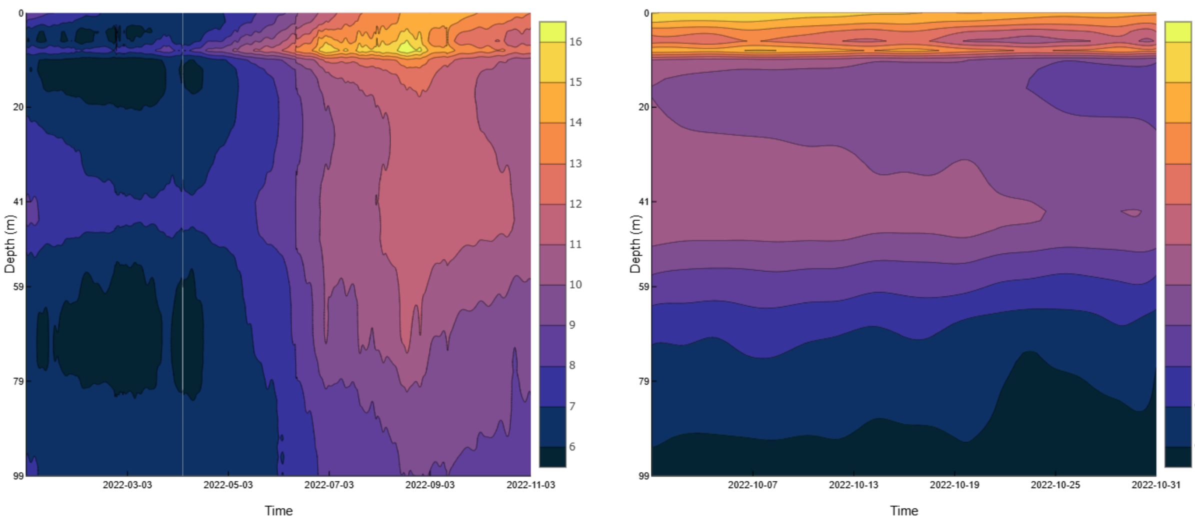
(a) 2022 BH1 **(b)** Oct 2022 BH1

Figure 18: Surface Plots of year 2022 and October (Oct) 2022 relative to Borehole 1 (BH1). The Temperature is in °C, while Depth values are in meters.



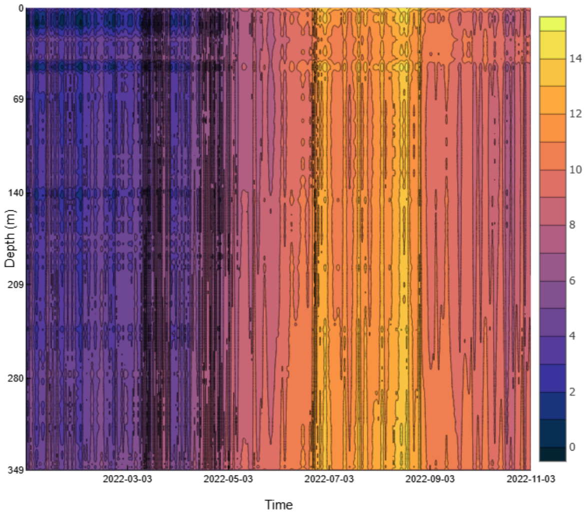
(a) 2022 BH10 **(b)** Oct 2022 BH10

Figure 19: Surface Plots of year 2022 and October (Oct) 2022 relative to Boreholes 10 (BH10). The Temperature is in °C, while Depth values are in meters.

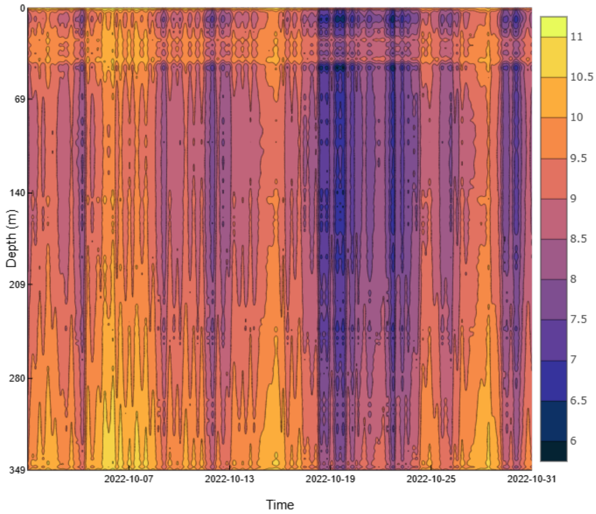


(a) 2022 BH1 **(b)** Oct 2022 BH1

Figure 20: Contour Plots of year 2022 and October (Oct) 2022 relative to Boreholes 1 (BH1). The temperature (°C) is represented by the color scale in the colorbar.

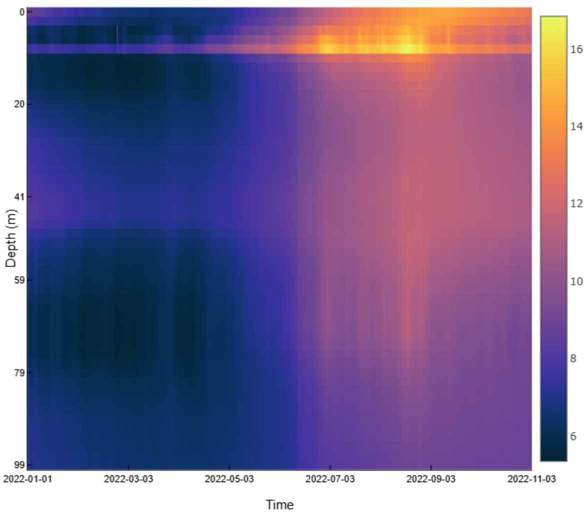


(a) 2022 BH10

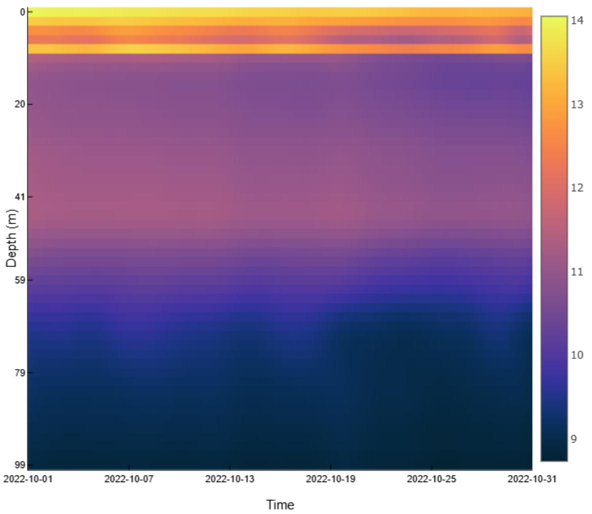


(b) Oct 2022 BH10

Figure 21: Contour Plots of year 2022 and October (Oct) 2022 relative to Boreholes 10 (BH10). The temperature (°C) is represented by the color scale in the colorbar.

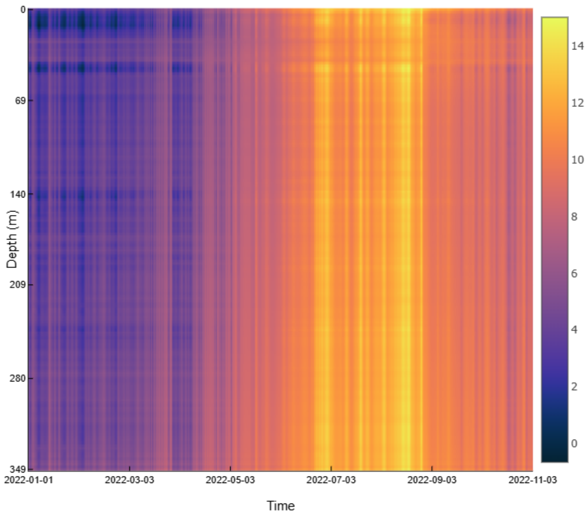


(a) 2022 BH1

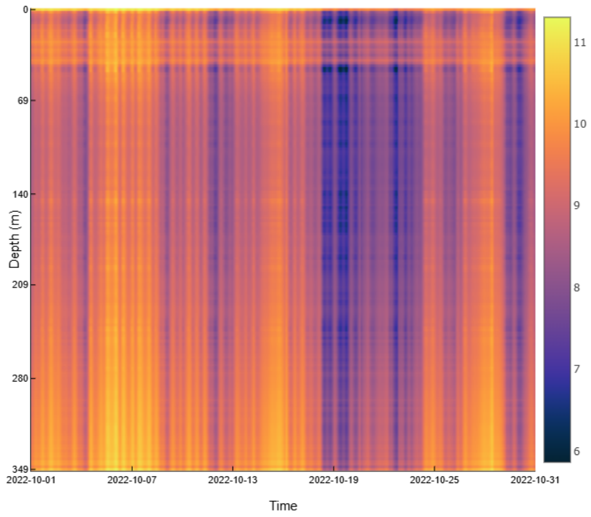


(b) Oct 2022 BH1

Figure 22: Heatmaps of year 2022 and October (Oct) 2022 relative to Boreholes 1 (BH1). The temperature (°C) is represented by the color scale in the colorbar.



(a) 2022 BH10



(b) Oct 2022 BH10

Figure 23: Heatmaps of year 2022 and October (Oct) 2022 relative to Boreholes 10 (BH10). The temperature (°C) is represented by the color scale in the colorbar.

These types of graphs are useful for intuitive visualization and qualitative considerations. General trends and initial observations helped identify valuable areas for deeper study and investigation. Further analysis, reported in Chapter 4, were carried on focusing on the identification of temperature patterns and general trends, along with heat transfer modeling, to examine how neighboring boreholes interact with one another.

3.3 Implementation and Tools

The programming language used to develop the code and data-handling algorithms was Julia. Julia is particularly well-suited for this purpose, with over 8,000 registered packages and applications in machine learning and numerical computing. Thanks to its just-in-time (JIT) compilation, Julia achieves execution speeds comparable to C++, eliminating the need for a dual-language approach where Python is used for ease of development and C++ for performance. The multiple dispatch system in Julia selects the most specific function based on input types, optimizing execution. It also supports multithreading and distributed computing, making it ideal for large-scale data processing. In conclusion, Julia is a flexible and dynamic language with a rich ecosystem and strong community support, making it a powerful tool for scientific computing and data-intensive applications [66], [67], [68].

An additional implemented tool was the PostgreSQL Database, an open source relational database management system. It supports different data types, functions, operators and indexing method for fast and easy access. The tables can be created, selected, exported and modified directly from the code with specific queries, while PGAdmin is a support application that provides a user friendly interface to achieve the same results [69], [70], [71].

3.4 Challenges and Solutions

One of the main challenges encountered was dealing with an outdated Windows system that lacked up-to-date protocols, making file and folder retrieval both complicated and slow. To address this, HFS (HTTP File Server) was installed, allowing dynamic sharing of selected folders. However, files could only be easily fetched if their full path, down to the exact file, was specified. Given that 11GB of files needed to be retrieved, manually accessing them one by one was not feasible. However, HFS still proved useful in scenarios where only the most recently produced file was needed, such as for real-time investigations or ongoing tests. For bulk transfers, USB sticks were used to copy and paste the folders generated by HaloDTS onto another device. From there, they were stored in a local directory to establish the necessary connection to the database. This process was extremely slow due to the system's limitations and patience was the only adopted approach.

Adding to the complexity, the file contents were diverse, some contained multiple sections mixed together, while others varied in format, making organization difficult. Arranging data and metadata in separate tables, sorting the data by channel, datetime, temperature and depth proved to be effective, since the time and channel acted as selection criteria while length and temperature were the main analysis objects.

An additional challenge consisted in determining the criteria to correctly identify the cable segment that measured borehole temperature, rather than picking up readings from elsewhere. Another issue arose when plotting data with missing values. Simply interpolating across gaps was not reliable, so different approximation and interpolation techniques were explored, identifying splines as the most flexible and suited one.

Distinguishing between actual temperature fluctuations and noise in the data was another challenge. Since detailed information about the heat pump's load and activation periods was lacking, pinpointing exact variations was tricky. Fortunately, for a qualitative analysis, extreme precision was not necessary. The general temperature trends remained intact with splines, even without perfect accuracy.

Chapter 4

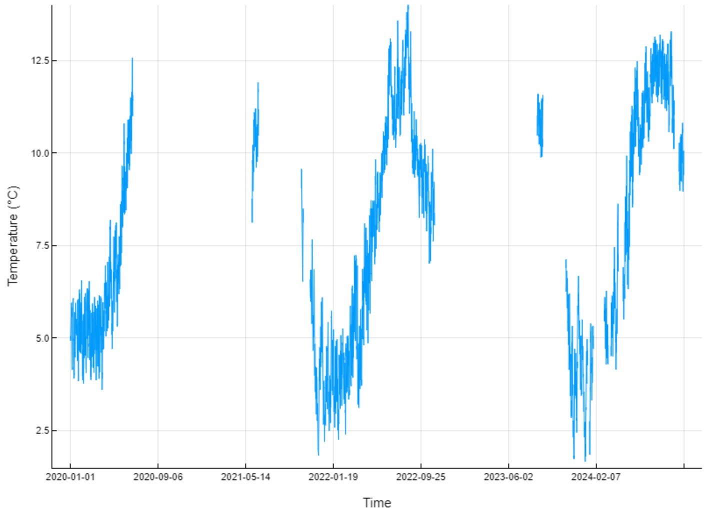
Data Analysis and Heat Transfer Modeling

4.1 Qualitative Analysis of Borehole Temperature Data

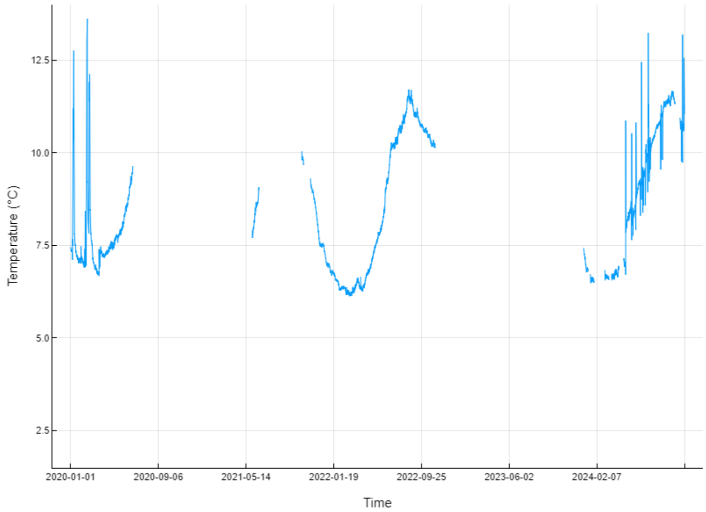
Due to the lack of important entities such as flow rate and loads values, only qualitative analsis is feasible. Nevertheless, it is possible to study various aspects related to soil temperature variations. All the results and plots produced in the following paragraphs, are obtained starting from the spline objects described in the previous chapter.

4.1.1 Temperature Patterns and Trends

One of the most interesting aspects lies in recognizing seasonal temperature trends. In this regard, the first observations can be made based on temperature over time graphs, considering, for a clear visualization, the average temperature over the depth.



(a) Average Temperature of BH10



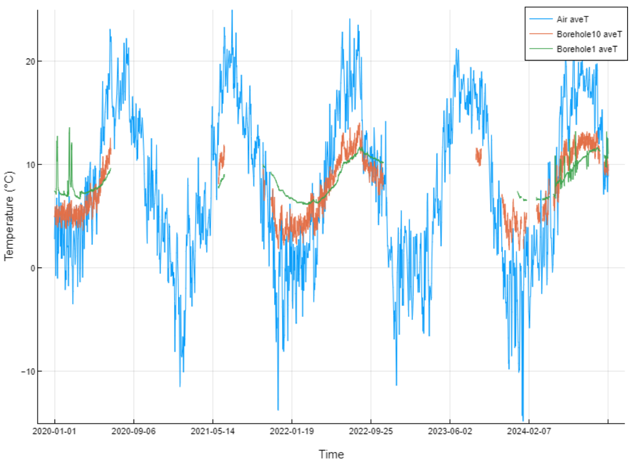
(b) Average Temperature of BH1

Figure 24: Average Boreholes Temperature from the 1st of January 2020 to the 14th of October 2024.

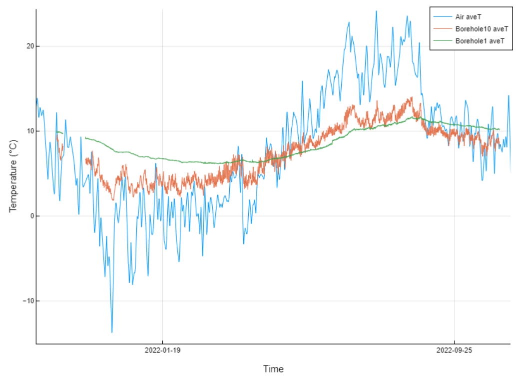
The graphs display large gaps corresponding to periods without data collection. Unfortunately, the missing measurements always refer to the same part of the year. However, it is possible to observe a sinusoidal pattern that repeats approximately annually (the longest uninterrupted period ranges from November 2021 to October 2022 and shows an almost complete temperature cycle).
Figure Figure 24b shows periods of unexpected oscillations at the beginning of 2020 and in 2024. These are likely due to the fact that the research heat pump connected to Borehole 1 was operated for temporary tests (in 2024, it is confirmed that these tests occurred). The annual trend, as expected, aligns with the annual alternation between summer and winter. In fact, the boreholes temperature decreases during the heating operation periods of connected heat pumps, which occurs throughout the Swedish winter, and subsequently, the soil temperature increases during the summer period. This pattern is more visible in the case of Borehole 10, which is effectively connected to heat pumps and therefore directly susceptible to variations in external temperature. Borehole 1, on the other hand, exhibits the same periodicity but less distinctly,

as its temperature decrease or increase is a reflection of the heat extracted or injected by the neighboring boreholes, not directly by the heat pumps.
For this reason, it is reasonable to expect a delay in temperature change between Borehole 1 and 10, as well as its compression.

To further investigate the relationship between ground temperature and external air temperature, air temperature data from Stockholm corresponding to the period of DTS data collection has been gathered from the Open-Meteo API, which integrates observations from weather stations, aircraft, buoys, radar, and satellites, along with mathematical models to estimate missing values [72]. The temperature values retrieved from the website represent the daily average temperature for each day. Cubic spline interpolation is applied to the three datasets (the outdoor temperature, and the two Boreholes' temperatures), with the smoothening parameter set to zero. The following graph shows the external temperature and the average temperature (averaged along the entire length of the boreholes) over time.



(a) From January 2020 to October 2024



(b) From November 2021 to October 2022

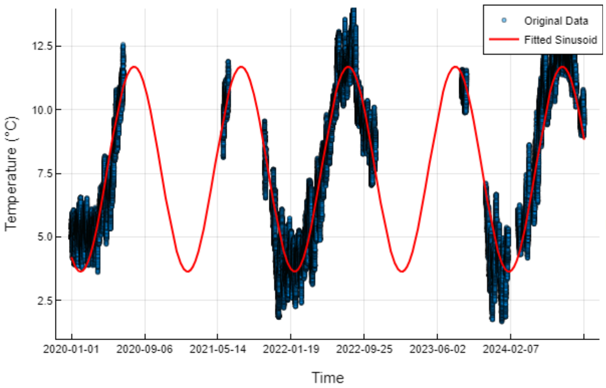
Figure 25: Average Boreholes Temperature and Air Temperature in Stockholm.

It is evident that when the external temperature drops below a certain value, the heat pumps are activated, extracting heat from the ground to transfer it to the buildings. Although Borehole 1 is not involved in heating or domestic hot water production, it quickly follows the trend of Borehole 10, but with a certain delay and to a lesser extent. This indicates that there is mutual interaction between the various components of the borehole field.

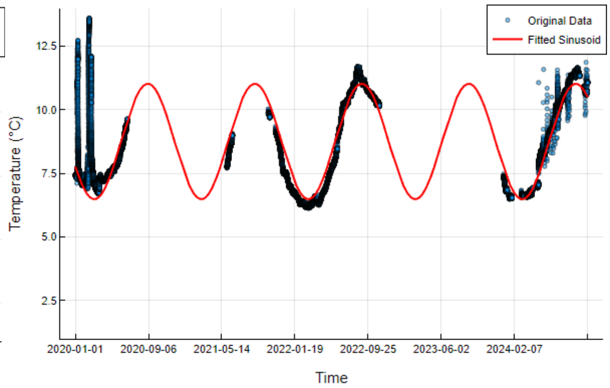
To quantify and better evaluate some properties of the temperature profiles, the data were approximated using simple sinusoidal functions, reflecting the dominant periodic behavior observed after removing daily and hourly fluctuations. It is reasonable to assume a dominant period of one year, a full seasonal cycle. The temperature profiles were modeled using the sine function

$$T(t) = A + B \cdot sin(C \cdot t + \phi) \tag{25}$$

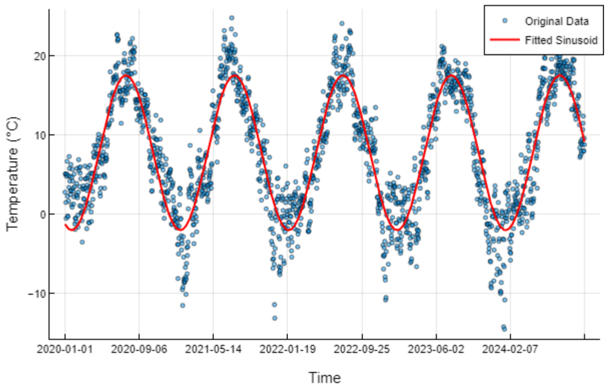
where T is the temperature, t is time, C is known and equal to one year period $C = \frac{2\pi}{365 \cdot 24 \cdot 3600} Hz$, while A, B and ϕ are obtained using the least squares method by fitting the sine function to the real data. A deeper, quantitative study regarding periodicity and phases of the temperature signals is reported in paragraph 4.2. The best-fit sinusoidal functions, resulting from this early-stage analysis, are shown in the following figures.



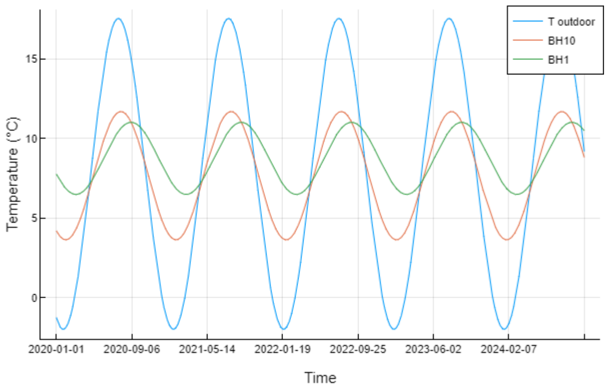
(a) Best Sin Fit to BH10 data



(b) Best Sin Fit to BH1 data



(c) Best Sin Fit to Outdoor data



(d) Best Sin Fits

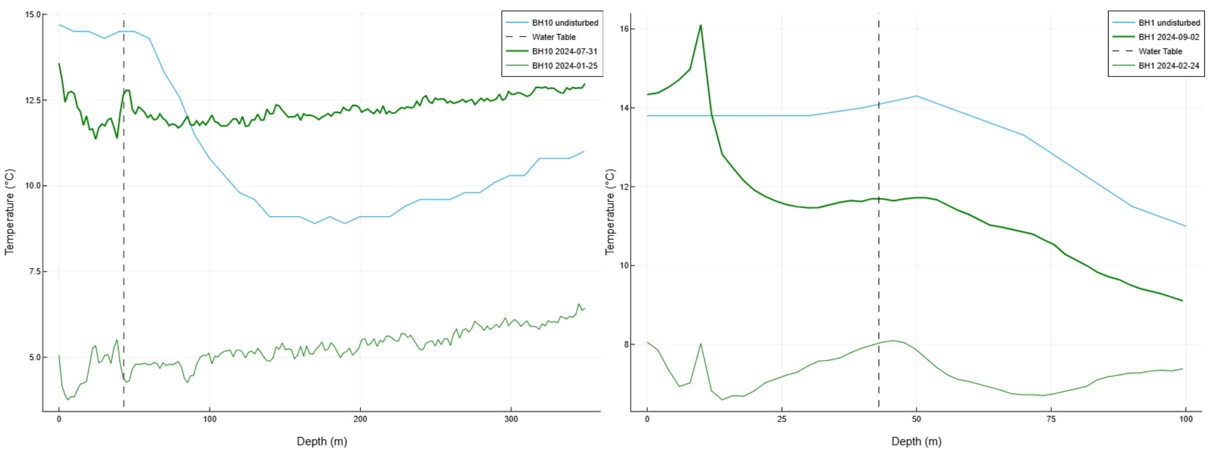
Figure 26: Best Sinusoidal Fits to Temperature Data.

The amplitude (B) of the various signals provides insights into the damping effect in Borehole 1 compared to Borehole 10. The sinusoidal function representing Borehole 1 has a total amplitude of approximately $2.245°C$, while Borehole 10 exhibits a larger amplitude of $4.035°C$. The external air temperature fluctuates with an amplitude of $9.77°C$. The soil proves to be a natural damper in the transfer of temperature shifts; although it follows the outdoor temperature variations, these shifts are always compressed and remain closer to the soil's mean temperature value. This effect is even more pronounced when considering the temperature of Borehole 1, which is further damped due to the heat transferred from adjacent soil zones, making its temperature variations more attenuated.

Regarding the phase shift (ϕ), the Outdoor temperature serves as leading signal, as expected.

Borehole 10 is delayed by only 8.5 days, while Borehole 1 phase shift relative to the outdoor temperature is 40.18 days.
The mean values (A) of the three fitted sinusoidal functions are very close to one another, respectively, $8.70°C$ for Borehole 1, $7.66°C$ for Borehole 10 and $7.78°C$ for the outdoor temperature. The small difference could be related to several factors. First of all, the length difference between the boreholes plays a role, as the shallower borehole is more sensitive to surface temperature fluctuations and the dry section constitutes almost half of its well. Secondly, the research borehole is the only one featuring a coaxial pipe structure, influencing the heat transfer. Additionally, at greater depths, the soil's thermal inertia is higher, meaning it resists temperature changes more effectively. This makes the temperature profile in the deeper borehole more stable and reflective of the external temperature, which explains the alignment with the air temperature. Moreover, the influence of the other boreholes might play a role that is not quantifiable, not to mention that this analysis is still an approximation.

The files related to the deviation measurements provided, in addition to the borehole inclination and their position every 10 meters, an associated temperature value. The measurements were taken 1–2 days after the drilling of the borehole field, on May 12, 2016. Therefore, we can consider this temperature value as the pre-operation baseline, representing the undisturbed ground temperature. Regarding the instrumentation, it is important to note that it was not intended, required, or well-suited for providing accurate temperature measurements, and no specification on that accuracy is provided. The sensors were in a pressure housing, delaying the outside temperature's effect. Depending on the surface-to-borehole temperature difference, stabilization could take 5–10 minutes, by which point, a large borehole section might already have been surveyed. Despite these uncertainties, the data could still provide interesting insights. The following plots enable the comparison between the recorded temperature values with those from the two boreholes at the time corresponding to the sine function peak in 2022, as evaluated previously, as well as at the minimum immediately preceding it.



(a) BH10 **(b)** BH1

Figure 27: Undisturbed Temperature and estimated minimum and maximum Temperature profiles.

Since Borehole 1 (*BH1*) is inactive, meaning it reflects the natural temperature heterogeneity of the ground, it provides a clearer signal. From Figure 27b, it can be inferred that the temperature is gradually decreasing, due to the heat extraction of the heat pumps from the borehole field.

4.1.2 Comparative Overview of Boreholes 1 and 10

The active borehole is influenced by heat pump operations, causing temperature oscillations that match the operational intervals of the heat pump. These fluctuations can be observed on a daily, if not hourly, scale. Conversely, the temperature of the inactive borehole remains much more stable, producing a series of significantly smoother plots (Figure 18, Figure 20, Figure 22).
To better illustrate the thermal interactions between boreholes within the field, animations were created to relate the temperature derivative over time (dT/dt) to its absolute temperature (T). In the first approach, the average temperature along the borehole length was considered. Each animation frame represents a specific point associated with the (*dT/dt*, *T*) pair, and a temporary trace of the point's trajectory is shown. In the second approach, the entire borehole length was analyzed, resulting in a curve instead of a single point. Selected frames from the animations are shown below, with the active borehole (Borehole 10) marked in red and the inactive borehole (Borehole 1) in blue.

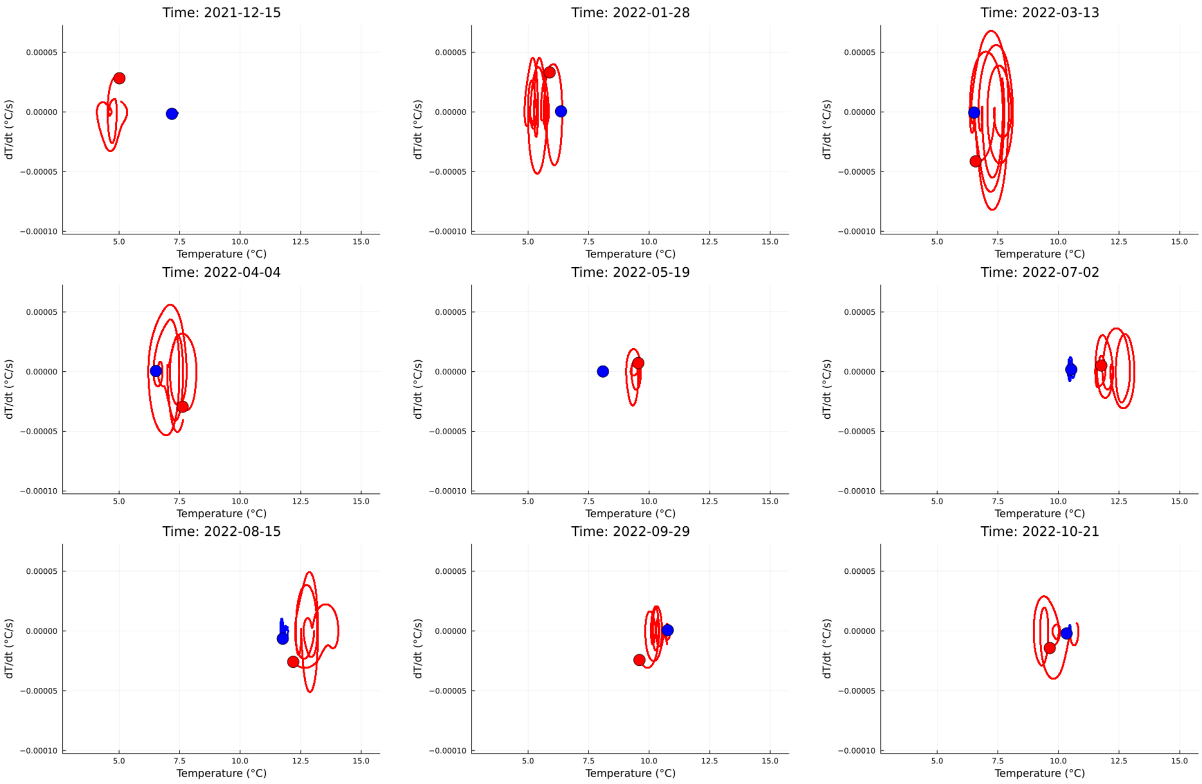


Figure 28: Time evolution of average Temperature.

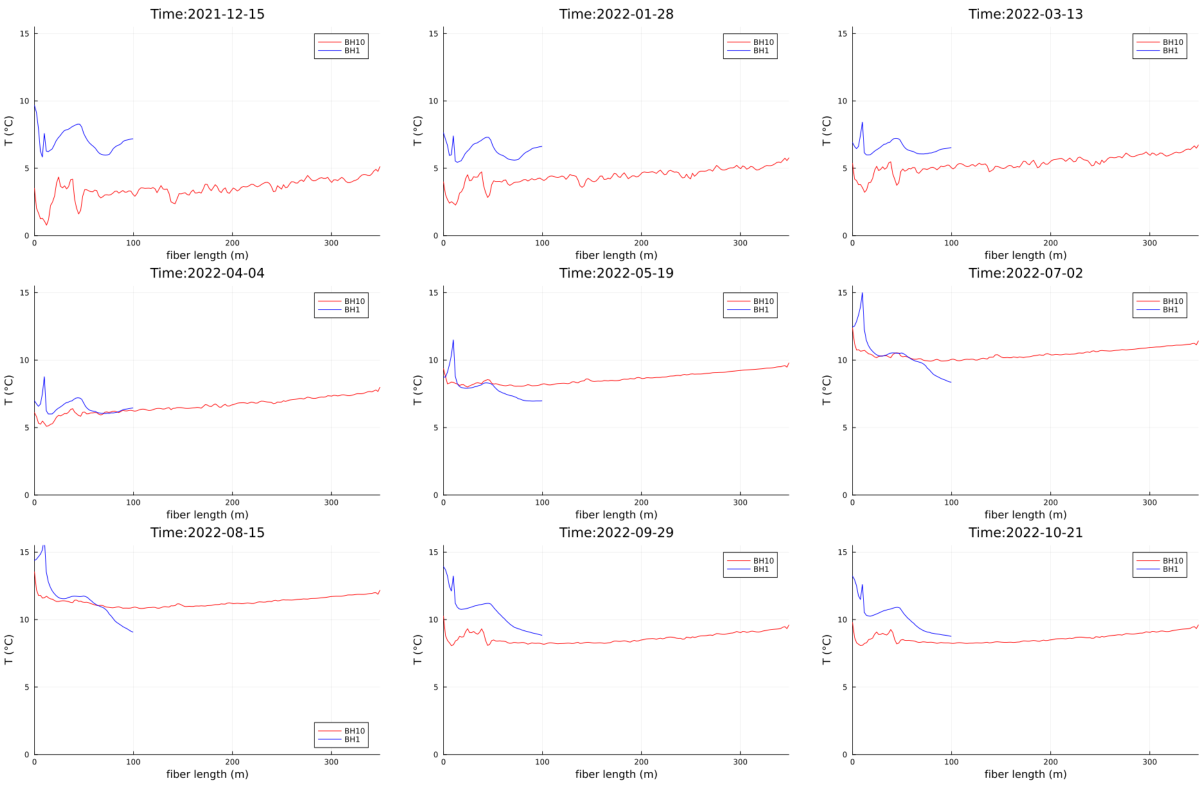


Figure 29: Time evolution of Temperature along the length.

The analysis reveals that the blue marker (Borehole 1) follows the red marker (Borehole 10) with a certain delay but moves more slowly. As a result, its trajectory is not clearly visible, showing minimal variations in *dT/dt*, especially when compared to the red marker in Figure 28. This suggests that the activity of Borehole 10 influences Borehole 1, but only in terms of long-term temperature trends. Short-term fluctuations, corresponding to brief operational cycles, are not transmitted.
Similar conclusions emerge from the full-length plot analysis presented in Figure 29.

4.1.3 Subsurface Variability and Water Table Effects

Temperature profiles over depth can provide valuable insights into lithological characteristics.

There is an unexpected temperature peak near the surface of Borehole 1, which is clearly visible in Figure 29, as well as in the contour plots and heatmaps from the previous chapter (Figure 20, Figure 22) at approximately 10 meters below the surface. The cause of this anomaly remains unidentified but is consistently present throughout the entire analysis period. It might be interesting to investigate the lithology at this depth, as the presence of a specific material or geological feature could explain this temperature anomaly.

A notable feature, especially evident in heatmaps and contour plots, is a distinct behavioral change around -45m, which can be reasonably attributed to the water table. It is important to mention that the water table level varies over time due to rainfall, snowmelt, and other atmospheric or geological events [73], [74]. However, for this analysis, we refer to documented data indicating the water table at -43m [22], [72].

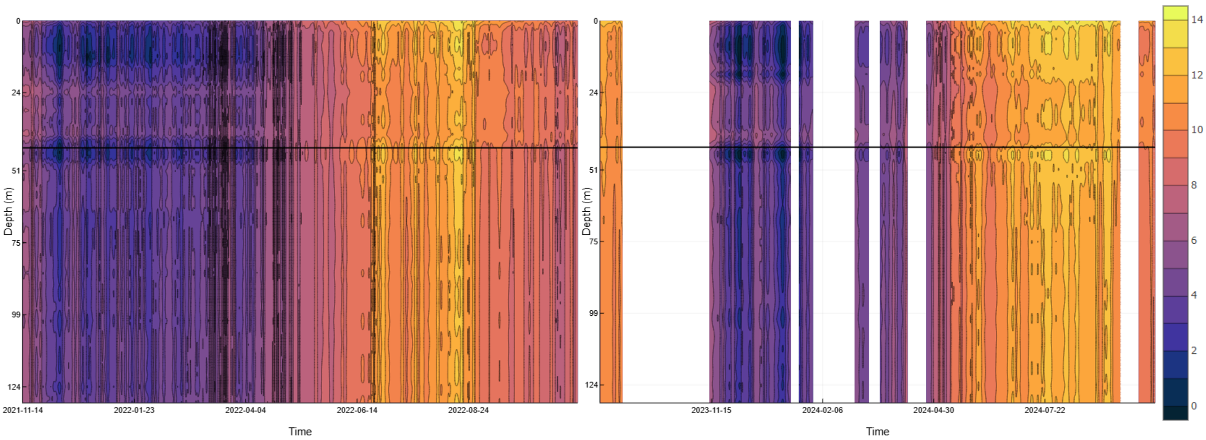


Figure 30: Contour plots of the upper part of Borehole 10 with water level Line. The temperature (°C) is represented by the color scale in the colorbar.

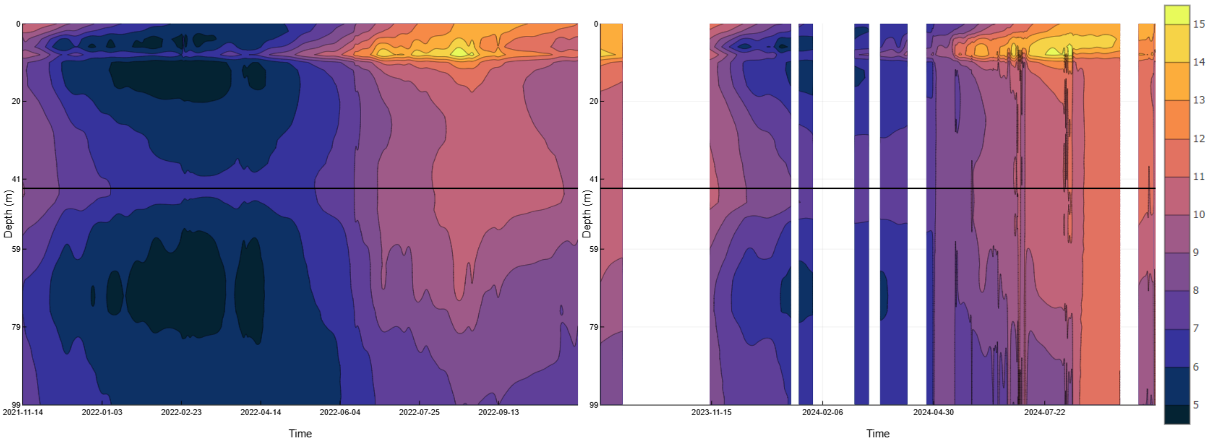


Figure 31: Contour plots of Borehole 1 with water level Line. The temperature (°C) is represented by the color scale in the colorbar.

The presented contour plots cover the period from November 2021 to October 2024, although they contain several gaps. A black line indicates the presumed water level. The plots provide greater insight in the case of inactive Borehole 1, where temperature variations decrease near the water table and never reach extreme values, indicating lower thermal conductivity of the ground. In contrast, for Borehole 10 (BH10), the temperature profile is much more uniform due to the action of pumps circulating water through the pipes. This circulation introduces advection as an additional heat transfer mechanism, resulting in a more homogeneous thermal distribution. As a consequence, the distinct behavior observed in BH1 is not present here. However, although heat transfer generally depends on depth, in this case, the sharp transition suggests with reasonable certainty that the observed thermal behavior is primarily caused by the presence of the water table. The dry section, which is grouted rather than filled with water, does not exhibit behavior significantly different from the wet section. Although the transition line between the sections is visible, the temperature profiles above and below it are relatively simmetrical. The only notable exception is the shallow layer just below the surface, which shows significant temperature fluctuations, with extremes of both heat and cold. This suggests the possibility of imperfect insulation from the ground surface.

Additionally, Borehole 1 exhibits an unusual, sharp temperature change above the expected water level, the cause of which is challenging to determine. Important considerations include the construction characteristics of the borehole, as it is an ungrouted coaxial borehole, as well as the positioning of the fiber optic cable, both of which could affect the recorded temperature profile.

Aside from the groundwater level, there are other depths where more abrupt changes are observed. Such observations may provide insights into the different geological layers and their varying characteristics.

The following plots refer to the same time span as Figure 30 and Figure 31. They are produced by subtracting the average temperature value over the borehole length from the measured temperature at each time step.

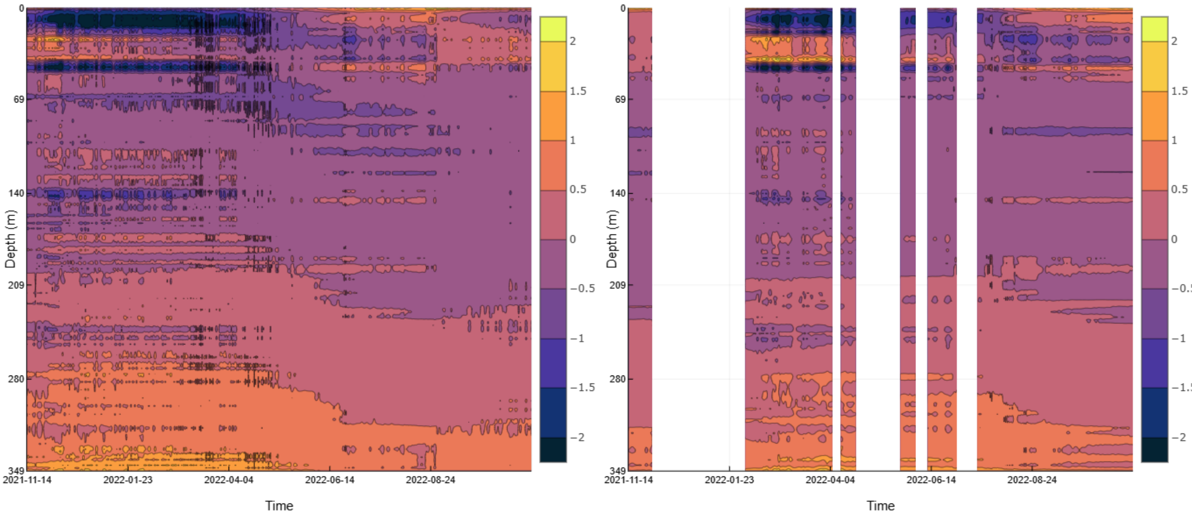


Figure 32: Contour plots of Borehole 10 Temperature Deviation from the Average Over Depth. The colorbar indicates $T - T_{ave}$.

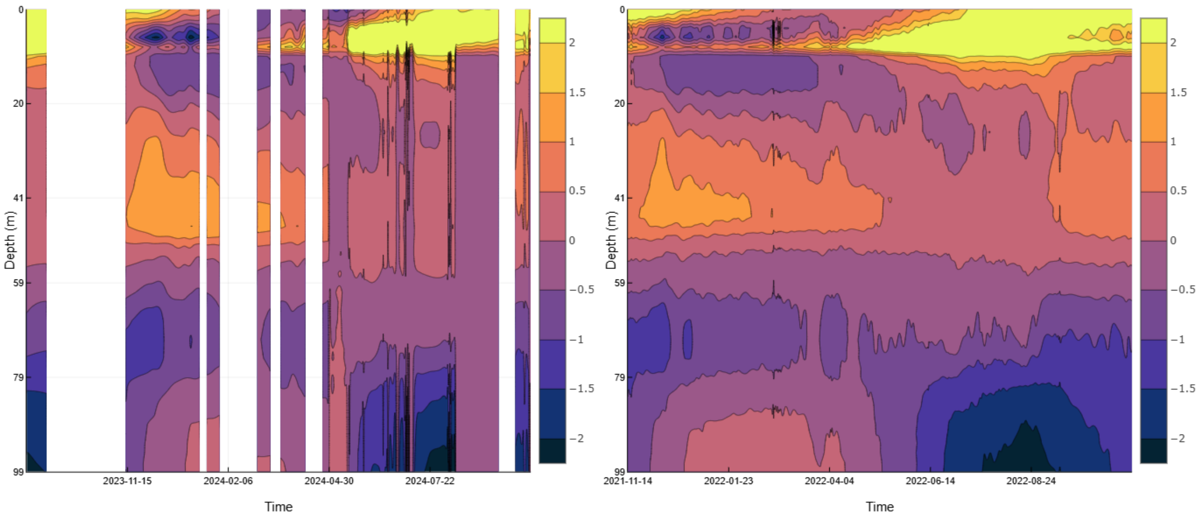


Figure 33: Contour plots of Borehole 1 Temperature Deviation from the Average Over Depth. The colorbar indicates $T - T_{ave}$.

The results show how much each area of the ground deviates from the average, and especially for how long. It is thus possible to evaluate recovery times, indicating that some zones in the soil recover more slowly than others. In the surface layer (0-50 depth) rapid temperature fluctuations due to external influences can be observed, with notable anomalies in early and mid-2022. The mid layer (100–250) shows moderate stability with persistent cool anomalies. In the deeper levels (250–349) prolonged warm anomalies from mid-2022, indicating slow thermal changes and heat retention.

Regarding thermal recovery, surface layers quickly return to average temperatures but are highly sensitive to external conditions. While deeper layers exhibit prolonged temperature deviations, suggesting slower heat dissipation due to higher thermal inertia or material differences. This may be due to differences in the lithology of the various layers at different depths, such as the presence of water below 43 m. Other influencing factors could be that in the more superficial part, insulation from external conditions (some sections are beneath buildings, others under an external cemented garden) might not be perfect, thus affected by external temperature, as well as the closer proximity between boreholes in the upper section.

4.2 Discrete Fourier Transform Analysis

4.2.1 Seasonal and Periodic Behavior

The frequency analysis performed using Fourier Transforms confirmed that seasonal cycles are the primary drivers of the periodic temperature variations observed in both boreholes and external air. The first peak, which corresponds to the lowest frequency and longest period, is the same across all three datasets, with a period of approximately 350 days, just under a year. This alignment reinforces the connection between external temperature variations and the ones of borehole 10, as well as the connection between the two boreholes' temperature. For simplicity and lack of specific data, Borehole 10 is assumed as representative of the other 10 active boreholes in the field, neglecting the length difference and assuming the loads are equally distributed. Borehole 1, influenced by 11 similar neighboring boreholes, exhibits the same seasonal pattern as Borehole 10. This suggests that low-frequency oscillations are transmitted effectively from nearby boreholes.

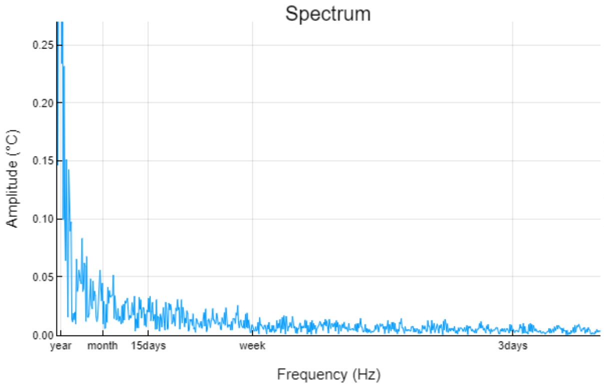
At higher frequencies, the correlation between Borehole 10 and external temperature remains strong, with a nearly perfect parallel trend. However, Borehole 1 does not show similarly intense peaks at high frequencies, essentially, the 350-days period is the shortest among the dominant ones. When analyzing frequencies corresponding to daily periods, with a tolerance of 3 hours, a significant difference in amplitude is observed between the two boreholes. Specifically, the amplitude, determined by averaging the amplitudes of frequencies corresponding to periods of 24 ± 3 hours, for Borehole 10 is 0.010213 °C, whereas for Borehole 1, it is only 0.000787 °C, representing a two orders of magnitude difference. This stark contrast highlights a stronger thermal response in BH10, which is likely due to its direct interaction with the heat source, while BH1 exhibits a more attenuated response. This indicates that the soil acts as a low-pass filter, allowing long-term (annual) temperature variations to propagate while attenuating short-term fluctuations. This conclusion aligns well with temperature plots, which show significantly more frequent oscillations in Borehole 10, whereas Borehole 1 reveals a much smoother temperature profile.

Phase values provide insight into the time lag between different signals, allowing for a comparison with the qualitative results discussed in the previous section. The phase relative to the common dominant frequency of 350 days is equal to 1.83654 rad for Borehole 10, 1.33391 rad for Borehole 1 and 2.03643 rad for the outdoor temperature, meaning that the boreholes temperature profiles are delayed. The phase difference between the active borehole and the external temperature is $\frac{(2.03643-1.83654)\cdot 349.599}{2\pi}$ days $\approx$ 11 days, slightly more than 8.5 days as previously estimated; while Borehole 1 is shifted of $\frac{(2.03643-1.33391)\cdot 349.599}{2\pi}$ days $\approx$ 39 days in respect to outdoor air, closely aligned with the estimated value.

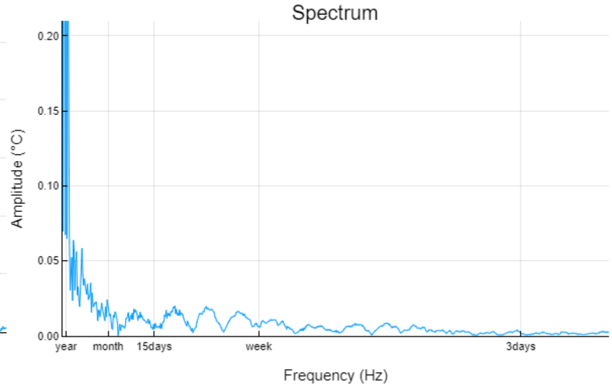
The fact that the dominant period is not exactly a year but few days shorter is unexpected, since seasons follow an yearly cycle. Although the borehole temperature data's frequency spectrum could be distorted by the linear interpolation applied to fill in gaps, the periodicity is also confirmed for external temperature data. In this regard, several aspects need to be considered: slow and gradual variations in temperatures could influence the main periodicity, and especially anomalies or climatic events (such as particularly warm winters or shorter summers) may slightly alter the annual cycle. Calendar variations such as leap years also play a role. Moreover, the sampling frequency (daily) can introduce aliasing that shifts the dominant component. Most importantly, with 4 years of data, the spectral resolution might not be perfect due to the limited time span, leading to an imperfect estimation of the dominant frequency. The analyzed data confirms the primary influence of the seasonal periodicity, however, a 365-day period remains the most reasonable one.

4.2.2 Frequency Domain Insights

The Discrete Fourier Transform was applied to the complete dataset for both boreholes, as well as for external temperature data in Stockholm referring to the same time span. The Fourier Transform is a mathematical tool, widely used in signal processing, that expresses any complex signal as a sum of sinusoids and decomposes any time-domain signal into its frequency components [75], [76]. The result is a spectrum that shows the amplitude of each frequency component, helping to identify dominant periodicities and periodic patterns within the data. Julia's FFT (Fast Fourier Transform) package, built on top of FFTW (Fastest Fourier Transform in the West), was used to apply the Discrete Fourier Transform to the temporal variations in temperature (again averaged over depth for the boreholes) and to extract the frequency components, their amplitudes and phases [77]. In this case, the average over each length value was computed from the matrix of values filtered with Kalman filter. The spline approximation was introduced on the resulting vector, however, in correspondence with the intervals of missing values, the interpolation proved to be very inaccurate, both in absolute terms and in terms of trend. Therefore, a simple linear interpolation was adopted for the Fourier analysis to prevent stronger approximations from distorting the frequency spectrum, although this method still carries the risk of altering the frequency spectrum. Moreover, the same analysis was conducted by limiting it to the interval between November 2021 and October 2022 (the longest uninterrupted period), and, as expected, all the lowest frequencies did not appear. However, the main frequency of 350 days was close to the one obtained in this case, which was around 360 days. The difference could be due to the fact that the considered interval was still shorter than a full period and not sufficient to reveal precise frequencies. Meanwhile, the phase shift between the various signals was almost confirmed. The following plots display the amplitudes of the three signals' frequency components.

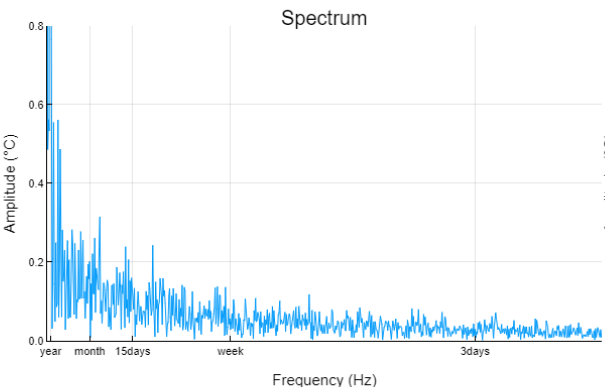


(a) Frequency Spectrum of BH10 Temperature signal

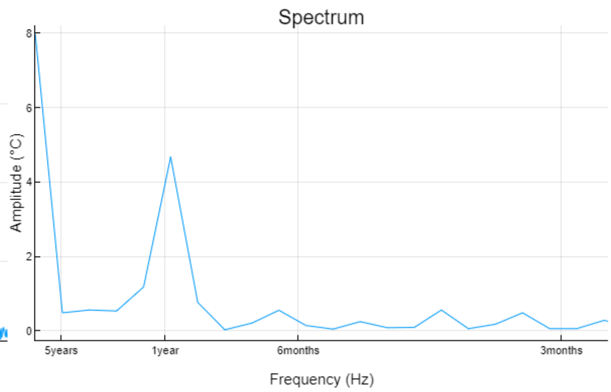


(b) Frequency Spectrum of BH1 Temperature signal

Figure 34: Frequency Spectrum of Boreholes Temperature signals



(a) Full Frequency Spectrum of Outdoor Temperature signal



(b) Low Frequency Spectrum of Outdoor Temperature signal

Figure 35: Frequency Spectrum of Outdoor Temperature signals

Once the initial peak, related to an infinite period, was excluded and only the positive frequencies (as the negative ones are symmetrical) were considered, the amplitudes, frequencies, and phases of the first four peaks in terms of amplitude were evaluated and are shown in the following table.

Source	Peak Frequency (Hz)	Peak Phase (rad)	Peak Amplitude (°C)	Peak Period (sec)	Peak Period (days)
Air	3.30877e-8	2.04601	4.67674	3.02227e7	349.8
Air	2.64702e-8	-0.738257	1.18075	3.77784e7	437.25
Air	3.97052e-8	1.89249	0.76593	2.51856e7	291.5
Air	1.32351e-8	-0.608862	0.56193	7.55568e7	874.5
BH10	3.31068e-8	1.83654	1.19432	3.02053e7	349.599
BH10	1.32427e-8	-2.13009	0.75692	7.55133e7	873.997
BH10	1.98641e-8	1.03478	0.46874	5.03422e7	582.664
BH10	4.63495e-8	2.3331	0.41415	2.15752e7	249.713

Source	Peak Frequency (Hz)	Peak Phase (rad)	Peak Amplitude (°C)	Peak Period (sec)	Peak Period (days)
BH1	3.31068e-8	1.33391	0.71230	3.02053e7	349.599
BH1	1.32427e-8	-2.02322	0.46701	7.55133e7	873.997
BH1	1.98641e-8	0.306362	0.39331	5.03422e7	582.664
BH1	6.62135e-9	2.0009	0.33380	1.51027e8	1747.99

The dominant period, which is the strongest among the three datasets, is common to all three and is approximately 350 days. The other periodicities, consistently with harmonic patterns exhibit a degree of interdependence: 874 days is twice 437 days, and 437 days corresponds to $291 \times \frac{3}{2}$ days. Overall, these secondary periodicities are associated with nearby frequencies, suggesting the possible presence of frequency leakage. This phenomenon can cause the primary period to produce prominent neighboring peaks that are not genuinely distinct periodic components.

4.3 Borehole Heat Transfer Modeling

4.3.1 Borehole Heat Transfer Model

In order to verify and validate the results obtained, a physiscal model was developed to evaluate the temperature signal at a certain distance from a heat source. This analysis involves a simple infinite solid with homogeneous properties, and a heat source modeled as a point source. In this case, heat transfer is governed by the pure conduction equation, also known as the heat diffusion equation:

$$\frac{\partial T}{\partial t} = \alpha \nabla^2 T \tag{26}$$

where $\alpha = \frac{k}{\rho c_p}$ is the thermal diffusivity, dependent on the material properties: thermal conductivity k, density ρ and specific heat capacity c_p. Documentation provides values for design ground thermal conductivity $k = 3.4$W/(mK) and design volumetric heat capacity $\rho c_p =$ 2.2MJ/(kgK), while the density was set to the typical value $\rho = 800.0 kg/m^3$ [22].

For a point source in an infinite medium, the solution to the heat equation is provided by the Green's function, which represents how the temperature varies over time in response to the impulse. For a unit heat pulse at $t = 0$, the temperature change $T(r, t)$ at distance r is given by:

$$T(r, t) = \frac{1}{(4\pi\alpha t)^{3/2}\rho c_p} \exp(-\frac{r^2}{4\alpha t}) \tag{27}$$

The goal is to model the heat injection or extraction as a sinusoidal function. The representation is chosen to match the daily load profile of a heat pump, with the period set to 24 hours: $q(t) = \sin \frac{2\pi}{24 \cdot 3600} t$, or the seasonal variation, with a second simulation where

$q(t) = \sin \frac{2}{\pi} 365 \cdot 24 \cdot 3600t$.
The applied heat load is not a single impulse but it varies with time. By the principle of superposition, applicable since the heat diffusion equation is linear, the temperature response at any point in the system results from the cumulative effect of all past heat inputs.
Mathematically, this is expressed as the convolution of the heat load with the impulse response.

In the developed model, $q(t)$ is approximated by a piecewise constant function, which places the system in the discrete domain. In this context, the corresponding expression for the impulse response is the differential of the step function T_{step}.
The step function represents the temperature response of a unit step load, and is obtained integrating the Green's function over time.

$$T_{step}(r,t) = \int_0^t \frac{1}{(4\pi\alpha\tau)^{3/2}\rho c_p} \exp(-\frac{r^2}{4\alpha\tau})\, d\tau \tag{28}$$

This integral can be solved by recognizing the complementary error function, defined as:

$$\text{erfc}(t') = \frac{2}{\sqrt{\pi}} \int_{t'}^{\infty} \exp(-u^2)\, du \tag{29}$$

By replacing $u = \frac{r}{\sqrt{4\alpha\tau}}$ and $t' = \frac{r}{\sqrt{4\alpha t}}$, the function limit $t' \to \infty$ becomes $t = 0$ and the equation is expressed as:

$$\text{erfc}(\frac{r}{\sqrt{4\alpha t}}) = \frac{2}{\sqrt{\pi}} \int_0^t \frac{2r\alpha}{(4\alpha\tau)^{3/2}} \exp(\frac{r}{\sqrt{4\alpha\tau}})\, d\tau \tag{30}$$

Thus, the solution for the temperature distribution when a unit step heat load is applied can be derived directly from the complementary error function.

$$T_{step}(r,t) = \frac{1}{4r\pi k} \text{erfc}(\frac{r}{\sqrt{4\alpha t}}) \tag{31}$$

Finally, the temperature variation over time can be expressed as:

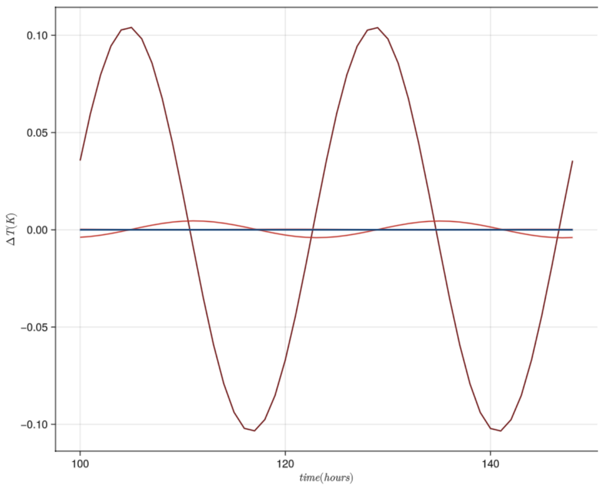
$$T[n,r] = q[n] * (T_{step}[n+1,r] - T_{step}[n,r]) \tag{32}$$

The difference operator for discrete convolution D, is defined as $DT[n] = T[n+1] - T[n]$, leading to the final temperature expression:

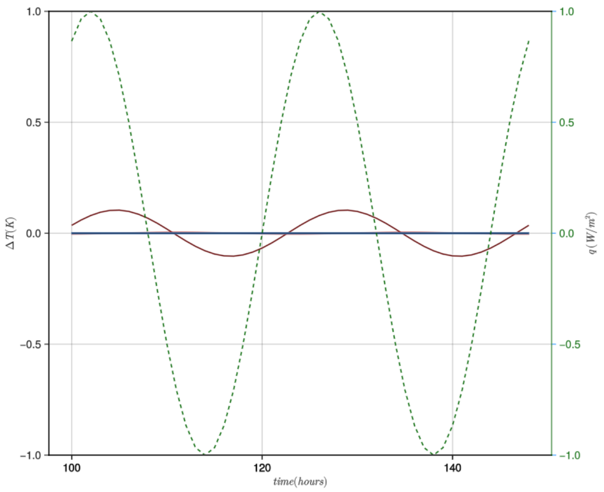
$$T[t,r] = q[t] * DT_{step}[r] \tag{33}$$

4.3.2 Distance-Based Temperature Attenuation

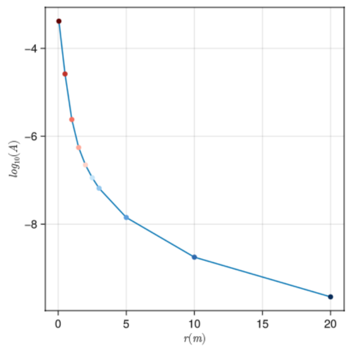
The results from the previously described simulations are presented for distances of $[0.05, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 5.0, 10.0, 20.0]$meters in the following graphs.



(a) Temperature variations at various distances from the heat source

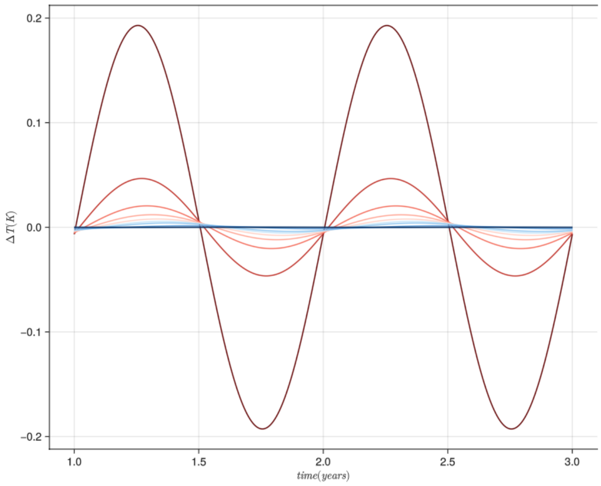


(b) Temperature variations and load over time

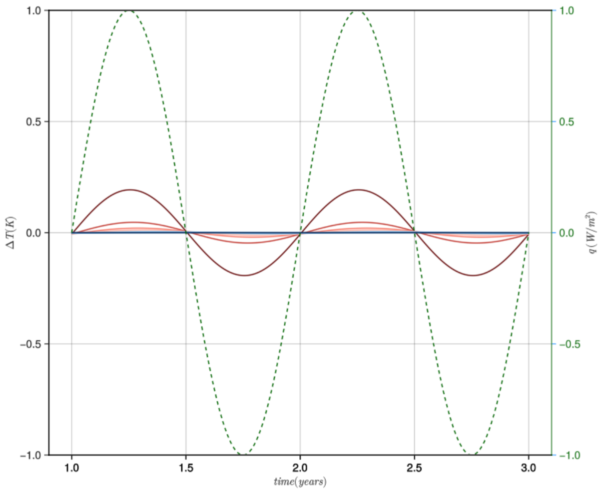


(c) Logarithmic scale of the Amplitudes (A) of the temperature variations at different radial distances

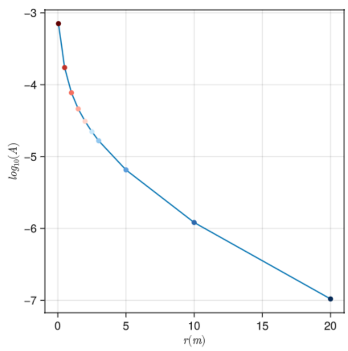
Figure 36: Temperature Profiles for 24-hour periodic sinusoidal heat load



(a) Temperature variations at various distances from the heat source



(b) Temperature variations and load over time



(c) Logarithmic scale of the Aamplitudes (A) of the temperature variations at different radial distances

Figure 37: Temperature Profiles for 1-year periodic sinusoidal heat load

Figure 36 corresponds to the temperature response of a daily sinusoidal load application, while Figure 37 represents the response of an yearly sinusoidal load. In Figure 36b and Figure 37b, the applied load is plotted alongside the temperature variation at various distances, represented by different colors. It is evident that only two temperature signals are distinguishable, the one at 0.05m and 0.5m, while the others have significantly lower amplitudes, making them difficult to observe.
In the long-term analysis, more distances become visible, as the damping effect is reduced, but it remains a dominant factor in heat transfer attenuation.
Figure 36c and Figure 37c display the amplitudes of the temperature signals at each considered distance, on a logarithmic scale, mantaining the same color scheme as in the temperature and heat load plots. As expected, the amplitude decreases rapidly with distance. This highlights the strong attenuation of heat transfer through the soil.

In the real temperature data provided by the DTS, the correlation between active and inactive boreholes remains quite strong, particularly for input signals with long periods. It is important to note that the damping effect, along with the phase shift (introducing a delay), is clearly present and visible. However, it does not reduce the temperature amplitude to near zero: Borehole 10

and Borehole 1 still show comparable values.
This can likely be attributed to the fact that real soil is neither perfectly homogeneous nor ideal, as assumed in the model. Additionally, near the surface, the boreholes are relatively close to each other, and there is more than just one point source of heat, leading to thermal interference. A distance distribution analysis reveals that the closest borehole to Borehole 1 is Borehole 5, with a separation of only 1.63m at a depth of 4m. In comparison, the shortest distance to Borehole 10 occurs at 66m depth, where the distance is 2.32m away.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis presents a reproducible method for processing data from Sensornet DTS systems. The approach enables the transformation of raw text file data into interpolated and smoothed data matrices using a PostgreSQL database, JuliaLang, and Python.

Distributed temperature sensing systems (DTS) are based on the Raman scattering effect. They include optical fibers that send light signals at various lengths, and detectors receiving the scattered ligths from which the temperature is derived.
HaloDTS, the software that runs the temperature distributed sensing systems, based on optical fibers and raman scattering methos, processes each temperature measurement and stores it in a text file. A new file is produced with every measurement, resulting in unstructured, noisy and often not directly meaningful dataset.
For the data to be analyzed effectively, it is useful to organize it, specifically, by datetime and the depth of the fiber corresponding to each temperature reading. Typically, handling large amounts of data is necessary, making database storage crucial. In this case, all the files produced from the beginning of 2020 to October 2024 were uploaded to PostegrSQL database, as three different tables: one for the metadata section, containing reference information; one for the actual temperature data; and one for configuration files with calibration parameters.
The recorded temperature values must be filtered to properly capture the portion of the fiber within the object of interest. This thesis presents a method to identify that meaningful section, evaluate temperature inaccuracies, and offer a solution for smoothing the results when precise measurements are not required, but rather a qualitative observation. The application of Kalman Filtering techniques provides a method for noise reduction, while spline fitting allows an additional, controlled smoothing.

Additionally, temperature data collected from 2020 to 2024 relative to two boreholes at KTH campus in Stockholm, despite time gaps in the dataset, provided valuable insights into temperature patterns, periodicity, and subsurface variability.

The two monitored boreholes are part of the same borehole field, consisting of 11 closely spaced boreholes near the surface, which diverge at greater depths. Since one of the two was not connected to any active heating system, unlike all the others, it was possible to study the interaction between undisturbed ground (represented by this inactive borehole) and the borehole field. This setup is uncommon, as the ground is rarely equipped with deep distributed sensors.
The datasets from the two monitored boreholes exhibit a sinusoidal seasonal pattern. The active borehole shows much faster temperature variations, but its seasonal profile remains clearly recognizable, as for the inactive Borehole. Comparing this profiles to the daily average outdoor temperature in Stockholm provided a visual representation of their correlation. The heat pumps activate when temperature reduces, typically in Autumn, causing the borehole temperature to decrease with a certain delay.
For this reason, these three raw temperature datasets were fitted to sinusoidal functions with a one-year period to estimate amplitudes and phase shifts.
To better quantify the correlation between the two boreholes, and between the boreholes and

the outdoor temperature, a Discrete Fourier Transform analysis was performed.
It revealed that the dominant frequency is identical in both boreholes and in the external air, confirming that outdoor temperature is the primary driving signal, although indirectly, as heat pump operations mediate its effect.
There is a delay of slightly more than a week between outdoor temperature variations and subsurface temperature fluctuations. However, the soil acts as a thermal buffer, reducing the amplitude of these fluctuations by approximately 50%. These results align with the sinusoidal functions mentioned earlier.
The inactive borehole can be considered representative of undisturbed ground. In this case, the temperature response exhibits a delay of over a month, with a more pronounced damping effect, yet it still follows the seasonal temperature trend. The soil prevents the daily and hourly temperature fluctuations observed in active boreholes from propagating fully to the inactive borehole, despite their proximity.
Various plots, such as contour plots and heatmaps, were produced, revealing a sharp local change in heat transfer dynamics at the water table level. However, this did not significantly affect the overall heat transfer, only locally, where the soil type changes. Moreover, the soil is inherently heterogeneous, as evidenced by the non-uniform temperature variations across different depths. Some parts recover faster than others, though the specific reasons for this behavior remain to be identified.
Finally, a numerical model was developed to investigate how ground temperature responds to a point heat source. The model was simplified by assuming an idealized, homogeneous, and infinite soil medium. The results indicate that for daily temperature variations, the thermal response remains significant only within a 1-meter radius. For annual sinusoidal heat loads, the temperature effect extends further but remains notable only up to a distance of approximately 2 meters. This closely aligns with the real data, where only long-term trends are transmitted over to the inactive borehole and dampened.

5.2 Future Research Directions

The thesis describes a data processing framework, offering a method to easily store and manipulate data produced by similar systems. This is reproducible and can be adopted in the future for similar setup.
Specifically, the historical temperature data can be compared with current measurements when the borehole is active, serving as a reference for analysis.

Having information on the thermal load and mass flow rate would enable the direct correlation of temperature variations with the amount of heat excahnged, bypassing the outdoor temperature, which adds a level of abstraction. In general, equipping the research borehole with a comprehensive sensor set, including a flow meter and a power meter, would provide a more detailed perspective and enable deeper analyses, including comparative studies of U-tube and coaxial layouts.

Moreover, extending the data collection period beyond five years would allow for a more robust evaluation of long-term temperature trends and help determine whether there is a gradual decline in subsurface temperature.

The integration of machine learning techniques, based on collected data, would constitute a powerful instrument to develop predictive models specific for that installation, based on real data.

Ringraziamenti / *Acknowledgments*

Sono parecchie le persone senza le quali avrei combinato ben poco.

Primi fra tutti, mamma e papà. Vi ringrazio perchè non solo avete reso ogni evento o esperienza concretamente possibile, ma soprattutto perchè mi avete fornito tutti gli strumenti perchè fossi io a scegliere le esperienze a cui dedicarmi. E vi siete fidati. Dopo avermi trasmesso i vostri valori, mi avete concesso mille viaggi per capirne l'importanza, e farli miei. La cosa più generosa e magica che avete fatto però, è stata sfornarmi una sorella, la C.

Abbiamo giocato, imparato, riso, pianto, litigato, odiato e amato tanto insieme. Per questo sei da sempre la persona a cui tengo più al mondo. Poi ognuna ha reso unico il proprio percorso, ci siamo allontanate, avvicinate, intrecciate e da C sei diventata anche Clara, e ora ho l'immensa fortuna che la persona a cui tengo di più è anche la persona di cui ho maggiore stima. Non mi stancherò mai di imparare da te, che mi hai già dato tanto.

Grazie ai nonni, perchè mi riportate sempre alle basi, mi avete fatto conoscer la terra, il sapore e il valore delle cose semplici.

Grazie alla mia cagnolina, la Perli, che non lo sa quanto rallegra iniziare la giornata col suo scodinzolio, eppure ogni mattina mi regala il suo sorrisetto e il ritmico battere della coda sul divano.

Grazie Fre, per esserti assunta il noioso ruolo di mia personale coscienza, grazie del supporto, del conforto, delle risate, della pazienza, della tua lucidità e della tua confusione, grazie di essermi amica. Sei un pilastro della mia vita e se ci penso mi sento gigante, mi fai essere me stessa a potenza.

Grazie Virgi, da cui ho preso ispirazione così tante volte che mi sento che ci siamo plasmate insieme, ed è un vanto sapere che la mia personalità ti assomiglia.

Grazie Gio, che non dimentichi mai di concedermi un sorriso, mentre mi ricordi che basta una colazione in hotel per sentirsi a posto con l'universo.

Grazie a Claire e a Sophi, che in poco tempo siete diventate compagne fidate, rifugio sicuro.

Grazie ad Amanda, Emma ed Eleonora che hanno reso un semestre di studi il più ricco e dolce di sempre.

Grazie alle Montagneros, un branco di donne dalla personalità scoppiettante e invidiabile, ispirate qualsiasi sentimento nello spettro dei sentimenti, e forse qualcuno ancora da definire, avrò grande cura delle memorie condivise.

Grazie a Rizz, Jaco, Cate, Chiara, Sanci, Agni, Marghe, Ceci, Manu, Pincho, Rosa e Ale, che bello scrivere i vostri nomi e avere la certezza che se volessi mai preparare una tesi tra 25 anni, tornerei a scriverli.

Grazie a Claudia, la mia relatrice, per la sua disponibilità e flessibilità, che mi hanno permesso di intraprendere un percorso di tesi all'estero.

Grazie ad Adrien e ad Alberto, che mi hanno accompagnato in questa fase finale del mio percorso, guidandomi con competenza, ma sempre con spirito amico.

Grazie ai miei compagni di merende e di pranzi in Dipartimento al KTH, per la leggerezza e la genitlezza con cui mi avete accolto.

Insomma, concludo quello che è diventato un inno all'amicizia: nonostante abbia scritto io la tesi, vi ringrazio tutti per averle dato un senso.

There are many people without whom I would have accomplished very little.
First and foremost, Mom and Dad. Thank you not only for making every event and experience practically possible but, above all, for giving me all the tools to choose for myself the experiences I wanted to dedicate myself to. And for trusting me. After passing on your values, you allowed me countless journeys to understand their importance and make them my own. But the most generous and magical thing you did was bringing into the world my sister, C.
We have played, learned, laughed, cried, fought, hated, and loved so much together. That is why you have always been the person I care about the most in the world. Then each of us carved out our own unique path, we drifted apart, came back together, intertwined, and from C, you also became Clara. Now, I have the immense fortune that the person I care about the most is also the one I admire the most. I will never tire of learning from you, for you have already given me so much.
Thank you to my grandparents, who always bring me back to my roots, who have taught me to know the land, the taste, and the value of simple things.
Thank you to my little dog, Perli, who has no idea how much joy it brings me to start the day with her wagging tail, yet every morning, she gifts me her little smile and the rhythmic thump of her tail against the couch.
Thank you, Fre, for taking on the dull role of my personal conscience, for your support, comfort, laughter, patience, clarity, and even your confusion. Thank you for being my friend. You are a pillar in my life, and when I think about it, I feel invincible. You make me feel myself squared.
Thank you, Virgi, from whom I have drawn inspiration so many times that I feel we have shaped each other, and it is an honor to know that my personality resembles you.
Thank you, Gio, who never forgets to offer me a smile and reminds me that sometimes, all it takes is a hotel breakfast to feel in harmony with the universe.
Thank you, Claire and Sophi, who, in such a short time, have become trusted companions and a safe haven.
Thank you to Amanda, Emma, and Eleonora, who made a semester of studies the richest and sweetest of all.
Thank you to the Montagneros, a pack of women with dazzling and enviable personalities, capable of inspiring every feeling within the spectrum of feelings, and perhaps even some yet to be defined. I will cherish the memories we share.
Thank you to Rizz, Jaco, Cate, Chiara, Sanci, Agni, Marghe, Ceci, Manu, Pincho, Rosa, and Ale. So nice to write your names and knowing with certainty that if I ever wanted to write a thesis in 25 years, I would write them again.
Thanks to Claudia, my supervisor, for her availability and flexibility, which allowed me to undertake a thesis journey abroad.
Thank you to Adrien and Alberto, who accompanied me through this final stage of my journey, guiding me with expertise but always with a friendly spirit.
Thank you to my Fika companions at the KTH Department, for the warmth and kindness with which you welcomed me.

And so, I conclude what has become an ode to friendship: though I may have personally written this thesis, I thank all of you for giving it meaning.

References

[1] "2023 Emilia-Romagna floods - Wikipedia." https://en.wikipedia.org/wiki/2023_Emilia-Romagna_floods.

[2] European Environment Agency, "Alluvione dell'Emilia-Romagna del 2023 - Wikipedia." https://it.wikipedia.org/wiki/Alluvione_dell%27Emilia-Romagna_del_2023, 2024.

[3] "Climate change: What the EU is doing - Consilium." https://www.consilium.europa.eu/en/policies/climate-change/#EU-countries.

[4] "Share of energy consumption from renewable sources in Europe." https://www.eea.europa.eu/en/analysis/indicators/share-of-energy-consumption-from, Oct. 2024.

[5] "Europe – Countries & Regions," *IEA*. https://www.iea.org/regions/europe.

[6] "Home - Eurostat." https://ec.europa.eu/eurostat/.

[7] Eurostat, "NRG_D_HHQ: Energy statistics - quantities (annual data)." European Commission, 2024.

[8] S. Gehlin, O. Andersson, and J.-E. Rosberg, "Country Update for Sweden 2020."

[9] A. Manzella *et al.*, "Geothermal Energy Use, Country Update for Italy."

[10] EHPA, "Executive-summary_EHPA-heat-pump-market-and-statistic-report-2024." https://www.ehpa.org/wp-content/uploads/2024/08/Executive-summary_EHPA-heat-pump-market-and-statistic-report-2024.pdf, 2024.

[11] T. Hastie, "The elements of statistical learning: Data mining, inference, and prediction." Springer, 2009.

[12] Y. Song, *Data-Driven Strategies for Heat Pump Systems: A journey from inadequate data towards knowledge-based services*. Stockholm: KTH Royal Institute of Technology, 2024.

[13] M. Krieger, K. A. Kurek, and M. Brommer, "Global geothermal industry data collection: A systematic review," *Geothermics*, vol. 104, p. 102457, Sep. 2022, doi: 10.1016/j.geothermics.2022.102457

[14] J. De Wit and E. Hallgrimsdottir, *Geothermal Data Management: Good Practices and the Role of Regulators and Developers*. World Bank, Washington, DC, 2023. doi: 10.1596/39800

[15] I. Staffell, D. Brett, N. Brandon, and A. Hawkes, "A review of domestic heat pumps," *Energy & Environmental Science*, vol. 5, no. 11, p. 9291, 2012, doi: 10.1039/c2ee22653g

[16] Fürtbauer, Daniel and Cheng, Chuan, "InvestigationofPipelineHeatLosstoGroundInfluencedbyAdjacentPipelines." Apr. 2019.

[17] G. A. Mouzeviris and K. T. Papakostas, "Comparative analysis of air-to-water and ground source heat pumps performances," *International Journal of Sustainable Energy*, vol. 40, no. 1, pp. 69–84, Jan. 2021, doi: 10.1080/14786451.2020.1794864

[18] S. Gehlin, "Thermal response test: Method development and evaluation."

[19] A. Priarone, S. Morchio, M. Fossa, and S. Memme, "Low-Cost Distributed Thermal Response Test for the Estimation of Thermal Ground and Grout Conductivities in Geothermal Heat Pump Applications," *Energies*, vol. 16, no. 21, p. 7393, Nov. 2023, doi: 10.3390/en16217393

[20] F. Tinti, R. Bruno, and S. Focaccia, "Thermal response test for shallow geothermal applications: A probabilistic analysis approach," *Geothermal Energy*, vol. 3, no. 1, p. 6, Dec. 2015, doi: 10.1186/s40517-015-0025-5

[21] W. Mazzotti, J. Acuña, A. Lazzarotto, and B. Palm, "Project number at Swedish Energy Agency 40934-."

[22] A. Lazzarotto, W. M. Pallard, T. Richert, and A. Vautrin, "Improved Borehole Technology for Geothermal Heat Pumps Development."

[23] A. McDaniel, J. Tinjum, D. J. Hart, Y.-F. Lin, A. Stumpf, and L. Thomas, "Distributed thermal response test to analyze thermal properties in heterogeneous lithology," *Geothermics*, vol. 76, pp. 116–124, Nov. 2018, doi: 10.1016/j.geothermics.2018.07.003

[24] Y.-L. Nian, X.-Y. Wang, K. Xie, and W.-L. Cheng, "Estimation of ground thermal properties for coaxial BHE through distributed thermal response test," *Renewable Energy*, vol. 152, pp. 1209–1219, Jun. 2020, doi: 10.1016/j.renene.2020.02.006

[25] J. Acuña and B. Palm, "Distributed thermal response tests on pipe-in-pipe borehole heat exchangers," *Applied Energy*, vol. 109, pp. 312–320, Sep. 2013, doi: 10.1016/j.apenergy.2013.01.024

[26] L. C. B. Silva, M. E. V. Segatto, and C. E. S. Castellani, "Raman scattering-based distributed temperature sensors: A comprehensive literature review over the past 37 years and towards new avenues," *Optical Fiber Technology*, vol. 74, p. 103091, Dec. 2022, doi: 10.1016/j.yofte.2022.103091

[27] C. A. Parra-Murillo, M. F. Santos, C. H. Monken, and A. Jorio, "Stokes–anti-Stokes correlation in the inelastic scattering of light by matter and generalization of the Bose-Einstein population function," *Physical Review B*, vol. 93, no. 12, p. 125141, Mar. 2016, doi: 10.1103/PhysRevB.93.125141

[28] W. M. Mazzotti Pallard, A. Lazzarotto, J. Acuña, and B. Palm, "Calibration and Uncertainty Quantification for Single-Ended Raman-Based Distributed Temperature Sensing: Case Study in a 800 m Deep Coaxial Borehole Heat Exchanger," *Sensors*, vol. 23, no. 12, p. 5498, Jun. 2023, doi: 10.3390/s23125498

[29] M. B. Hausner, F. Suárez, K. E. Glander, N. V. D. Giesen, J. S. Selker, and S. W. Tyler, "Calibrating Single-Ended Fiber-Optic Raman Spectra Distributed Temperature Sensing Data," *Sensors*, vol. 11, no. 11, pp. 10859–10879, Nov. 2011, doi: 10.3390/s111110859

[30] J. Acuña, B. Palm, R. Khodabandeh, and K. Weber, "Distributed temperature measurements on a u-pipe thermosyphon borehole heat exchanger with CO2," in *Refrigeration science and technology proceedings*, KTH, Applied Thermodynamics and Refrigeration, 2010.

[31] J. Acuña, *Distributed thermal response tests: New insights on U-pipe and Coaxial heat exchangers in groundwater-filled boreholes*. Stockholm: KTH Royal Institute of Technology, 2013.

[32] J. S. Selker *et al.*, "Distributed fiber-optic temperature sensing for hydrologic systems," *Water Resources Research*, vol. 42, no. 12, p. 2006WR005326, Dec. 2006, doi: 10.1029/2006WR005326

[33] J. M. Bahr and D. J. Hart, "Distributed temperature sensing (DTS) as a hydrostratigraphic characterization tool."

[34] V. F. Bense *et al.*, "Distributed T emperature S ensing as a downhole tool in hydrogeology," *Water Resources Research*, vol. 52, no. 12, pp. 9259–9273, Dec. 2016, doi: 10.1002/2016WR018869

[35] R. Richter, C. Hutengs, C. Wirth, L. Bannehr, and M. Vohland, “Detecting Tree Species Effects on Forest Canopy Temperatures with Thermal Remote Sensing: The Role of Spatial Resolution,” *Remote Sensing*, vol. 13, no. 1, p. 135, Jan. 2021, doi: 10.3390/rs13010135

[36] C. A. Keller *et al.*, “Fiber optic distributed temperature sensing for the determination of the nocturnal atmospheric boundary layer height,” *Atmospheric Measurement Techniques*, vol. 4, no. 2, pp. 143–149, 2011, doi: 10.5194/amt-4-143-2011

[37] A. L. McDaniel, “District-Scale Ground Heat Exchange Performance with Fiber Optic Distributed Temperature Sensing.”

[38] C. S. Lowry, J. F. Walker, R. J. Hunt, and M. P. Anderson, “Identifying spatial variability of groundwater discharge in a wetland stream using a distributed temperature sensor,” *Water Resources Research*, vol. 43, no. 10, p. 2007WR006145, Oct. 2007, doi: 10.1029/2007WR006145

[39] T. I. Coleman, B. L. Parker, C. H. Maldaner, and M. J. Mondanos, “Groundwater flow characterization in a fractured bedrock aquifer using active DTS tests in sealed boreholes,” *Journal of Hydrology*, vol. 528, pp. 449–462, Sep. 2015, doi: 10.1016/j.jhydrol.2015.06.061

[40] O. Khankishiyev, S. Salehi, G. Hasanov, and Z. Hu, “Application of Distributed Temperature Sensing (DTS) in Geothermal Wells.”

[41] S. S. Naicker, “Performance Analysis of a Large-Scale Ground Source Heat Pump System.”

[42] C. Natale, M. Dongellini, C. Naldi, and G. L. Morini, “Experimental study on a Dual-Source Heat Pump in ground mode to assess the soil thermal response by means of a Distributed Temperature Sensing system,” *Journal of Physics: Conference Series*, vol. 2685, no. 1, p. 012036, Jan. 2024, doi: 10.1088/1742-6596/2685/1/012036

[43] M. S. Araujo, H. A. Spidle, S. P. Siebenaler, S. G. Blaisdell, and D. W. Vickers, “Application of Machine Learning to Distributed Temperature Sensing (DTS) Systems,” in *Volume 3: Operations, Monitoring, and Maintenance; Materials and Joining*, Calgary, Alberta, Canada: American Society of Mechanical Engineers, Sep. 2018, p. V003T04A024. doi: 10.1115/IPC2018-78640

[44] S. P. Szklarz *et al.*, “Ensemble-based assisted history matching of DTS monitoring data in HT-ATES systems: Application to a real-life case study,” *Geothermal Energy*, vol. 12, no. 1, p. 46, Dec. 2024, doi: 10.1186/s40517-024-00329-y

[45] S. Gelin, “Annex 52 - Instrumentation Guideline,” Heat Pump Centre, 2021. doi: 10.23697/TGR4-QN89

[46] “Home – Project Haystack.” https://www.project-haystack.org/.

[47] “Introduction.” https://brickschema.org/.

[48] Anonymous, “DTS Data Processing,” *Center for Transformative Environmental Monitoring Programs*. https://ctemps.org/data-processing, Oct. 2015.

[49] “GitHub - klapo/pyfocs: Processing functions for Fiber Optic Distributed Sensing (FODS) data.” https://github.com/klapo/pyfocs?tab=readme-ov-file.

[50] M. Domanski and F. D. Day-Lewis, “DTSGUI.” U.S. Geological Survey, 2019. doi: 10.5066/P91Z7ZAZ

[51] B. Des Tombe, B. Schilperoort, and M. Bakker, “Estimation of Temperature and Associated Uncertainty from Fiber-Optic Raman-Spectrum Distributed Temperature Sensing,” *Sensors*, vol. 20, no. 8, p. 2235, Apr. 2020, doi: 10.3390/s20082235

[52] “Outpost-os/python-dts-utils.” Outpost Operating System, Sep. 2024.

[53] A. Vautrin, W. M. Pallard, J. Acuna, and A. Lazzarotto, "Assessing coaxial collector performances: Field data and analysis," in *2024 Research Conference Proceedings of the IGSHPA*, International Ground Source Heat Pump Association, May 2024. doi: 10.22488/okstate.24.000040
[54] "Wells." https://www.sgu.se/en/products/maps/map-viewer/groundwater-map-viewers/wells/.
[55] "GoogleEarth," *Google Earth*. https://www.google.com/intl/it/earth/.
[56] "Optimal flow for maximum performance of geothermal heat pumps," *KTH*. https://www.liveinlab.kth.se/en/projekt/r-d-projects/optimalt-flode-for-maximal-prestanda-av-bergvarmepumpar-1.1298631.
[57] Devico, "DeviDrill: Steerable wireline core barrel."
[58] G. Florides and S. Kalogirou, "Measurements of Ground Temperature at Various Depths."
[59] "The climate - data, analyses and research | SMHI." https://www.smhi.se/en/about-smhi/what-we-do/the-climate-data-analyses-and-research-1.83761.
[60] S. Reece and S. Roberts, "An introduction to Gaussian processes for the Kalman filter expert," in *2010 13th International Conference on Information Fusion*, Edinburgh: IEEE, Jul. 2010, pp. 1–9. doi: 10.1109/ICIF.2010.5711863
[61] D. Simon, *Optimal State Estimation: Kalman, H∞, and Nonlinear Approaches*, 1st ed. Wiley, 2006. doi: 10.1002/0470045345
[62] L. R. Weill and P. N. De Land, "The Kalman filter: An introduction to the mathematics of linear least mean square recursive estimation," *International Journal of Mathematical Education in Science and Technology*, vol. 17, no. 3, pp. 347–366, May 1986, doi: 10.1080/0020739860170311
[63] J. Kung and G.-C. Rota, "A practical guide to splines," *Advances in Mathematics*, vol. 32, no. 1, p. 81, Apr. 1979, doi: 10.1016/0001-8708(79)90033-1
[64] P. Dierckx, *Curve and surface fitting with splines*, Repr. in Monographs on Numerical Analysis. Oxford: Clarendon, 1996.
[65] G. Micula and S. Micula, *Handbook of Splines*. Dordrecht: Springer Netherlands, 1999. doi: 10.1007/978-94-011-5338-6
[66] "The Julia Programming Language." https://julialang.org/.
[67] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, Jan. 2017, doi: 10.1137/141000671
[68] J. Eschle *et al.*, "Potential of the Julia Programming Language for High Energy Physics Computing," *Computing and Software for Big Science*, vol. 7, no. 1, p. 10, Dec. 2023, doi: 10.1007/s41781-023-00104-x
[69] P. G. D. Group, "PostgreSQL," *PostgreSQL*. https://www.postgresql.org/, Jan. 2025.
[70] "pgAdmin - PostgreSQL Tools." https://www.pgadmin.org/.
[71] S. Juba, A. Vannahme, and A. Volkov, *Learning PostgreSQL: Create, develop, and manage relational databases in real-world applications using PostgreSQL*, 1st ed. in Community Experience Distilled. Place of publication not identified: Packt Publishing, 2015.
[72] P. Zippenfenig, "Open-Meteo.com Weather API." Zenodo, Dec. 2024. doi: 10.5281/ZENODO.7970649

[73] X. Yu, R. Ghasemizadeh, I. Y. Padilla, D. Kaeli, and A. Alshawabkeh, "Patterns of temporal scaling of groundwater level fluctuation," *Journal of Hydrology*, vol. 536, pp. 485–495, May 2016, doi: 10.1016/j.jhydrol.2016.03.018

[74] N.-C. Chen *et al.*, "Investigation and Estimation of Groundwater Level Fluctuation Potential: A Case Study in the Pei-Kang River Basin and Chou-Shui River Basin of the Taiwan Mountainous Region," *Applied Sciences*, vol. 12, no. 14, p. 7060, Jul. 2022, doi: 10.3390/app12147060

[75] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967, doi: 10.1109/TAU.1967.1161901

[76] R. Pintelon and J. Schoukens, *System identification: A frequency domain approach*. John Wiley & Sons, 2012.

[77] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005, doi: 10.1109/JPROC.2004.840301