Chair of Applied Geophysics 1:
Computational Geoscience,
Geothermics and Reservoir
Geophysics (CGGR)

MASTER OF SCIENCE IN APPLIED GEOPHYSICS

MASTER THESIS

The present work was submitted to the Chair of Computational Geoscience, Geothermics and Reservoir Geophysics

---

Application of the Non-Intrusive Reduced-Basis Method in Magnetotellurics for High-Temperature Geothermal Systems

Nadja Lindner
Student No. 438378

Dr. Denise Degen
Dr. Alexander Grayver
Prof. Florian Wellmann, Ph. D.

August 2, 2023

Application of the Non-Intrusive Reduced-Basis Method in Magnetotellurics for High-Temperature Geothermal Systems

Master of Science Thesis

for the degree of Master of Science in Applied Geophysics
by

Nadja Lindner

Student No. 438378

August 2, 2023

IDEA LEAGUE
JOINT MASTER'S IN APPLIED GEOPHYSICS

Delft University of Technology, The Netherlands
ETH Zürich, Switzerland
RWTH Aachen, Germany

Dated: *August 2, 2023*

Supervisors:

Dr. Denise Degen

Dr. Alexander Grayver

Prof. Florian Wellmann, Ph. D.

Committee Members:
Bold names represent Examiners.

Dr. Denise Degen

Dr. Alexander Grayver

Prof. Florian Wellmann, Ph. D.

Dr. Deyan Draganov

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Declaration of Academic Integrity

Lindner, Nadja
Name, Vorname/Last Name, First Name

438378
Matrikelnummer (freiwillige Angabe)
Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende ~~Arbeit/Bachelorarbeit/~~ Masterarbeit* mit dem Titel
I hereby declare under penalty of perjury that I have completed the present paper/bachelor's thesis/master's thesis* entitled

Application of the Non-Intrusive Reduced-Basis Method in Magnetotellurics for High-Temperature Geothermal Systems

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without unauthorized assistance from third parties (in particular academic ghostwriting). I have not used any other sources or aids than those indicated. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Aachen, 03/08/2023
Ort, Datum/City, Date

N. Lindner
Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
§ 156 StGB (German Criminal Code): False Unsworn Declarations
Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence
(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding three years or to a fine.
(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Aachen, 03/08/2023
Ort, Datum/City, Date

N. Lindner
Unterschrift/Signature

Abstract

Magnetotellurics is a widely-used geophysical method for characterizing variations in electrical properties of the subsurface. Its ability to achieve large penetration depths and its sensitivity to fluids make it particularly advantageous for assessing the potential of high-temperature geothermal systems. However, its computationally demanding nature poses challenges in data interpretation, especially for 3D applications. Thus, efficient subsurface characterizations are essential for geothermal projects, as it directly impacts resource- and technical potential, planning of heat extraction and safety control.

We propose the combination of the non-intrusive reduced-basis method with magnetotellurics simulations to address these challenges. The non-intrusive reduced basis method aims to construct surrogate models of reduced order, based on input training data obtained by GoFEM, a solver specified for geoelectromagnetic applications. We demonstrate the effectiveness of our approach with numerical experiments on the real-valued magnetotellurics parameters resistivity and phase, as well as on the complex-valued impedance. Results show significantly lower computational costs compared to conventional magnetotellurics simulation approaches, while keeping accuracy level high. When the impedance is split into its real and imaginary component during the construction of the surrogate model, we are successfully below the desired error tolerance of maximum 10^{-4} globally and 10^{-2} locally. The transformation of the complex-valued impedance into a real-valued matrix currently exceeds this threshold, but it still remains a possible approach due to its higher model consistency. The advantage of obtaining a large speed-up by a factor of 10^4 in forward simulations due to the non-intrusive reduced-basis method can possibly be used in probabilistic approaches like in the Markov Chain Monte Carlo method, for the resistivity and phase, as well as for the impedance directly. During this research, we discuss the beneficial use of the non-intrusive reduced-basis method for magnetotellurics, as well as current limitations, while outlining future research directions. The non-intrusive reduced-basis method shows great potential for improving magnetotellurics data interpretation and advancing geophysical modeling for complex-valued parameters in practical applications.

August 2, 2023

Acknowledgements

I would like to express my genuine gratitude to everyone who has enabled me to complete my journey of this master thesis enthusiastically and well-guided, especially the following ones:

First and foremost, I have much appreciation for my primary supervisor, Dr. Denise Degen, for her everlasting guidance and inspiring expertise. My questions and problems were constantly supported and valued, and I felt taken-seriously and cherished. Her constructive feedback and immediate support were keeping my motivation high during the entire thesis, and I am very grateful to have had her as my supervisor.

I extend my appreciation to Dr. Alexander Grayver for his valuable suggestions and inputs, which significantly enhanced the quality of my research and helped me focus on important matter. His expertise in magnetotellurics and geothermal applications eased my entry into this fascinating field.

Furthermore, I would like to express gratitude to Prof. Florian Wellmann, Ph. D., for his support and his innovative ideas when I faced difficulties. His extensive research experience, analytical thinking and positive attitude widened my perspective and enriched my research.

I would like to extend my appreciation to my family, friends, and office mates for their encouraging words when needed, patience and support during this demanding period, and understanding in situations where I personally might had lacked of it. I am thankful for their acceptance for research discussion during dinner, late-night coding sessions, and coffee-breaks when distraction was needed. Their support has provided me with a positive attitude throughout this thesis.

Last but not least, I would like to thank you, dear reader, for showing interest in my project, on which I have worked with great passion and dedication. I hope you enjoy reading this thesis, gain knowledge, or perhaps even question certain aspects. If the latter is the case, I invite you to get in touch with me, as research is a collaborative world of knowledge exchange.

RWTH Aachen University Nadja Lindner
August 2, 2023

Table of Contents

Acronyms

AI	artificial intelligence
BC	boundary condition
RB	reduced basis
CVNN	complex-valued neural network
DoF	degrees of freedom
E	electric
EGS	enhanced geothermal systems
EM	electromagnetics
FE	finite element
FEM	finite element method
GP	gaussian process
H	magnetic
LHS	latin hypercube sampling
MCMC	Markov chain Monte Carlo
ML	machine learning
MSE	mean squared error
MT	magnetotellurics
NI-RB	non-intrusive reduced basis
NN	neural network
PDE	partial differential equation
PINN	physical-informed neural network
PB-ML	physics-based machine learning
POD	proper orthogonal decomposition
RB	reduced-basis
ROM	reduced order modeling
SA	sensitivity analysis
SVD	singular value decomposition
TE	transverse electric
TM	transverse magnetic
UQ	uncertainty quantification
Z	impedance

Nomenclature

$\vec{\boldsymbol{\Sigma}}$	$\operatorname{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_{N_s}\}$
δ	Skin depth [m]
ϵ	Error tolerance
ϵ_0	Electric permittivity [$\approx 8.8541 \cdot 10^{-12}\,\mathrm{A\,s\,V^{-1}\,m^{-1}}$]
$\mathbf{W}$	Complex-valued matrix
μ_0	Magnetic permeability [$4\pi \cdot 10^{-7}\,\mathrm{V\,s\,A\,m^{-1}}$]
ω	Angular frequency [rad $\mathrm{s^{-1}}$]
$\overrightarrow{\mu}$	Parameter
∂_t	Temporal partial differentiation
ϕ	Impedance Phase [rad]
ψ_i	Basis functions
ρ	Electrical resistivity [Ω m]
σ	Electrical conductivity [S $\mathrm{m^{-1}}$]
θ_{rb}	Reduced coefficients
ε	Electric permittivity [F $\mathrm{m^{-1}}$]
$\vec{B}$	Magnetic flux density [$\mathrm{V\,s\,m^{-2}}$]
$\vec{D}$	Electric flux density [$\mathrm{A\,s\,m^{-2}}$]
$\vec{E}$	Electric field intensity [$\mathrm{Vm^{-1}}$]
$\vec{E}_{\mathrm{x}}$	Electric field intensity in the x-direction [$\mathrm{Vm^{-1}}$]
$\vec{E}_{\mathrm{y}}$	Electric field intensity in the y-direction [$\mathrm{Vm^{-1}}$]
$\vec{E}^{+}$	Vector constant of the electric field [$\mathrm{Vm^{-1}}$]
$\vec{E}^{-}$	Vector constant of the electric field [$\mathrm{Vm^{-1}}$]
$\vec{H}$	Magnetic field intensity [$\mathrm{A\,m^{-1}}$]
$\vec{H}_{\mathrm{x}}$	Magnetic field intensity in the x-direction [$\mathrm{A\,m^{-1}}$]
$\vec{j}$	Electric current density [$\mathrm{A\,m^{-2}}$]
$\vec{J}^{e}$	Source volume densities of the electric current
$\vec{J}^{m}$	Source volume densities of the magnetic current
$\vec{S}$	Snapshot matrix
$\vec{U}$	Orthogonal matrix
$\vec{u}_h(\boldsymbol{\mu})$	Snapshot for any parameter μ
$\vec{u}_{rb}$	Reduced Solution
$\vec{Z}$	Orthogonal matrix
H_{z}	Magnetic field intensity in the z-direction [$\mathrm{A\,m^{-1}}$]
ξ	Eigenvalues
f	Frequency [$\mathrm{s^{-1}}$]
k	Wavenumber [$\mathrm{m^{-1}}$]
N	Total number of training samples
q	Electric charge density [$\mathrm{A\,s\,m^{-3}}$]
r	Size of the reduced basis
T	Phase [s]
V_{rb}	Reduced space
Z	Impedance [Ω]
z	Depth [m]

Chapter 1

Introduction

The growing importance and need of renewable energy is nowadays well-known (Farghali et al., 2023). Global warming, diminishing fossil fuel sources and the current energy crisis have led to the starting shift from conventional energy to the renewable sector. Geothermal energy is nearly inexhaustible, reliable and predictable, which is beneficial compared to other renewable energy sources (Kulasekara and Seynulabdeen, 2019). For these reasons, research about an efficient use of geothermal reservoirs has increased over the last decades, currently prioritizing heat generation rather than electricity production (Baillieux et al., 2014; Jolie et al., 2021; Kabeyi, 2019; Käufl et al., 2020; Tseng et al., 2020). In order to extent the application of geothermal systems beyond their current limits, research is additionally conducted on enhanced geothermal systems. These systems include a host rock of too low permeability for geothermal heat extraction, which it is artificially increased by creating fractures to enhance fluid flow (Olasolo et al., 2016). To decrease risk of borehole failures and strong seismic activities caused by fracturing, further detailed knowledge about geothermal systems is required.

This knowledge can be obtained by, e.g., geological or geophysical measurements. One of this geophysical techniques is magnetotellurics (MT), which reveals variations of the electrical conductivity in the subsurface. It is a passive technique, as it uses the Earth's natural electric (E) and magnetic (H) fields as a source. This source is of very low frequency in a range of milli- to kilohertz, enabling a deep penetration depth of hundreds of kilometers (Siniscalchi et al., 2021). This advantage, and, additionally, its sensitivity to the presence of fluids due to strong electrical conductivity contrasts are two main benefits making MT a suitable geophysical method for geothermal systems (Peacock et al., 2012). Other geophysical measurement techniques are either incapable of reaching the typical depth of geothermal systems of a few kilometers, or of revealing the availability of fluids. The mentioned advantages of MT lead it its capabilities of estimating temperature, porosity, and permeability due to their dependency on the electrical conductivity, which is beneficial for analysing geothermal reservoirs potential (Munoz, 2014; Ussher et al., 2000). In the last decades, MT has been frequently used for geothermal applications, in low- and intermediate- (Bertrand et al., 2013; Heise et al., 2016; Hill et al., 2015), as well as in high-temperature geothermal systems (Munoz et al., 2010; Peacock et al., 2012, 2013).

However, geoscientific measurements and applications often lack of data, leading to problems in simulations and data interpretation (Degen et al., 2022). This also applies for MT applications, making it challenging for machine learning (ML)-based approaches, where large data sets are needed to obtain accurate results (Ahmed et al., 2023). ML implementations allows to simulate data and extract relevant information from it autonomously and accurately with low computational costs when the data set generation is cheap (Mahesh, 2020). Thus, applying ML in geosciences has the chance to approach data in a different, efficient way. Therefore, there is an urgent need to overcome the problem of data sparsity in scientific and engineering applications, in order to enable ML approaches with reduced computational costs. Related to the application in geothermal, it would allow us to increase out detailed knowledge of the subsurface required for optimal geothermal energy production (Swischuk et al., 2019). The use of the non-intrusive reduced basis (NI-RB) method in MT applications has the potential to tackle these problems (Degen et al., 2022). It is an approach combining physics-based models with data-driven ML, taking advantage of each of these techniques individually. While maintaining a large interpretability and transparency from the physics-based approach, and allowing low computational costs due to the combination of proper orthogonal decompositions (PODs) with ML techniques, we enable an efficient modelling technique (Wang et al., 2022b) by the construction of surrogate models based on forward simulations (Degen et al., 2022). In this project, these forward simulations are obtained by the software GoFEM, a scalable and robust solver specifically created for geoelectromagnetic applications (Grayver and Kolev, 2015). The chance of enhancing the computational approach to handle low-density MT data, and enabling low-cost simulations with keeping the accuracy high, can widen the knowledge needed for efficient green energy production from geothermal systems.

The response obtained by MT includes the complex-valued impedance, as well as the real-valued parameters resistivity and phase measured by the electromagnetic field. These three properties are of interest to derive the electrical conductivity distribution of the subsurface. For this reason, the goal of this project is to construct surrogate models for simulations of the impedance, resistivity and phase. The implementation of complex-valued variables, in our case the impedance, in ML techniques requires adaption of traditional methods, such as of the structure of neural networks (NNs) (Lopez-Pacheco and Yu, 2022). Splitting complex-valued parameters into real and imaginary parts and treating them separately as two reduced models has been one possible approach used in science (Lopez-Pacheco and Yu, 2022). Another possibility is the use of complex-valued neural networks (CVNNs), which have specified training methods, as activation functions, back-propagation operations, and gradient descent algorithms adapted to the complex domain (Lopez-Pacheco and Yu, 2022). Nowadays, the use of CVNNs gains in interest, especially in the application of radio frequency signal-, image-, and audio signal processing (Bassey et al., 2021). In MT, research about modelling complex-valued impedance using CVNNs has been conducted, but with lacking accuracy in case of sparse data. Imamura and Schultz (2020) show an application of CVNNs for the impedance in MT applications, which shows a successful implementation when having large data sets. However, research of CVNNs is still sparse nowadays, especially in the area of MT. As we combine such a ML approach with the NI-RB method in this project, we need to ensure the correct coupling of both applications, which can be challenging and requires the right implementation of CVNNs, if applied. Another approach is the transformation from complex to real values (Barwick, 2016). The real and imaginary parts are reorganized in a matrix, which is in the real domain. These three possibilities are tested throughout this thesis.

In this project, we overall try to enhance the performance and accuracy of simulations of the resistivity, phase and impedance of the MT response. We use the NI-RB method to overcome the common MT data sparsity, while keeping the model error low. The combination of the NI-RB method and MT measurements has not been tested before, giving this project an experimental character. However, reduced-basis (RB) methods have been successfully applied for MT applications, focusing on real-valued properties as the resistivity and phase (Manassero et al., 2020). The intrusive RB-method has the ability of maintaining the laws of physics and of having a defined error bound (Degen et al., 2022). The current research of RB methods in electromagnetics (EM) applications focuses mostly on the Maxwell Equations in the frequency domain (Benner and Hess, 2017; Hess and Benner, 2013). The traditional RB method works very well for this application, however, the potential is a real-valued property, thus making its application simpler than the complex-valued impedance. Both of the potential and the impedance are obtained from the Maxwell Equations, but the potential focuses on the electrical field (Benner and Hess, 2017). The implementation of the impedance, instead of the resistivity and phase, in RB methods is rather new, making this project of exploratory nature. The NI-RB method, compared to the intrusive RB method, has a larger flexibility in its applications, as these do not have to be implemented in the NI-RB method. It uses purely the forward simulations as an input for the surrogate model construction, and thus allows adaptions to more realistic scenarios due to less geological assumptions made in the data generation. Additionally, it can handle non-linear and hyperbolic problems. For these reasons, it is of interest to introduce the NI-RB method in MT applications. If successful, the project enables a wider use of MT simulations for the detection of structural interfaces in the subsurface, as the complex-valued impedance can be used with relatively low computational costs due to the implementation of ML.

During this study, the focus lies on the construction of 2D MT surrogate models created by the NI-RB method (Degen et al., 2022). As a guideline, this following research question is leading through this project, and is tried to be answered along the process:

> *How can we apply the NI-RB method to the real-valued phase and resistivity, as well as the complex-valued impedance of the MT responses, in order to stay below the desired maximum local error of 1 %?*

We try to answer this research question in Chapter 4, after having achieved the desired results. As a preliminary step, we formulate a working procedure for this problem statement in order to provide an indication of the study direction:

> *The phase and resistivity are assumed to be usable in the NI-RB method without requiring too many adjustments of the current available library. For the impedance, we first try to split this parameter into its the real and imaginary component. However, as previously mentioned, this approach does not fulfill the ultimate goal of treating the impedance as a complex-valued variable in its entirety. The latter approach would allow us to train a single surrogate model, whereas a split would require two separate models, one for the real and the imaginary component. As a disadvantage, training two reduced models compared to just one leads to an increase model inconsistency. Thus, in its final stage we aim for having a single model treating the impedance as one complex-valued parameter.*

> *For this reason, an approach needs to be found that ultimately has the ability of dealing with complex values. The application method of gaussian process (GP) regression and neural network (NN) is established in the NI-RB method, thus it needs to be decided which approach is more suitable. Structures of codes and its functions need to be adapted accordingly to the use of complex values.*

This hypothesis aims to demonstrate the possible approach of this thesis. It still needs to be considered that research often includes changes in approaches, thus variations can occur.

Thus, the general aim of this project is to combine the NI-RB method as a computational approach taking advantage of physics-based models and data-driven ML approaches, and the geophysical measurement technique MT, which enables revealing variations of the electrical conductivity in the subsurface. This research has the ultimate goal to accelerate simulations and therefore to ease investigations of high-temperature geothermal systems, as the common field can lead to interpretation difficulties. The combination of the NI-RB method with MT would allow simulations of high accuracy, independent of the sparsity of the data.

As an overall structure of this project, this thesis is be divided into four main chapters. Theoretical and methodological aspects of the research are covered in Chapter 2. It gives a background about geothermal systems and MT, introduces the simulation technique and shows the functionality of physics-based machine learning (PB-ML) and the NI-RB method in particular. Chapter 3 presents the research findings, including analyses phase, resistivity and impedance, as well as MT responses across different frequency ranges. It also includes a comparison of computational costs between the NI-RB method and GoFEM. In Chapter 4, a thorough discussion of the results is provided, contextualising them with in the broader research question and exploring potential applications. An outlook with continuing research ideas are mentioned in Chapter 5, along with concluding words.

Chapter 2

Theory and Methodology

This chapter provides an overview of the main methods addressed in this thesis. It introduces geothermal systems, the theory of MT, the software functionality applied in this project for simulations, as well as the background of the NI-RB method. It is worth noting that each section is not intended to fully cover the theory and methodology of the corresponding topic, but rather focuses on the specific knowledge required for this research.

2-1 Geothermal System

Geothermal systems besides other renewable energy sources have a growing importance in today's energy sector (Farghali et al., 2023), caused by the current energy crisis, diminishing resources, and the need for lowering CO_2 emissions due to global warming (Mohtasham, 2015). Enhancing the efficiency of heat extraction in geothermal energy production has gained priority for society. Compared to other renewable energies such as wind or solar energy, geothermal reservoirs represent a predictable and reliable source, as they are not weather-dependent and thus provide constant energy (Kulasekara and Seynulabdeen, 2019). To ensure safe and economically sustainable heat extraction, a better understanding of these complex systems is crucial. This section aims to give a broad overview over the reservoirs.

2-1-1 Current State

A shift to more renewable energies and an increase of related technological developments are nowadays urgently needed due to the previously mentioned rapid climate change and current energy crises worldwide (Farghali et al., 2023). These reasons have led to a growing importance of geothermal system during the last decades, as it has the potential to provide up to 150 GWe of renewable energy until the year 2050 (Jolie et al., 2021). Currently, it supplies over 90 countries with approximately 95.1 GWe used for both electricity and heat (Jolie et al., 2021), whereas the current focus of research lies on heat production with the need of further improvements in electricity recovery (Kabeyi, 2019). The geothermal heat source is nearly inexhaustible as it is present in large quantities in both the crust and deeper parts of the Earth (Barbier, 2002). However, efficient and safe heat extraction including a carefully chosen selection of appropriate well locations and pressure monitoring require sufficient knowledge about the local subsurface (Reinecker et al., 2019).

Thus, geothermal energy is a promising sector of renewable energy, which has led to a large increase in research during the last years. It also has been greatly financially supported by governments in various countries, such as in Germany (Agemar et al., 2014), China (Zhang et al., 2019), Australia (Bahadori et al., 2013), Indonesia (Alhamid et al., 2016) and Turkey (Kilic, 2016). Due to this support, the success rate of geothermal projects has increased, as the understanding of the geological sites and the applied technology has been improved. We could learn from failed projects, where seismic event of magnitudes M_L 3.4 in Basel, Switzerland, (Deichmann and Giardini, 2009) and 5.8 in Pohang, South Korea(Ellsworth et al., 2019), were caused due to an case-specific unsuitable high-pressure water injection, endangering the life of human's beings. As an example for well-working geothermal use, the German city Munich with over 1.5 million inhabitants plans to be the first European city with a district-heating network purely supplied by geothermal energy by 2040 (Farquharson et al., 2016), thanks to the well-suited geological conditions of this city in the carbonate Malm area (Farquharson et al., 2016). The United States have a current electricity-generating capacity over 3.8 GW installed (Kumar et al., 2022). Possibilities enabled by the use of enhanced geothermal systems (EGS) are currently explored, as they aim to allow efficient heat extraction in geological systems of too low natural permeability for an economically sustainable energy production compared to traditional, hydrothermal system s(Olasolo et al., 2016). To artificially increase the permeability, the host rock is mechanically or chemically fractured (Caulk et al., 2016), followed by fluid injection for heat or energy extraction (Lu, 2018). However, there are challenges, such as dealing with the presence of highly corrosive and hostile fluids in areas with critical temperatures (Kumar et al., 2022). Additionally, some geothermal reservoirs are located in hard and abrasive igneous or metamorphic rocks, making the understanding of the stress orientation crucial to stabilize the borehole properly (Kumar et al., 2022). Therefore, room of improvement is still large, and so further research is needed to enhance the use of geothermal energy beyond the current state.

2-1-2 Geological Background

Geothermal systems are closely linked to the geology present and the processes that led to the current geological state. Depending on the temperatures and pressure conditions dominating in the desired depth, these reservoirs can be subdivided into different types, which are also influenced by the geological settings, and the present formations and structures. The following sections give some insight about the Earth's geological structure and its connection to geothermal systems.

2-1-2-1 The Earth's structure and its heat flow

The Earth is divided into five layers, based on their chemical and physical properties like density and composition (Brown, 2012) (Figure 2-1). The innermost layer is the inner core, which is a solid layer consisting of mostly iron. Despite its high temperature, it is solid due to extremely large pressure. The outer core consists of iron in the liquid state. This liquid behavior creates the magnetic field of the Earth. The lower and upper mantle are both solid and consist of silicates, with the lower mantle being more rigid and the upper one more ductile (Brown, 2012). The crust is the outermost layer, and can be divided in thicker and less dense continental, and a denser and thinner oceanic crust. The geothermal energy is extracted from the Earth's crust, but its energy source comes from the deeper layers, thus being of larger importance to understand the potential of geothermal energy. Therefore, it is important to keep the concept of a layered earth in mind, focusing on where the source is located and where the heat is extracted.

The Earth's crustal mean heat flow is around 91.6 mW m^{-2} with variations depending on the local geological environment (Elders and Moore, 2016). In the continental crust the heat flow is generally lower than in the oceanic crust, as it shows an average of 70.9 and 105.4 mW m^{-2}, respectively (Elders and Moore, 2016). The Earth's heat sources is primarily derived from the radioactive decay of elements such as ^{238}U,^{232}Th, ^{40}K, as well as the residual heat generated during the planets initial accretion and its subsequent gravitational differentiation into core, mantle, and crust (Elders and Moore, 2016) (Figure 2-1). Therefore, the Earth's main heat is stored in the deep subsurface with a minimum depth of 5 km (Jolie et al., 2021). The heat can reach the surface by hydrothermal convection, where water is the transport medium of heat through permeable rocks and their fractures (Jolie et al., 2021). The heat flow produced by these sources gives rise to an average geothermal gradient of about 30 °C km^{-1} in stable areas, with a maximum of around 500 °C km^{-1} in tectonically active zones (Elders and Moore, 2016).

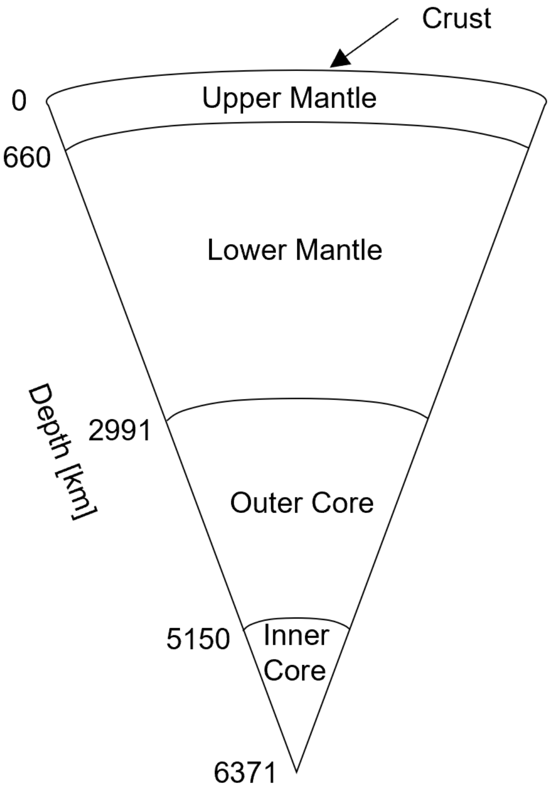

Figure 2-1: Sector of the Earth, showing its main five layers (modified after Brown (2012)).

2-1-2-2 Influence of the Geology on Geothermal Systems Types

Geothermal systems can be divided according to their reservoir temperature, their enthalpy including the pressure, or exergy focusing on the available energy (Zarrouk and McLean, 2019). Usually, these geothermal systems are divided into three temperature classes of (1) *low-temperature* (below 125 °C), (2) *intermediate-temperature* (between 125 and 225 °C), and

(3) *high-temperature* reservoirs (above 225 °C) (Jolie et al., 2021; Zarrouk and McLean, 2019), where low- and intermediate-temperature systems are non-magmatic and high-temperature magmatic systems. The presence of sufficient fluids in intermediate geothermal systems allows the use of binary power plant technologies for power generation (Jolie et al., 2021). High temperature reservoirs enable the implementation of flash power plant technologies besides the binary option (Jolie et al., 2021). Examples for low-temperature geothermal systems are, e.g., the Têt fault geothermal system in the Pyrenees (Taillefer et al., 2017), the Groß Schönebeck site in Germany (Munoz et al., 2010) and the IRETHERM project in Ireland (English et al., 2023). The r4d project in Mongolia is an intermediate-temperature system with past volcanic activities and temperatures of up to 95 °C in hot springs at the surface (Bignall et al., 2005; Käufl et al., 2020). High-temperature geothermal systems are, e.g., the Coso geothermal field in the USA (Newman et al., 2008) and projects in Aluto, Tulu Moye and Corbetti at the East African Rift (Mazzarini and Isola, 2010; Samrock et al., 2018; Hübert et al., 2018).

These different types of geothermal systems are influenced by geological settings, formations and structures present at each location (Jolie et al., 2021). Their impact is explained in the following paragraphs.

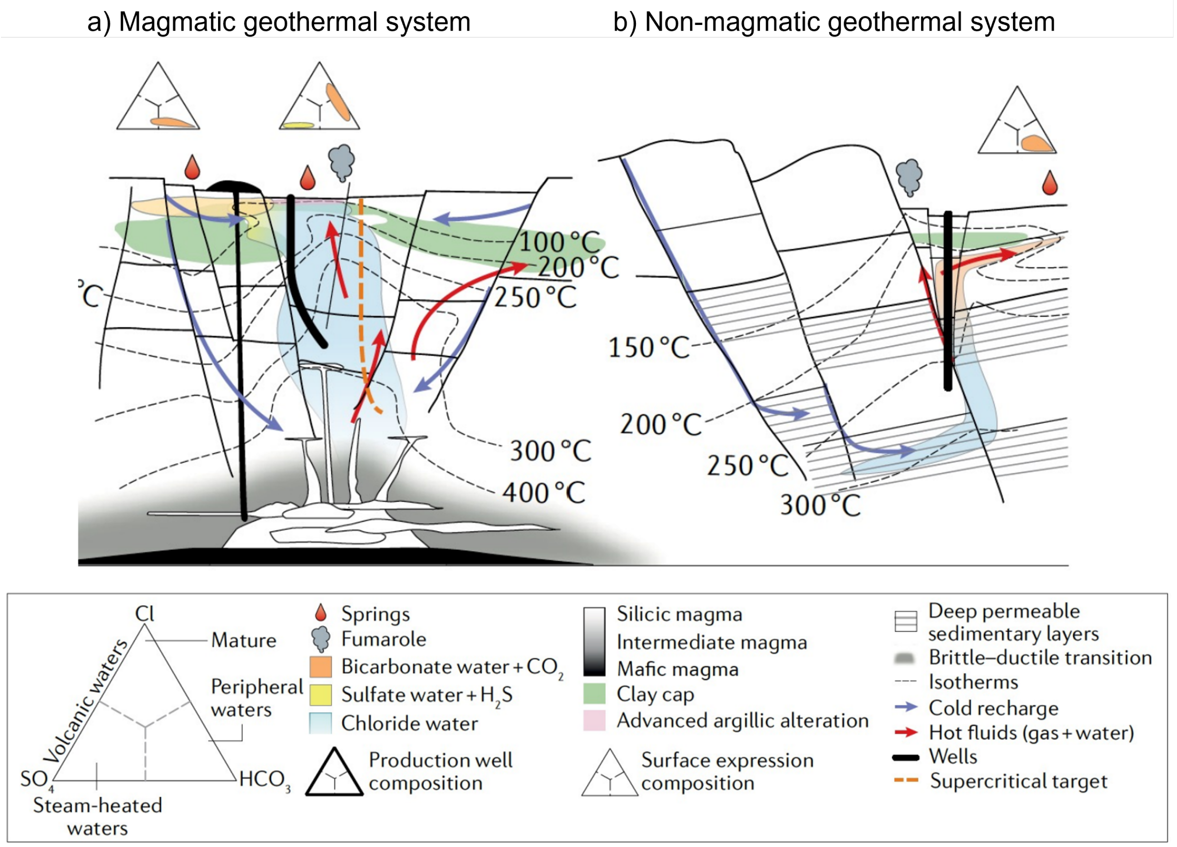

Figure 2-2: Schematic model of magmatic (a) and non-magmatic (b) geothermal systems (modified after (Jolie et al., 2021). a) Magmatic geothermal are of high-temperatures and the conductivity is dominated by hydrothermal alteration and melt from the magma chamber. b) Non-magmatic are of low- and mid-temperatures, where the conductivity is dominated by the porosity and pore fluids.

Geological settings have a large impact on the geothermal reservoir productivity (Jolie et al., 2021). In a stable area with a temperature gradient of 30 °C km^{-1}, heat transfer mostly corresponds to conduction (Jolie et al., 2021). In tectonically active regions with a temperature gradient of up to 500 °C km^{-1}, convection is dominating, allowing a faster heat transfer. In high-temperature geothermal systems, the conductivity is dominated by hydrothermal alteration and melt from magma chamber. On the other hand, the conductivity in non-magmatic systems is strongly influenced by permeability, porosity and the pore fluids. These characteristics lead to a heat flow per area that can be up to 10 times larger in volcanic areas than in non-magmatic provinces (Jolie et al., 2021). Permeable pathways enable fluid, solute and heat transfer (Luhmann et al., 2017). Oceanic crust often consist of basalts with a lower natural permeability of around $1.5{\cdot}10^{-16}\,\mathrm{m}^2$, while siliciclastic volcanic rocks allow a better heat extraction due to their higher permeability of minimum $10^{-15}\,\mathrm{m}^2$. In Figure 2-2, the difference of the geological setting of high- and mid- and low geothermal systems is shown in a schematic way.

Geological formations define important rock parameters as porosity and permeability which vary as a function of lithology, depth, stress regime and degree of hydrothermal alteration (Jolie et al., 2021). The permeability of a rock can vary over time, as secondary permeability caused by fractures, faults or rock dissolution can increase a formations primary permeability, while precipitation of secondary alteration minerals can lead to an decrease (Jolie et al., 2021).

Permeable structures like faults and fractures are the main factor on fluid flow rates in mid- to high temperature convective geothermal systems (Jolie et al., 2021). Such reservoirs often occur along faults, which allow average yield flow rates of over $0.1\,\mathrm{m}^3\,\mathrm{s}^{-1}$. If fractures are not present, pores in the permeable host rock are used as fluid paths instead.

2-1-2-3 Magmatic Geothermal Systems

Since the focus of this thesis is on magmatic geothermal systems, we introduce them in this Section in more detail. A traditional magmatic geothermal system is a hydrothermal system, which is dominated by water or steam (Munoz, 2014). In a basic matter it consists of a heat source, an upflow zone in which the heat is transported upwards, as well as an impermeable clay cap confining the system. A MT-based illustration of such a high-temperature geothermal systems, is shown in Figure 2-3. It represents the Aluto-Langano geothermal system at the Ethiopian Rift, describing the fluid- and heat flow in these reservoirs (Samrock et al., 2023). Based on such illustrations, we develop a simplified, schematic graphic shown in Figure 2-4, where the the heat source is set to a depth of 6 km. This depth can vary between usually 3 to 6 km (Samrock et al., 2018). Magmatic geothermal systems include large electrical conductivity contrasts, which are of advantage when applying the geophysical method of MT (Hill et al., 2015). The surrounding rock often consists of rather resistive fresh volcanic lithology (Samrock et al., 2018). The clay cap is more conductive, containing of the argillic smectite-illite alteration at temperatures below 200 °C. The reservoir is more resistive and measures temperatures above 200 °C, leading to a propylitic chlorite-illite alteration zone. The conductivity range of the heat source strongly depends on whether the magmatic system is active, which would lead to a higher conductivity as cooled intrusions. Possible electrical conductivity ranges of these described units can be found in Table 2-1.

Table 2-1: Typical Electrical Conductivity ranges chosen for each unit in a geothermal model (after Samrock et al. (2018)). These ranges are also used in the simulations of this project.

Unit	**Electrical conductivity [S m^{-1}]**
Air	$1 \cdot 10^{-12}$
Surrounding rock	0.001 - 0.02
Clay-cap	0.5 - 2
Upflow zone	0.001 - 0.5
Source	0.2 - 0.5

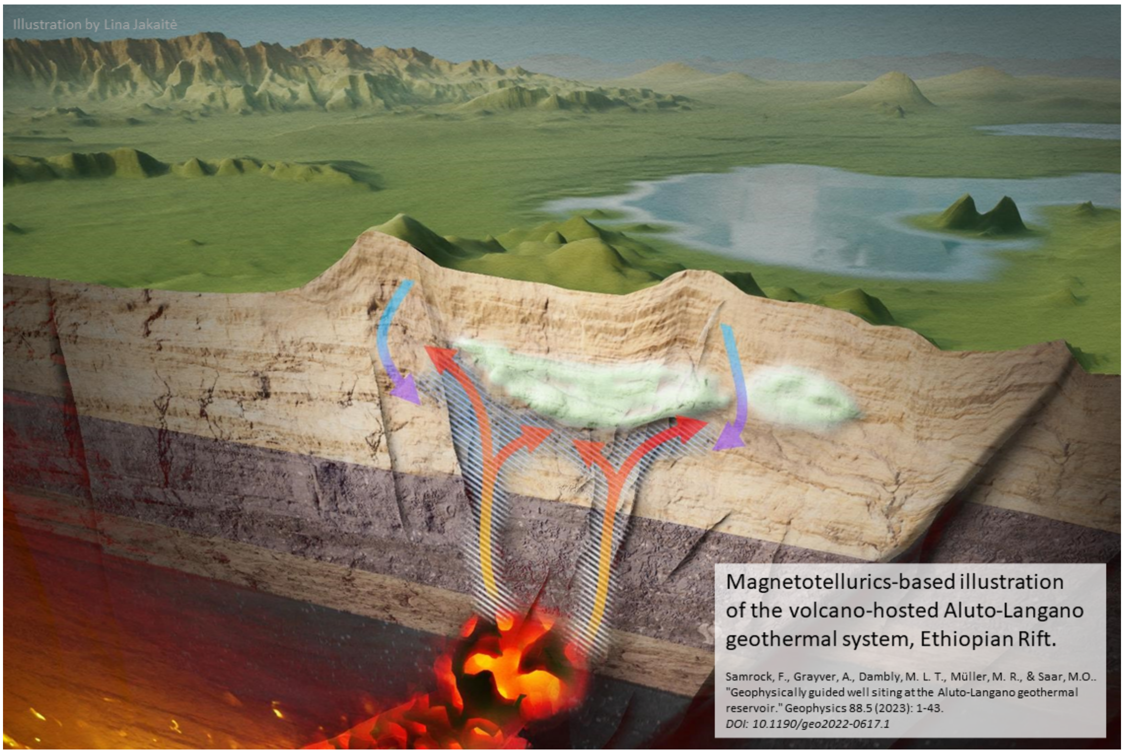

Figure 2-3: MT-based illustration of the volcano-hosted Aluto-Langano geothermal system, Ethiopian Rift. (Samrock et al., 2023).

Geothermal fluids show typically a high conductivity, which is due to their high salt content (Munoz, 2014). Fluid-, as well as the matrix conductivity, is temperature-dependent. The *clay cap* shown in Figure 2-4 changes its mineral content when the temperature exceeds 180 °C, which is due to the increasing replacement of smectite with illite. This process only happens when the temperature increases and is not reversible when the system cools down again (Munoz, 2014). Therefore, the alteration of the clay cap can also be considered as a *maximum thermometer*, showing the hottest temperature present in that region. This phenomena gives one reason for the importance of improving our understanding in clay caps as it gives information about the present or maximum temperature. The *upflow zone* defines the region in which heat is transported upwards, via water or vapour. When it reaches the nearly impermeable clay cap, the medium is redirected in a more horizontal flow behavior, described by the term *outflow*. Hydrothermal fluids are often mixed with meteoric water when approaching the surface, and hardly recirculate back in larger depths (Peacock et al., 2016). The isotherms included in Figure 2-4 are of importance when planning the drilling location (Jolie et al., 2021). The aim is to reach high temperatures in shallow depths, in order to maximise the economical outcome while keeping costs low. In case of very high-enthalpy regions as present in, e.g., Iceland, special equipment is required that can withstand these high temperatures (Song et al., 2022).

2-1-3 Geophysical Background

Geophysical methods response to different properties of the subsurface, such as to the water content, temperature variations or density contrasts.

In order to find the most suitable technique to extract heat from a geothermal reservoir, it is crucial to be aware of the relevant physical parameters of the area of interest beforehand. Temperature, pressure, porosity, permeability and the chemical content of the fluid (salinity) are usually the five main physical properties of interest for geothermal applications (Kana et al., 2015). Besides these factors, there are other parameters that increase ones knowledge about geothermal systems and that are therefore useful to explore, such as: fluid flow rate, heat flow, mineralogy, saturation, the density of the host rock and the fracture network. By defining these properties, we can outline bodies of geothermal systems like the heat source, the upflow zone and the clay cap in a more accurate way, to locate aquifers in that area and to decide for suitable drilling locations. As each system has its own shape and extent, its detailed layout needs to be investigated individually, in order to gain knowledge about the systems quality, size and depth (Kana et al., 2015). The whole project may not be economically feasible if the expenses required to access the heat outweigh the financial gains of production due to low permeability or temperature, as well as a small reservoir volume. These properties can be investigated with geophysical methods in an efficient way with relatively low costs (Aretouyap et al., 2016).

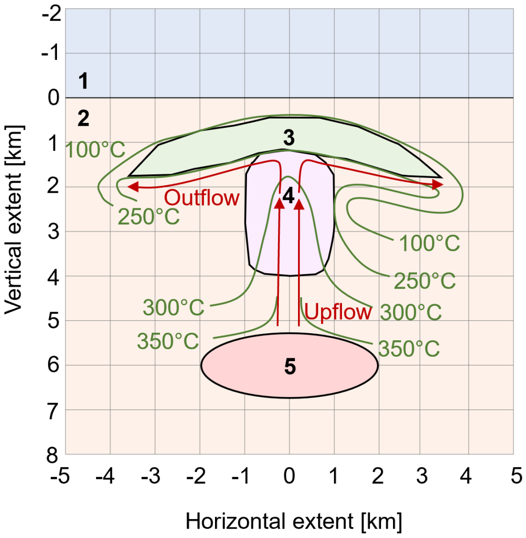

Figure 2-4: Simplified magmatic geothermal system including typical isotherms, upflow and outflow regions and possible depth ranges (the depth of the heat source normally varies between 3 to 6 km) (modified after (Samrock et al., 2018)). Blue (1): air layer, light orange (2): surrounding rock, green (3): clay cap, pink (4): upflow zone, red (5): heat source.

In order to provide an overview about possible geophysical measurements responding to different target parameters, we introduce here shortly four geophysical methods that can be applied to investigate geothermal systems. **Seismic** measurements are defined as indirect methods, as they give information about the geological structure of a geothermal systems (Kana et al., 2015). In this method the acoustic wave velocity distribution and their attenuation are measured, allowing us to see anomalies of the Earth. It enables the detection of geological layers, discontinuities, porosity, fluid-filled zones and in this way even temperature contrast, which is of interest in geothermal applications (Kana et al., 2015). **Electromagnetics** rely on the large contrast in electrical conductivity present in geothermal systems, showing generally higher electrical conductivities than the surrounding host rock (Munoz, 2014). MT is the only EM method allowing a penetration depth of several kilometers *and* being sensitive to fluids, which is of advantage for the investigation of deep geothermal systems (Munoz, 2014; Aretouyap et al., 2016) (more details about MT can be found in Section 2-2). **Gravimetrics** measures the changes in gravity depending on the variations in density in the subsurface. As, e.g., fluid-rich rocks of high porosity show lower densities as rocks with only few pores, these changes in density can give insights into the geothermal potential of the subsurface (Kana et al., 2015). **Direct current** methods rely on the induction of an electrical current into the subsurface, from which an electrical potential field develops, measurable with electrodes at the surface (Kana et al., 2015). As a result we obtain the resistivity distribution, which describes the geometry of a geothermal system.

The focus on this study lies on the simulation of MT, due to its previously mentioned advantages. For this reason, the next Section will introduce MT in more detail, and give a theoretical background.

2-2 Magnetotellurics

MT is a geophysical method used to study the distribution of the electrical conductivity of the subsurface. It is a passive EM method, using the Earth's E and H fields as a natural source. Its theory is based on Maxwell's equation, describing the propagation of these fields, but thus also including certain limitations that need to be considered. This section focuses on the theory behind MT, including the case of 1D and 2D, and differentiating between the transverse electric (TE) and transverse magnetic (TM) mode. Possible MT applications are explained.

2-2-1 Theory

The Earth's H field is described as

$$\vec{H} = \vec{B}{\mu_0}^{-1}, \tag{2-1}$$

where $\vec{H}$ is the magnetic field intensity [$\mathrm{A\,m^{-1}}$], $\vec{B}$ the magnetic flux density [$\mathrm{V\,s\,m^{-2}}$], and μ_0 is the magnetic permeability [$4\pi \cdot 10^{-7}\,\mathrm{V\,s\,A\,m^{-1}}$] (Kuvshinov et al., 2019).

The Earth's E field stands for the electric field intensity $\vec{E}$ [$\mathrm{Vm^{-1}}$], described as

$$\vec{E} = \vec{D}{\epsilon_0}^{-1}, \tag{2-2}$$

where $\vec{D}$ is the the electric flux density [$\mathrm{A\,s\,m^{-2}}$] and ϵ_0 the electric permittivity [$\approx 8.8541 \cdot 10^{-12}\mathrm{A\,s\,V^{-1}\,m^{-1}}$] (Kuvshinov et al., 2019). The amplitude, phase and directional relation between these E and H fields are dependent on the electrical conductivity (Vozoff, 1991). Therefore, the response of these fields given by subsurface structures allows the visualization of the distribution of the electrical conductivity.

The spatio-temporal behavior of EM fields can generally be described by the Maxwell's equations (Eq. (2-3)-(2-6)), while the frequency-domain is obtained via the Fourier transform (Eq. (2-7)-(2-10)),

$$\nabla \times \vec{H} = \vec{j}, \tag{2-3}$$

$$\nabla \times \vec{H} = (\sigma - \mathrm{i}\omega\varepsilon_0)\,\vec{E} + \vec{j}^{cxt}, \tag{2-7}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{2-4}$$

$$\nabla \times \vec{E} = \mathrm{i}\omega\mu_0\vec{H}, \tag{2-8}$$

$$\nabla \cdot \vec{B} = 0, \tag{2-5}$$

$$\nabla \cdot \vec{H} = 0, \tag{2-9}$$

$$\nabla \cdot \vec{D} = q, \tag{2-6}$$

$$\nabla \cdot \vec{E} = \frac{q}{\varepsilon_0} \tag{2-10}$$

where $\vec{j}$ corresponds to the electric current density [$\mathrm{A\,m^{-2}}$], ∂_t to the temporal partial differentiation, q to the electric charge density [$\mathrm{A\,s\,m^{-3}}$], and $\omega = \frac{2\pi}{T}$ to the angular frequency [rad $\mathrm{s^{-1}}$] (Kuvshinov et al., 2019). Ampère's law (Eq. (2-3)) describes the generation of magnetic fields by currents, and Faraday's law in Eq. (2-4) introduces the induction of electric fields by variation over time of magnetic fields. Gauss' law for magnetism (Eq. (2-5)) describes the magnetic field as *solenoidal*, meaning that $\vec{B}$ has no source or sink and no magnetic monopoly exist. This does not hold for electric fields, which are described in the Gauss' law for electric fields in Eq. (2-6), including the parameter q describing that electric field lines have a start and end at charges.

The parameters influencing EM fields can be included in the Maxwell equations as

$$-\nabla \times \vec{H} + (\sigma + \varepsilon\partial_t)\,\vec{E} = -\vec{J}^e, \tag{2-11}$$

$$\nabla \times \vec{E} + \mu_0\partial_t\vec{H} = -\vec{J}^m, \tag{2-12}$$

with the electrical conductivity σ [S $\mathrm{m^{-1}}$], the electric permittivity ε [F $\mathrm{m^{-1}}$] and the temporal partial differentiation ∂_t (Slob et al., 2007). The properties $\vec{J}^e$ and $\vec{J}^m$ stand for the source volume densities of the electric and magnetic current, respectively. Eq. (2-11) and Eq. (2-12) describe the electric and magnetic fields assuming a non-vacuum environment and an immediate response of the Earth to electromagnetic variations, while including material parameters of the subsurface. From these rewritten Maxwell equations, we can derive the propagation of EM fields in a medium over space and time by the so-called Helmholtz differential equations, which state for E field 1D that

$$\frac{\partial^2 \vec{E}}{\partial z^2} + k^2\vec{E} = 0, \tag{2-13}$$

with

$$k^2 = \underbrace{\omega^2 \epsilon_0 \mu_0}_{(1)} - \underbrace{i\omega\mu_0\sigma}_{(2)}, \tag{2-14}$$

with the frequency ω [rad s^{-1}] and the wavenumber k [m^{-1}] (Basu and Dhasmana, 2022; Kuvshinov et al., 2019). The ratio of the two terms (1) and (2) in Eq. (2-14) influence the regime of propagating EM fields. In case of (1) >> (2), Eq. (2-14) simplifies to a wave equation depending on the electrical permittivity which can be present in the application of, e.g. a ground penetrating radar. If (1) << (2) the process of diffusion

$$\nabla^2(\vec{E}, \vec{H}) - i\omega\mu_0\sigma\ (\vec{E}, \vec{H}) = 0, \tag{2-15}$$

is dominating, strongly depending on the electrical conductivity σ. This diffusion regime of Eq. (2-15) describes MT, as it yields electromagnetic induction (Chave and Jones, 2012).

The penetration depth of E and H fields, also called *skin depth*, is dependent on the frequency and the electrical conductivity. It is defined as the distance over which electromagnetic field decays by e^{-1} $\approx$ 0.37, and can be specified as

$$\delta(\omega) = \sqrt{\frac{2}{\omega\mu_0\sigma}} \approx 503\sqrt{\rho T} = 503\sqrt{\frac{1}{\sigma f}}, \tag{2-16}$$

with the skin depth δ [m], the electrical resistivity ρ [Ω m] being the inverse of the electrical conductivity, and period T [s] the inverse of frequency f [s^{-1}](Chave and Jones, 2012). It is to note that this Eq. (2-16) only holds for a plane-wave source in a homogeneous medium. Hence, deep structures in the subsurface with low electrical conductivity can be investigated with a low frequency signal.

The linear relationship between the magnetic and electric field can be described by the transfer function, stating that

$$\vec{E} = \vec{Z}\vec{H}, \tag{2-17}$$

where the impedance tensor Z relates E and H, containing information about the unknown parameters of the subsurface (Chave and Jones, 2012). Z is complex-valued and the tensors shape reveals details about the dimensions of the underground. This impedance is usually the goal of EM measurements, but due to its complex values, it is obtained indirectly via the phase ϕ [rad] and the apparent resistivity ρ [Ω m]. Also Eq. (2-17) only holds for a plane-wave source in a homogeneous medium. The relation between the impedance and resistivity and phase, respectively, can be shown as (Kuvshinov et al., 2019):

$$\rho(\omega) = \frac{|Z|^2}{\omega\mu_0}, \tag{2-18}$$

$$Z = \operatorname{Re} Z + i \operatorname{Im} Z = |Z|(\cos\phi + i\sin\phi). \tag{2-19}$$

When replacing $|Z|$ in Eq. (2-19) with its definition from Eq. (2-18), we obtain

$$Z = \sqrt{\rho\omega\mu_0}(\cos\phi + i\sin\phi), \tag{2-20}$$

which is showing how the impedance can be calculated from resistivity ρ and phase ϕ.

In the case of **1D**, representing a layered Earth, we define the z-axis to point downwards. A plane wave assumption leads to excluding vertical components of both E and H fields and to assume constant E and H fields along any horizontal plane of the x- and y-axis (Kuvshinov et al., 2019). Additionally, we define both the E and H field to go to zero when z tends to infinity. These assumptions in the 1D case with horizontal layering lead to the following simplification of the Maxwell equations in the frequency-domain (Kuvshinov et al., 2019):

$$\nabla \times \vec{H} = \sigma\vec{E}, \tag{2-21}$$

$$\nabla \cdot \vec{H} = 0, \tag{2-22}$$

$$\nabla \times \vec{E} = i\omega\mu_0\vec{H}, \tag{2-23}$$

$$\nabla \cdot \vec{E} = 0. \tag{2-24}$$

Thus, the Helmholtz equation defined in (2-13) can be rewritten for the E field as the following,

$$\frac{\partial^2\vec{E}}{\partial z^2} + k^2\vec{E} = 0, \tag{2-25}$$

with the reminder of the propagation constant $k = \sqrt{i\omega\mu_0\sigma}$ in the case of diffusion (Kuvshinov et al., 2019).

The general solution of (2-25) is then

$$\vec{E}(z) = \vec{E}^{+}e^{ikz} + \vec{E}^{-}e^{-ikz}, \tag{2-26}$$

with $\vec{E}^{+}$ and $\vec{E}^{-}$ being the vector constants. For a uniform 1D Earth we can then conclude the solution to be

$$\vec{E}_{\mathrm{x}}(z) = \vec{E}_{\mathrm{x}}{}^{+}e^{ikz}, \tag{2-27}$$

$$\vec{E}_{\mathrm{y}}(z) = \vec{E}_{\mathrm{y}}{}^{+}e^{ikz} \tag{2-28}$$

for the propagation of the E field over depth (Kuvshinov et al., 2019).

In the case of a **2D** Earth, variations of the electrical conductivity happen now in horizontal direction of the y-axis in addition to the depth (z-axis) (Kuvshinov et al., 2019). In comparison to the 1D case previously described, the E and H fields are not constant anymore in the horizontal plane (spanned by x and y), but only in the x-direction, as they vary along the y-axis. This constant x-axis is also called the *strike direction* (Kuvshinov et al., 2019). A schematic overview of the electric conductivity model including its TE and TM modes is visible in Figure 2-5.

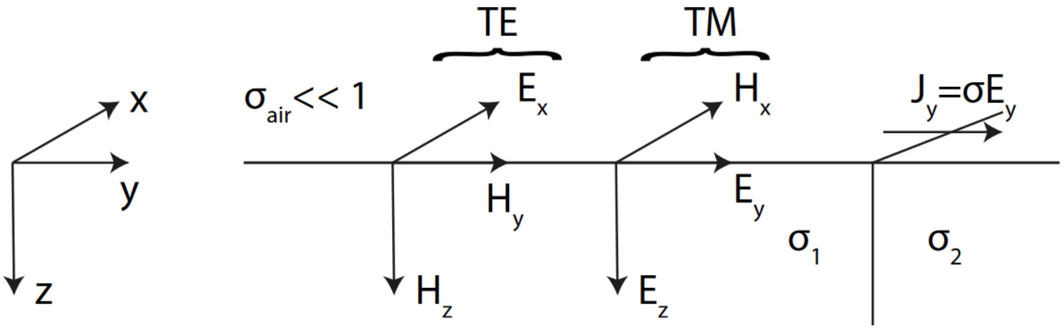

Figure 2-5: Electrical conductivity model and EM modes of TE and TM for a 2D Earth (Kuvshinov et al., 2019).

Rewriting the Ampère's law and Faraday's law of the Maxwell equations in the frequency domain (Eq. (2-7) and Eq. (2-8)) needs to be done component-wise for the TE and TM mode (which will be explained in more detail in chapter 2-2-2), assuming that displacement currents can be neglected (Kuvshinov et al., 2019). These rewritten equations look as the following:

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \sigma E_x \quad \text{(TE)}, \tag{2-29}$$

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = \mathrm{i}\omega\mu_0 H_x \quad \text{(TM)}, \tag{2-32}$$

$$\frac{\partial H_x}{\partial z} = \sigma E_y \quad \text{(TM)}, \tag{2-30}$$

$$\frac{\partial E_x}{\partial z} = \mathrm{i}\omega\mu_0 H_y \quad \text{(TE)}, \tag{2-33}$$

$$-\frac{\partial H_x}{\partial y} = \sigma E_z \quad \text{(TM)}, \tag{2-31}$$

$$-\frac{\partial E_x}{\partial y} = \mathrm{i}\omega\mu_0 H_z \quad \text{(TE)} \; . \tag{2-34}$$

Three of these equations define the TE and TM mode, respectively. Additionally, it is to note that both the strike electric field $\mathrm{E_x}$ and magnetic field $\mathrm{H_x}$ are involved (Kuvshinov et al., 2019). Thus, the two fully decoupled Helmholtz equations can be defined as

$$\nabla^2 E_x + k^2 E_x = 0 \; \text{(TE)} \; , \tag{2-35}$$

$$\nabla\dot{(}(\sigma \nabla H_x) + \mathrm{i}\omega\mu_0 H_x = 0 \quad \text{(TM)}, \tag{2-36}$$

together describing the EM plane wave propagation in a 2D case (Kuvshinov et al., 2019). It is to note that in this 2D case, no analytical solution of equations 2-35 and 2-36 can be derived, while this is possible for the case of 1D. Thus, in a 2D case, numerical methods are required.

2-2-2 Transverse Electric and Magnetic Mode

The EM field can be divided in two different main wave types: the TE and TM mode. These modes relate the orientation of the E and H fields of the EM waves with the direction of propagation (Chave and Jones, 2012). In EM methods, these two wave types must be analysed separately, as they respond differently to structures and ranges of electrical conductivities in the subsurface, thus complementing each other. In the following paragraphs, both modes are explained in more detail for 2D applications only. As a comparison of their response for the same subsurface, Figure 2-6 can be kept in mind (Berdichevsky et al., 1998).

The TE mode is related to the telluric current flowing along structures due to the connection to E-polarized waves generating transverse MT curves, following the strike direction (Berdichevsky et al., 1998; Gómez-Treviño et al., 2018). It does not charge structures, as its anomalies are of inductive nature (Berdichevsky et al., 1998). This mode is more sensitive to deep structures, and more robust to the 3D effects caused by resistive structures. The TE mode is not susceptible to the static shift which occurs in EM waves. When focusing on Figure 2-6, we see that the TE mode is much more insensitive to the anomaly present around 1 km depth, as it is hard to detect, especially in larger periods (Berdichevsky et al., 1998). However, assuming the same anomaly in bigger depths of over 100 km, it would more dominant in the TE mode than in the TM mode. During analysis, the frequency-axis can be assumed to be corresponding to depth, where smaller frequencies represent larger depths and vice versa (Chave and Jones, 2012). Additionally, the phase ϕ is shifted relatively between the TE and TM mode by $\frac{\pi}{2}$, which is due to the orthogonality of these two modes relative to each other (Chave and Jones, 2012). This phase ϕ measures the time lag between the E and the H field.

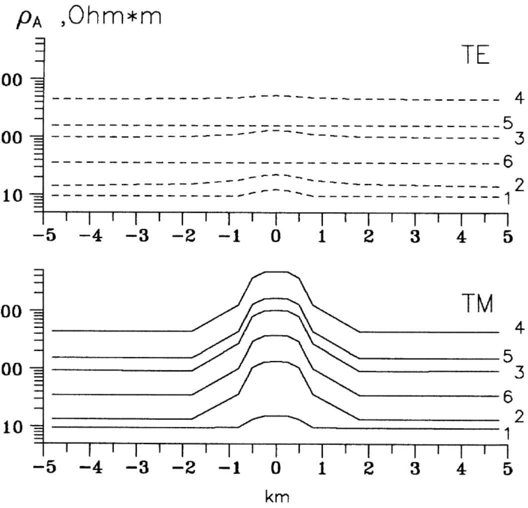

Figure 2-6: Model of 2D horst-like resistive elevation in a sedimentary cover, shown by apparent resistivity profiles for TE and TM modes, for the same subsurface. Curve 1 has a period T of 0.1 s, curve 2 of 1 s, curve 3 of 10 s, curve 4 of 100 s, and curve 5 of 1000 s s. Figure obtained from Berdichevsky et al. (1998).

The TM mode on the other side follows the telluric current across structures, as it is related to the H-polarized wave generating the transverse MT curves (Berdichevsky et al., 1998). Hence it is flowing across the strike direction (Gómez-Treviño et al., 2018). In contrast to the TE mode, the TM mode charges structures and its anomalies are of galvanic nature (Berdichevsky et al., 1998). It is more sensitive to structures closer to the surface and it is more robust to 3D effects caused by conductive structures. A downside is its strong influence to static shift. Looking at Figure 2-6, we want to draw attention to the clear representation of the anomaly in the lower surface (Berdichevsky et al., 1998). However, the static drift is more dominant in shallower depths, for which reason it needs to be considered in the TM response. Due to the static shift and the galvanic effect present in this mode, depth estimates via frequencies are not possible in the TM-mode (Chave and Jones, 2012). These effects can be neglected in the TE mode, however only in the case of 2D (Becken et al., 2008)

2-2-3 Limitations and Benefits

As a general **limitation**, the theory underlying MT described in chapter 2-2-1 follows some assumptions besides being governed by the Maxwell Equations (Eq. 2-3 to 2-6), that one should be aware of:

- MT is based on the plane wave EM field propagation assumption (Kuvshinov et al., 2019). This leads to two advantages: first, it eases the calculation of analytical solutions of a plane wave propagation in layered media, and second it represents well a natural EM field in MT. However, this assumption is violated in certain geographic locations as in equatorial and auroral electrojet zones, where concentrated sources in the ionosphere invalidate the wave being propagated in a planar form (Hermance and Peltier, 1970).
- The current density sources in the subsurface are assumed to be zero (Kuvshinov et al., 2019). The EM fields are induced by external variations, and produced and dissipated inside the Earth.
- The magnetic permeability μ_0 and the electric permittivity ϵ_0 in deep EM induction studies are assumed to be constant over time and space and equal to their values in free space(Kuvshinov et al., 2019).
- We follow the *quasi-static approximation*, which means that a neglect of displacement currents is valid since $\sigma >> \omega\epsilon_0$ (Kuvshinov et al., 2019).

Besides these limitation in the theoretic foundation of EM, noise occurring in EM fields influences the data quality. This noise is often divided in natural and human-made noise. Natural sources as radio noise vary over a large frequency range from 1 mHz to 300 GHz (Bianchi and Meloni, 2007). Events of low frequencies are slow variations in the Earth's magnetic field, but also geomagnetic pulsations of higher frequencies may affect EM measurements. Sudden atmospheric events such as lightning can produce large spikes in the data, but also changes in air temperature can influence measurements in shallow depths. Human-made EM noise is usually caused by human technologies, such as power lines, radio or TV communication installations (Bianchi and Meloni, 2007). The magnitude of the noise is dependent on the distance to this source, on the sources frequency and power. Experiments close to streets tend to have a higher noise source than ones in rural areas, as cars additionally cause EM and seismic noise. As technological innovations have increased and advanced in the last decades, this source strength is overall growing.

Another drawback of MT measurements are possible distortions of E and H fields, leading to a decrease in data quality. The galvanic distortion is one example for such a phenomena. It is caused by electrical charges being present at discontinuities or small-scale electrical conductivity gradients (Chave and Smith, 1994). These 'unexpected' charges have an influence on the E field, which also changes the H field due to its coupling. Galvanic distortions can be described as the superposition of frequency independent factors caused by these local property changes in the subsurface on the frequency dependent desired signal (Bibby et al., 2005). The removal of this effect is problematic as galvanic distortions operating on the impedance is underdetermined. It requires assumptions which might be arguable depending on the experiment it is applied to. A second distortion which might cause interpretation errors is the topographic distortion. Rough changes in elevation levels along an experiment require corrections on the impedance, as the density of telluric currents change depending on the topography (Vallianatos et al., 2002). Due to the increase in elevation, these currents tend to diverge and converge below hills and valleys, respectively. Topographic distortions might be either caused by galvanic or induced effects, where research shows that the former is mainly dominating at periods above 10 s and the latter at shorter periods below 10 s (Käufl et al., 2020).

As a **benefit**, MT allows a penetration depth of many kilometers (ranging from 50 m to 500 km) due to the use of the natural low-frequency range variations between 10^{-4} and 10^{4} Hz with the Earth's magnetic field as a source (Chave and Jones, 2012; Munoz, 2014; Siniscalchi et al., 2021). Such a depth can be also reached by seismics, but MT is highly sensitive to the presence of fluids (Peacock et al., 2012), which is of advantage for investigations of geothermal systems. In such reservoirs, conductivity varies over seven orders of magnitude, ranging from 0.01 to $10{\cdot}10^{4}$ mS m^{-1}. The application of MT has been proven to be successful in both magmatic (Heise et al., 2016; Bertrand et al., 2013; Hill et al., 2015) and non-magmatic geothermal systems (Munoz, 2014; Peacock et al., 2012, 2013).

Variations in electrical conductivity allow to estimate the reservoirs size and shape (Ishizu et al., 2021; Newman et al., 2008; Gao et al., 2018), its porosity and permeability (Munoz, 2014; Ussher et al., 2000) as well as to approximate the temperature range present (Azeez and Harinarayana, 2007; Harinarayana et al., 2006; Spichak et al., 2015). E.g., an increase in electrical conductivity below the conductive clay cap reflects an increase in temperature with depth, which often corresponds to a high-temperature geothermal system. This is due to the alteration of clay-minerals. In regions of temperatures below 150 °C smectite-illite alterations are common, which have a larger electrical conductivity due to their loosely bounded cations. At higher temperatures, chlorite-epidote species replace the smectite-illite composition, having a more stable crystal lattice and therefore leading to a lower conductivity. MT can also observe the presence of melt, and magmatic and meteoric brine in the lower regions of geothermal systems (Samrock et al., 2021). In non-magmatic geothermal reservoirs the porosity and permeability can be imaged with EM. High conductivity values normally correspond to saline fluids, which can help to identify fluid paths in the subsurface.(Munoz, 2014). However, it should be kept in mind that geophysical results are ambiguous, as different features can lead to the same conductivity anomaly. Therefore, cross-correlations with other geophysical methods can help to solve this non-uniqueness (Miller et al., 2022).

Besides for geothermal reservoirs, is MT widely applicable in many other different fields focusing on investigations of the subsurface. MT can be applied in hydrocarbon exploration, where 3D models can be used to estimate the volume of hydrocarbon in larger depths, trying to fit well log data (Hoversten et al., 2015; Mansoori et al., 2016; Azeez et al., 2011). A further area of application is groundwater estimation, where aquifers can be found in depths of several kilometers, underlayed by aquicludes (Asaue et al., 2012). Additionally, flowpaths of high conductivities can be detected. Also, MT can also be used to analyse the static water level in shallower depth of 25 m (Agyemang, 2022). Furthermore, MT can improve our knowledge about the upper mantle, thanks to its large penetration depth. It allows the estimation of melt composition, convection movement and thermal structures (Baba et al., 2010; Selway et al., 2019). MT also allows further insights into planetary exploration, as it has been applied on, e.g., Jupyter's moon Europa and on Mars for the detection of water (Grimm et al., 2021; Delory et al., 2007).

2-3 Simulations

Simulations are obtained by GoFEM, which allows efficient simulations using finite element (FE). As sampling strategy for the material properties in the simulations, we use latin hypercube sampling (LHS). The following paragraphs explain these elements in more detail.

GoFEM (Grayver and Kolev, 2015) is used to generate a mesh of a generalized, simplified, high-temperature geothermal system, and to run simulations of the propagation of EM fields through the predefined mesh. GoFEM represents a Python-environment enabling 3D geoelectromagnetic modelling using a parallel adaptive high-order FE method (Grayver and Kolev, 2015). FEs are areas or volumes of systems, which are interconnected with nodal points of elements (Bathe, 2007). In its methods, we solve partial differential equations (PDEs) numerically in the discretized manner of using these FEs, allowing to tackle problems too complex to be solved with analytical methods. FE methods are nowadays widely used, as they allow scientists and engineers to model and solve such complex problems, in a mathematical and numerical way. There are two different main approaches that can be used to define the variability of a system, the so-called degrees of freedom (DoF): (1) we define it via the number of nodes and the number of variables to solve, or (2) we include the number of elements instead of nodes (Larson and Bengzon, 2013). GoFEM takes advantage of the FE method, while using the element-wise approach. Furthermore, it is based on a distributed robust and scalable linear solver, using optimal block-diagonal and auxiliary space preconditioners. Local mesh refinements allow the reduction of computational costs while maintaining a high accuracy. It enables EM modelling while having high FE orders, a wide range of frequencies and conductivities, as well as a large number of DoF and unstructured and nonconforming locally refined meshes (Grayver and Kolev, 2015). Besides allowing the simulations of electromagnetic currents in the subsurface with varying properties, it also provides mesh generation. In this project, this is done for the electrical conductivity. A non-conforming hexahedral mesh with a locally finer grid at boundaries is defined. Figure 2-7 shows the GoFEM mesh generated in this project. It represents a simplified magmatic geothermal system consisting of an air layer above an elevation of 0 m, a surrounding rock below an elevation of 0 m, a clay cap, an upflow zone, as well as a heat source.

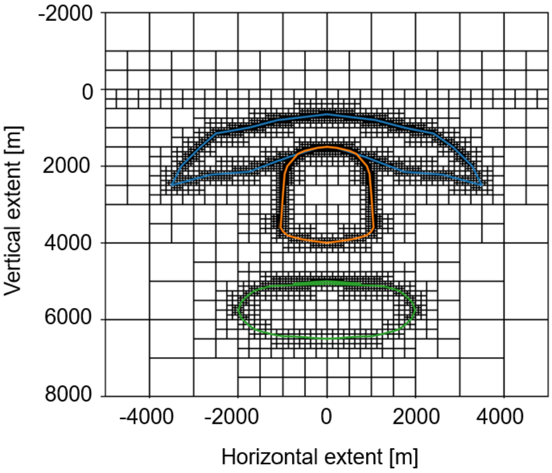

Figure 2-7: Mesh of a simplified magmatic geothermal created in GoFEM with a hexahedral configuration. Blue: clay cap, orange: upflow zone, green: heat source.

In order to ensure reliable and efficient GoFEM simulations by generating random variables for a problem, the statistical sampling method LHS is chosen (Huntington and Lyrintzis, 1998). After defining the number of parameters p and the number of random variables n that are needed for a specific problem, a $n \cdot p$ matrix is generated. This matrix preserves the marginal probability distribution of each parameter while being based on joint probability density functions (Huntington and Lyrintzis, 1998), containing values between 0 and 1. A advantage using LHS is the representation of each parameter in a fully stratified manner (McKay et al., 2000). Especially in high-dimensional systems, LHS gives a significant computational benefit, when comparing it to, e.g., standard Monte Carlo (Manteufel, 2000). Overall, it ensures a distribution of samples over the given parameter space with a efficient sampling strategy (Olsson et al., 2003), which is important for modelling applications. Figure 2-8 shows a scatterplot of a $10 \cdot 2$ matrix as an example, where 10 variables are defined cia LHS for the two parameters X and Y.

After the creation of the matrix and running the simulations, the obtained values are scaled to be merged into the desired range. The applied scaling method for this research is the min/max scaling, which defines new ranges as

$$x' = x(max(x) - min(x)) + min(x), \tag{2-37}$$

where *x'* is the rescaled value of a parameter, *x* the original value, and *min(x)* and *max(x)* the minimum and maximum value of the specific simulation values, respectively. This ensures the values to be between zero and one, which is of advantage for specific calculations and functions. The material properties themselves are standardized by subtracting their mean and dividing by the standard deviation.

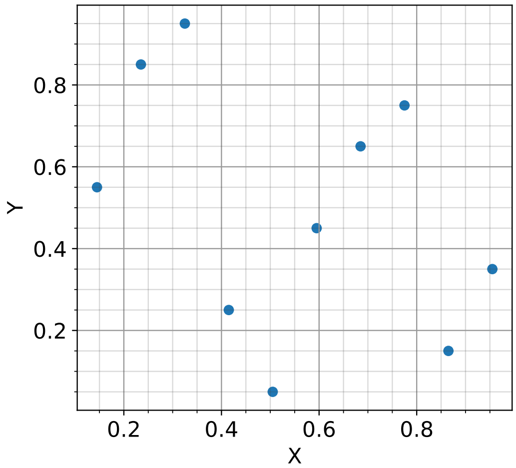

Figure 2-8: Scatterplot of a LHS with two parameters X and Y and 10 samples between 0 and 1.

As previously mentioned allows GoFEM, besides others, the mesh generation of subsurface structures (Grayver and Kolev, 2015). However, its main goal are simulations of electrical currents in the underground when having varying properties present, such as the electrical conductivity. The applied linear solver is algebraic, acting on the matrix-vector level. Therefore, information about boundary conditions (BCs), the discretization, and explicitly about the source strength are not required. This is of advantage in MT applications where the source magnitude is unknown. It is to note that for physically logical results, the implementations of, e.g., Neumann or Dirichlet BCs should be considered in the governing PDEs. The combination of a high-order finite element method (FEM) with a goal-oriented mesh refinement allows a satisfying accuracy with relatively little computational time and only few DoFs. GoFEM includes error indicators for each cell (Grayver and Kolev, 2015). This allows the identification of the most inaccurate solution, which can then be redefined until the desired accuracy is achieved. The focus is primarily placed on the receiver location, as increasing the precision at these points significantly enhances the results. GoFEM as a scalable and robust solver has an sufficient stability to solve the lowest order Nédélec elements. The robustness of this solver allows the preservation of a fast convergence, also for the case of large variations in model parameters and many unknowns being present, which is often the case in EM. As GoFEM is scalable, it can reduce the computational time proportionally to the number of processes with the use of parallelisation. The impedance is often the main parameter of interest in MT modelling, also in this project. This complex-valued parameter can, besides others, be calculated in GoFEM simulations for the MT response (Grayver and Kolev, 2015). Since the impedance is influenced by anisotropy of the subsurface (Martí, 2014), its measurements usually require two orthogonal source polarizations. This makes computations more expensive, which could be reduced in GoFEM by using secondary field formulations and homogeneous BCs, making the system matrix for the polarization equal to the one for the dual problems. In the application of MT, the combination of quadratic elements with adaptive local mesh refinement is a widely used and efficient solution, which additionally keeps the computational costs low.

2-4 Physics-Based Machine Learning

PB-ML describe the combination physics-based and data-driven ML approaches. It takes advantage of both techniques and approaches in order to overcome their individual limitations. The following sections give a background about this approach and compare two different methods: physical-informed neural networks (PINNs) and the NI-RB method. For a more extensive background of artificial intelligence (AI) and more specifically of ML, further information can be found in the appendix B-1-1 and B-1-2.

2-4-1 Background to Physics-Based Machine Learning

In PB-ML, physical laws are used as prior knowledge in the form of, e.g., PDEs, such as conservation of energy, mass or momentum. They have the goal to limit the output to a physically feasible range and to minimize the DoFs of ML approaches (Vadyala et al., 2022). This prior knowledge allows for a faster convergence of simulations and more accurate physical predictions (Vadyala et al., 2022). The implementation of physical knowledge aims to have a low dimensional representation of high-fidelity simulations which can then be used as a training data set for ML techniques (Wang et al., 2022a). These data sets train systems to solve desired models based on the previously defined PDEs (Vadyala et al., 2022). In the case of PB-ML, physics-based models are used to decrease the size of the black box, as more knowledge about the input and the function connecting the input to the output is given. The output is additionally constrained to a specific range, as they need to follow physical laws. Without this prior physics-based knowledge, this purely data-driven ML approach would require much larger data sets, which are rare in geoscientific and engineering applications (Swischuk et al., 2019). More details about data-driven ML approaches and physics-based models can be found in the appendix B-1-3 and B-3, respectively. The use of ML enables automatised simulations and inversions of data sets, as the computer is required to learn from the data provided (Willcox et al., 2021). It allows the handling of multi-dimensional and complex data sets more efficiently than purely physics-based model would enable (Lattimer et al., 2020). Therefore, PB-ML merges knowledge gained from data with the awareness obtained from physics-based models to increase the accuracy of the results while keeping the required amount of data low (Willcox et al., 2021).

By combining physics-based models and data-driven ML, the advantages of both these techniques are combined. The interpretability and transparency obtained from physics-based models is merged with the high computational speed and learning rate of the ML approach (Wang et al., 2022a), leading to a very efficient modelling technique. It increases the overall accuracy and ability to solve non-linear models by ML, while keeping physical laws consistent throughout the modelling by physics-based models. We therefore can preserve the main physical character, which is necessary for predictions and uncertainty quantification (UQ) (Degen et al., 2022).

PB-ML can be applied in supervised (Stern et al., 2021), unsupervised (Lattimer et al., 2020) and reinforcement learning (Peng et al., 2018) to predict and understand physical systems, guided by physical equations and laws. In this project, PB-ML is implemented within a general ML framework. Physical parameters such as density and electrical conductivity are used as labeled input to characterize phenomena such as fluid and electrical flows, which are fundamentally governed by the laws of physics. These PB-ML models have a wide applications range in many research areas, such as data storage, fluid dynamics, quantum mechanics and computational resources (Vadyala et al., 2022).

2-4-2 Physics-informed Neural Networks

PINNs are a type of PB-ML that combines physics and data-driven techniques (Degen et al., 2022) to minimize the residual loss of PDEs (Chen et al., 2021). PINNs are NNs that implement physics as a constrain via PDEs, BCs and initial conditions. NNs are trained by giving specific input-output-pairs, forcing a specific relationship given by the physical laws. It aims to minimize the norm, also called the *loss function*, of the residuals in the reduced order equation (Chen et al., 2021). These underlying physical laws are included as an additional term in this loss function of the NN during the training process, in order to approximate solutions that meet these physical constrains (Raissi et al., 2017; Kharazmi et al., 2019).

In general, NNs are inspired by the function and connectivity of the human brain and its nervous system (Gupta et al., 2013). It allows the processing of bigger data sets by the application of a large number of nodes (so-called *neurons*), each with its own output function (Wu and Feng, 2018). The nodes are connected through weighted connections, allowing the user to influence the system depending on the connection strength between cells. The black box is created by the presence of hidden nodes, which can lead to a problematic overfitting of the data (Liu et al., 2008).

The advantage of this method is its high flexibility and adaptability in function approximations, leading to a wide range of problems that can be tackled with PINNs. It can solve over-determined systems (Cai et al., 2021), which are a problem in NN without a physical background (Liu et al., 2008). However, there are specific problems in the use of PINNs. The presence of hidden layers increase the size of the black box, leading to processes which are challenging to fully understand. A user can influence NNs by adjusting so-called hyperparameters such as the range of layers or neurons desired, but the detailed workflow inside the NN is unknown. Furthermore, problems have occurred in simulations of highly non-linear problems (Chen et al., 2021) and computational fluid dynamics (Degen et al., 2022), like the Navier-Stokes equations using PINNs as a solver (Chuang and Barba, 2022). Also, a discrepancy in the convergence rate of PINNs loss functions has been detected (Wang et al., 2022b). A main restriction is that physics is simply treated as data or as constrain inside the BCs or initial conditions of the loss function (Degen et al., 2022). This is handled differently in, e.g., NI-RB methods, where the physics is implemented in the snapshots, as explained in more detail in the next Section 2-4-3. The physical restriction of PINNs causes the solution not to represent the original physical problem anymore. Additionally, modifications are required when aiming for a parameter estimation, as PINNs are typically purely used for state estimations problems, since the input-output relationship is not necessarily maintained (Degen et al., 2022). Due to the solution being enforced at specific points only, and the following model reconstruction being solely based on these selected points to obtain the entire state distribution for a single parameter realization only, an UQ is not possible with the traditional use of PINNs (Degen et al., 2022).

2-4-3 Non-Intrusive Reduced Basis Method

NI-RB is an alternative PB-ML technique to the previously mentioned PINNs, combining physical process models with data-driven ML techniques (Degen et al., 2022). It constructs reliable, scalable and interpretable surrogate models of high accuracy from physics-based input models, requiring few basis functions (Degen et al., 2022). The NI-RB method consists of an offline and online stage to construct and use these surrogate models, which are explained in the following paragraphs. A supporting workflow typical for the NI-RB method can be found in Figure 2-9.

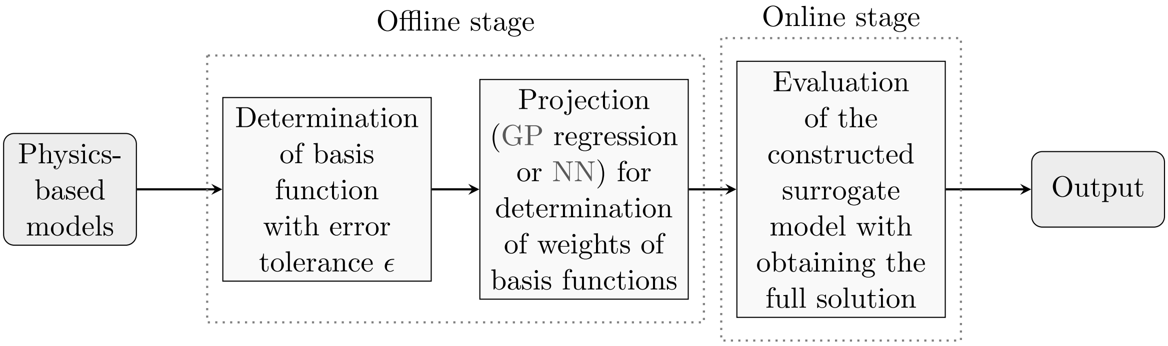

Figure 2-9: Workflow for the NI-RB method.

The surrogate model is constructed during the **offline stage**. This is an unique, non-repeated process and includes all computationally expensive steps (Degen et al., 2022). Two steps are required: the determination of the basis function and the calculation of their coefficients.

Prior to the first step, the chosen, physics-based PDEs are solved, giving the solutions as so-called *snapshots*, which are required to test and train the constructed model. The snapshot matrix $\vec{S} \in \mathbb{R}^{N_h \times N_s}$ includes all snapshots $\vec{u}_h(\vec{\mu})$ for any parameter $\vec{\mu}$ calculated by a fixed FE setting (Guo and Hesthaven, 2018). It is therefore given by

$$\vec{S} = \left[\vec{u}_h\left(\vec{\mu}^1\right) \middle| \vec{u}_h\left(\vec{\mu}^2\right) \middle| \cdots \mid \vec{u}_h\left(\vec{\mu}^{N_s}\right)\right]. \tag{2-38}$$

This snapshot matrix $\vec{S}$ shows the characteristic behavior of the problem (Degen et al., 2022; Hesthaven and Ubbiali, 2018), by representing a collection of input-output pairs of the simulation process. These solutions of the PDEs are computed with an external high-fidelity model for different inputs (Swischuk et al., 2019). Using snapshots instead of implying the physics in the data or as a constrain (as it is done in PINNs) enables a decrease in computational costs of this step while still capturing the physical behavior of the model.

After the creation of the snapshots, the next step of the offline stage takes place: the proper orthogonal decomposition (POD) is implemented in the form of singular value decomposition (SVD). It recovers the RB functions (Hesthaven and Ubbiali, 2018) by keeping the singular vectors corresponding to the largest singular values present in the matrix (Degen et al., 2023). Via the POD we obtain the basis functions that define parts of the model architecture, while a individual error tolerance ϵ can be defined, so that

$$\frac{\sum_{i=1}^{r} \xi_i^2}{\sum_{i=1}^{N} \xi_i^2} \leq \varepsilon, \tag{2-39}$$

with the eigenvalue ξ and the total number of training samples N (Degen et al., 2022). This error tolerance ϵ limits the physical plausible range of each parameter included.

SVD decomposes the rectangular matrix of the snapshots $\vec{S}$ obtained from the PDEs,

$$\vec{S} = \vec{U}\vec{\Sigma}\vec{Z}^{\mathrm{T}}, \tag{2-40}$$

with $\vec{U} \in \mathbb{R}^{N_h \times N_h}$ and $\vec{Z} \in \mathbb{R}^{N_s \times N_s}$ being orthogonal matrices, and $\vec{\Sigma} = \operatorname{diag}\{\xi_1, \xi_2, \ldots, \xi_{N_s}\}$ containing the singular values $\xi_1 \geq \xi_2 \geq \cdots \geq \xi_{N_s} \geq 0$ of the snapshot matrix $\vec{S}$ (Guo and Hesthaven, 2018).

This decomposition of the snapshots into its basis functions allows the reduction of the preliminary snapshot dimensionality, while considering constraints imposed on rows and columns of the matrix (Abdi, 2007). The basis functions focus on the more important features of the matrix, in order to meet the previously defined error tolerance ϵ. This leads to a reduction of computational costs (Chen et al., 2021), while keeping the accuracy sufficiently high for each application individually. These reduced basis functions received via the SVD are the most informative eigenvectors of the initial problem described by PDEs (Efendiev et al., 2011). An example of basis functions needed for a surrogate model construction can be found in Figure 2-10. In this example, which is received from the reduced model construction in this project, four basis functions are required to achieve a pre-defined error tolerance ϵ.

The second step describes the projection step, in which supervised ML is implemented in the form of GP regression or NNs (Degen et al., 2022; Wang et al., 2019; Chen et al., 2021). In the conventional RB, the Galerkin projection is used instead of ML techniques. More information about the RB method can be found in the appendix B-2. Since both the GP regression, as well as NNs are used in this project, a short explanation of both follows. *GP regression* is a non-parametric Bayesian approach towards regression problems to predict new parameter values according to previous ones, while assuming that the input-output pairs follow a prior of Gaussian function (Guo and Hesthaven, 2018; Schulz et al., 2018). Thus, it is guided by a probabilistic distribution over the given basis functions. We use this distribution as new inputs, as mentioned in Guo and Hesthaven (2018). It can include a wide variety of input-output relations as it applies a very large number of parameters while adapting the level of complexity based on the input data and via Bayesian inference (Schulz et al., 2018). Its advantage is in its lower computational costs compare to NNs since it requires less hyperparameters to be adapted. It can perform as accurate as NNs if the problem is not highly nonlinear. Thus, its main application is found in linear parameter variations (Raissi et al., 2017). *NNs* on the other side are more often used for highly non-linear PDEs, as they show a better performance and higher accuracy than GP regression for these types of problems (Degen et al., 2022). The working principle of NN is explained in the previous Section 2-4-2. In the case of the NI-RB method, often a *feedforward NN* is applied, which is defined by multiple layers of interconnected neurons present. The data goes from the input layer to the output layer without any feedback in between (Sazli, 2006). Neurons are only connected with neurons in the following layer, and not in the current layer itself (Hesthaven and Ubbiali, 2018). This creates a map between the input and output space, which is especially of advantage in case of continuous function approximations (Hesthaven and Ubbiali, 2018).

In this project, both GP regression and NNs are applied due to different approaches of handling MT parameters. The use of ML is only present as a projection method, in order to determine the weights of the basis functions. This has the advantage that the black box problem of ML only occurs in this step. Therefore, the approximation errors are mainly related to this weighting step (Degen et al., 2022). Within the process of constructing the surrogate model, the main characteristics of the response can be preserved.

Finally, the reduced model solution is defined as

$$\vec{u}_{rb}(\mu) = \sum_{i=1}^{r} \theta_{rb}^{(i)}(\mu)\psi_i \quad \in V_{rb}, \quad (2\text{-}41)$$

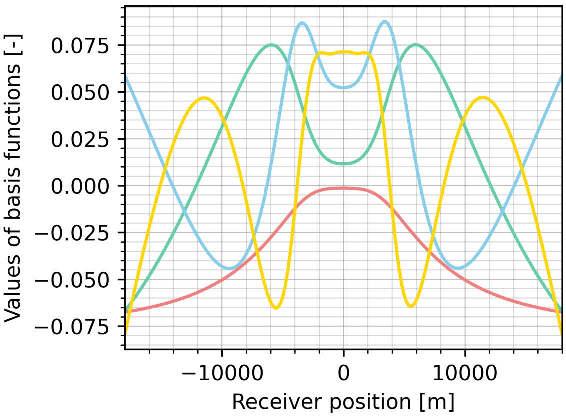

Figure 2-10: An example of four basis functions being required for a reduced model for its predefined error tolerance. These basis functions are obtained via the POD.

where $\mathbf{u}_{rb}$ is the reduced solution, r the size of the reduced basis, θ_{rb} the reduced coefficients or weights, and ψ_i the basis functions that create the reduced space V_{rb} (Degen et al., 2022). In order to validate the obtained result, a validation data set is additionally needed to the training data set, which size should have a minimum of 10% of the training data set (Degen et al., 2022). The training data set is used to construct the reduced model. This validation data set tests the approximation of the surrogate model with its predicted outcomes to the true solution of GoFEM, in our case. These predicted and true values are compared by the calculations of the mean squared error (MSE), yielding the Q2 score. This is similarly done with the training data, emerging in the R2 score. These two values, Q2 and R2, reflect the overall model performance and are thus aimed to be minimized by adjusting the models architecture and its hyperparameters accordingly.

As mentioned, this **offline step** is computationally expensive, which is caused by two factors. The first component is the construction of the training set, including solving numerous expensive finite element simulations (Degen et al., 2022). The second computationally complex part is the determination of the hyperparameters during the ML. The former factor can be reduced by an efficient sampling strategy as, e.g., the implementation of LHS with a preferably small range of values (Hesthaven and Ubbiali, 2018; Degen et al., 2022), while the latter can be improved by pre-processing input data or by using the optimization technique of a Bayes optimization scheme, to accelerate the convergence of the hyperparameter tuning (Degen et al., 2022).

During the **online stage** the surrogate model is used allowing an efficient computation process (Degen et al., 2022). The previously defined reduced order framework is used to find a suitable approximation to the true solution by a linear combination of the basis functions and coefficients specifically determined for the newly given parameter values, based on the model obtained from the offline stage (Hesthaven and Ubbiali, 2018).

Overall, by combining physical process models and data-driven ML, the NI-RB method represents a solution to overcome the limitations of these approaches when being used individually (Degen et al., 2022). By implementing PODs to decrease the dimensional complexity of a problem (Chatterjee, 2000), the physical behavior is extracted and preserved, instead of treating physical laws purely as constraints or data as it is done in PINNs (Degen et al., 2022). ML is purely implied to determine the weight of the characteristics, which restricts the error and the problem of having a black box in this process (Degen et al., 2022). Therefore, error deviations are mainly based on the combination of scalar values. Due to the use of ML, the NI-RB method maintains a high accuracy, while providing orders of magnitude computational gain and requiring less data points than needed in purely data-driven ML (Degen et al., 2022). Due to the theory standing behind the NI-RB method, this approach can be applied in a wide range of geoscientific problems, also including non-linearly coupled physical problems (Degen et al., 2022). Its solution allows multiple realizations of pre-defined parameter sets, which allows the performance of probabilistic analysis such as global sensitivity analysis (SA) and UQs (Degen et al., 2022). The four main advantages of the NI-RB method can be summarized as the following: (1) physical scalability to small- and large-scale processes, as well as performance scalability, (2) interpretability of the surrogate models by following physical laws in a consistent manner, (3) generalizability of the approach to different applications, and (4) robustness of the model behavior (Degen et al., 2023). A summarizing schematic graphic is displayed in Figure 2-11, comparing physics-based, data-driven and PB-ML methods, with the application of both data and models, as well as the benefits and disadvantagess of each technique (Degen et al., 2023).

As an initial example of using the NI-RB method in this project, we perform simulations based on thermal conductivities as input parameters with the open-source, parallel finite element, high-performance toolkit MOOSE (Lindsay et al., 2022). MOOSE is developed by the Idaho National Laboratory (Peterson et al., 2018). The mesh used for this model is created in the open-source finite element grid generator *GMSH* (Geuzaine and Remacle, 2009). These simulations are used as a training to become accustomed to the accuracy of the NI-RB method compared to the true solution of MOOSE, and to enhance the understanding of this approach before combining NI-RB method with the application of GoFEM, as this has not been linked before. As this trial is only preliminary work and not in the focus of this thesis, the results obtained are not shown.

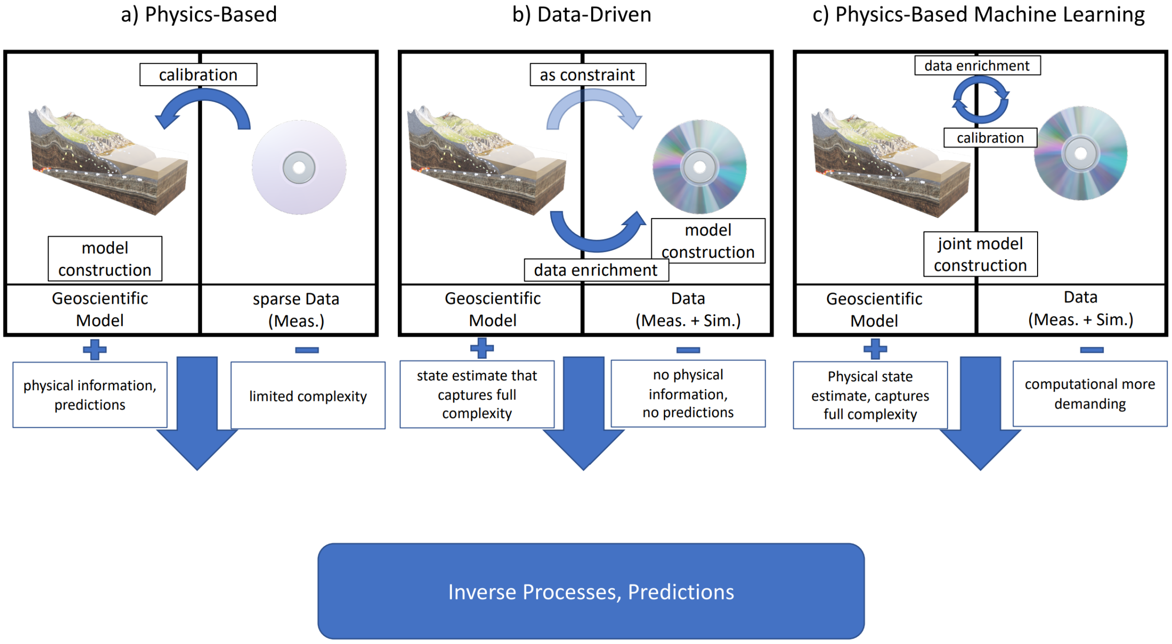

Figure 2-11: Schematic illustration of physics-based, data-driven and PB-ML techniques, including their approach of data- and model handling. A 'plus' and 'minus' signs describe main advantages and disadvantages of each method, respectively. *meas.* means *measured*, and *sim.* stands for *simulated*. The figure is taken from (Degen et al., 2023).

Chapter 3

Results

This section reveals the results obtained in this project. It is to remember that the chosen model consists of five bodies (Figure 2-7). The material properties of four bodies are varied over a wide electrical conductivity range of up to orders of magnitude (Table 2-1), while the air layer is kept constant. These training parameters representing electrical conductivities are picked via LHS (Section 2-3), to allow a near-random sampling strategy over the entire predefined range (a list of these values can be found in Section C). For the entire project, a set of 150 training parameters is used, as this ensures an sufficient amount of simulations to construct an accurate surrogate model using the NI-RB method. The validation parameters are determined for the same conductivity ranges with a random sampling strategy, to ensure independence of the training- to the validation data. 30 validation parameters are defined as such, since they typically should represent minimum 10 % of the amount of training parameters (Degen et al., 2022). These validation parameters aim to test the quality of the surrogate model performance on a data set other than the one used in the training. These parameter sets are used for forward simulations with GoFEM, to create 150 and 30 MT responses for the training and validation data set, respectively. An representative example of these forward simulations for the parameters phase and resistivity is shown in Figure 3-2. It is to note that for all analysis in this project a frequency of 1 Hz is used, with the exception of one frequency investigation of 1000 Hz in Chapter 3-1. These two frequencies of 1 and 1000 Hz are chosen as they include larger and smaller changes in values, respectively, thus aiming to test the NI-RB method in its extremes.

As an initial step, the focus is on the parameters phase and resistivity, as the NI-RB method implementations for real-valued problems already exist. We aim to test the potential of the NI-RB method in the application of MT with the current possibilities offered by the NI-RB method. We compare the projection method GP regression and NNs for their suitability. To ensure high validity of the NI-RB method over frequency ranges, we conduct a brief analysis of another frequency with weaker signal variations. Then, the focus is shifted to the complex-valued impedance, while simplifying its approach by treating it as a real-valued parameter. This is achieved by splitting the impedance into its real and imaginary parts, which are trained individually. We compare again the approach of GP regression with NNs.

This step is followed by an implementation of the impedance as a complex-valued parameter in its entirety, with the projection method of NNs. Due to issues encountered during this approach, we switch back to the GP regression, while splitting the real and complex parts after the POD. Ultimately, we transform the complex-valued impedance into real-values, to obtain a high model consistency and avoid issues arising from complex values. Frequency analysis of the surrogate model of the impedance would be also of interest, however this is not in the frame of this project. Figure 3-1 summarizes the described workflow.

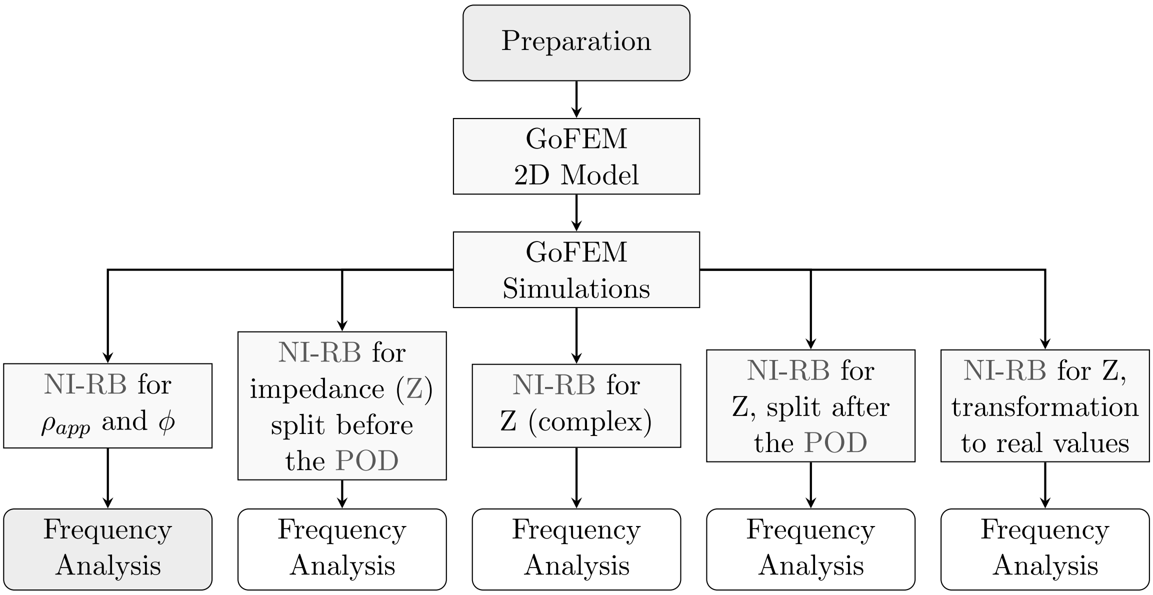

Figure 3-1: Workflow for the overall study. The frequency analysis is only conducted for the phase, but could be done additionally for the impedance (white boxes).

3-1 Phase and Resistivity

Resistivity ρ and phase ϕ are two key parameters of the MT response, providing information about the electrical conductivity distribution of the subsurface (Vozoff, 1991). GoFEM (Grayver and Kolev, 2015) allows simulations of these parameter, as it is specifically developed for electromagnetic problems. The first part of this project is to use the NI-RB method to construct a surrogate model for the resistivity and phase individually, for both the TE and TM modes. These four surrogate models are based on 150 forward simulations, which are validated with 30 additional simulations. The goal is to estimate the possible accuracy the NI-RB method with respect to the full solutions of GoFEM. For this MT application, we aim for an maximum local error tolerance of 1 % for the unscaled data set, which is a commonly set error tolerance in MT applications. Additionally, the POD error threshold of $1 \cdot 10^{-4}$ for the scaled data set is determined before defining the basis functions.

For each parameter (phase and resistivity for the TE- and TM-mode) one separate surrogate model needs to be trained. The NI-RB method (Degen et al., 2022) applied in this project uses several ML techniques as the projection method, while we employ the GP regression and NNs in this project. The advantage of GP regression is its faster execution of the ML

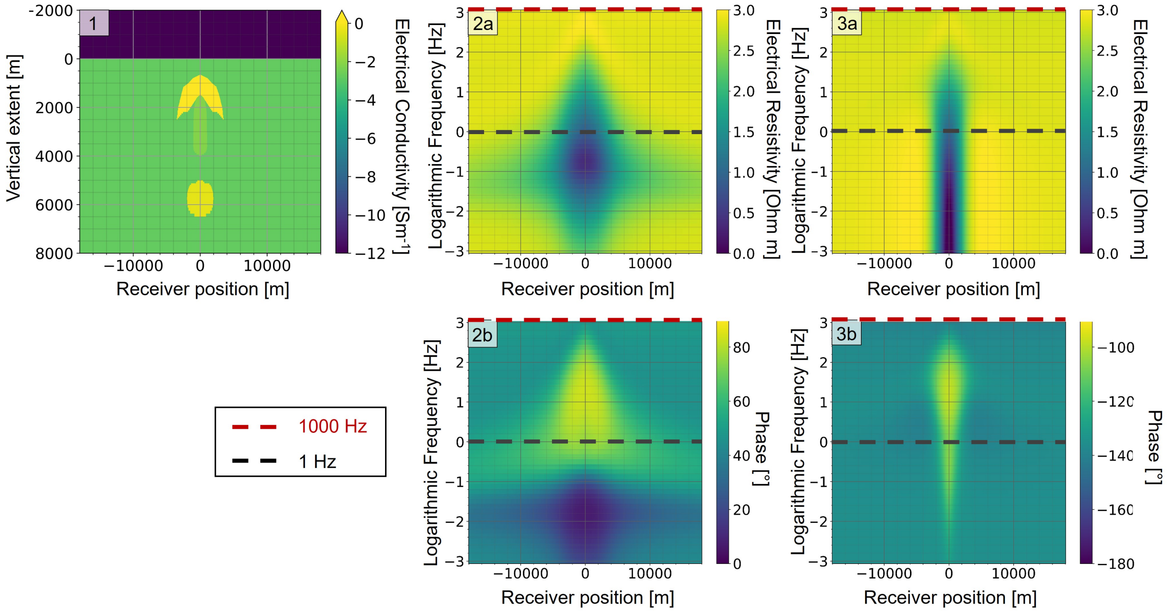

Figure 3-2: MT response of the forward simulations using GoFEM of the predefined mesh from Figure 2-7). (1) Conductivity distribution of the model. (2a) and (2b) shows the phase and resistivity of the TE mode, and (3a) and (3b) of the TM mode, respectively. The dark grey and red dashed lines mark 1 and 1000 Hz, respectively, which are frequencies used in the analysis of this project.

step, because for the NN, more hyperparameters need to be trained until the set is optimized sufficiently. However, NN has a wider range of applications, as it can deal with non-linear and hyperbolic problems, where GP regression lacks in its performance. For the approach of the real-valued resistivity and phase, both ML methods are tested and compared in Table 3-1. In order to understand the table better, a brief description for the adjustable architecture of the GP regression is given:

Length scale: represents the length scale of the kernel, defining over which distance the covariance between data points decreases.

Length scale bounds: describes the range between which the length scale can vary.

Nu: controls the smoothness of the kernel.

Number of restarts of optimizer: states how often the optimizer is restarted for the optimal hyperparameter search.

Optimizer: is a algorithm to adjust the model parameters for higher accuracy.

The following list explains the adjustable architecture and hyperparameters of the NNs:

Activation function: is a function applied to the output of every neuron.

Loss function: measures the difference between the predicted and the true output.

Optimizer: is a algorithm to adjust the model parameters for a higher accuracy.

R2/Q2 calculator: determines the model performance compared to the training (R2) and validation (Q2) data set.

Number of layers: is the number of hidden layers, excluding input and output layer.

Number of neurons per layer: represents the number of neurons per hidden layer, which can receive inputs and compute outputs based on weights and biases.

Number of epochs: defines the number of iterations over the complete NN.

Learning rate: controls the step size of the NN to adjust its weights, also being referred to as the speed of convergence.

Batch size: represents the number of training samples that are treated as one group to update the network's weights.

We would like to point out several aspects which we can conclude from Table 3-1:

- The *number of basis functions* varies between four and six, while the phase requires two more basis functions than the resistivity. Therefore, the phase seems to be more complex in its shape (please note the area around 0 m on the x-axis in Figure 3-3 (1a) compared to (2a), confirming this assumption). However, this range of basis functions is overall low, considering that the material properties of four bodies inside the geothermal reservoir are varied. Thus, each body might require only one basis function for the resistivity, but further analyses would be require to prove this assumption. Taking into account the preliminary definition of 150 training parameters, we can conclude that due to the low number of basis functions probably 50 samples would be sufficient for the construction of the surrogate model.
- The *hyperparameters* of the NN show overall the complexity of the surrogate model. It is to note that especially the TM mode requires only one to two hidden layers, keeping the model structure rather simple and calculations fast. As the NI-RB method reduces dimensionality, we usually require a less complex NN architecture for this method than when applying a pure NN-bases approach Degen et al. (2022). This allows us to reach such a simple architecture and lower training times, as we can see in these models.
- The *training and validation data global model error* of both the GP regression and NN show a magnitude ranging from 10^{-9} to 10^{-7} for the scaled data set, calculated with the MSE. The global model error represents the global discrepancy between the predictions made by the reduced model and the true solution, in this case the NI-RB method and GoFEM, respectively. In this project this error is defined as the MSE between the reduced model and the training and the validation data, respectively. We can see slightly smaller errors in GP regression of the training data than in the NN. The validation data error is similar for both approaches.
- The *maximum local error* reveals the the maximum discrepancy between the surrogate model and the true solution, without averaging as it is done in the MSE calculations for the global error. Thus, it is more sensitive to outliers and reveal the largest inaccuracy

of the surrogate model. Our aim is to keep it below 1 % to ensure an overall high precision without outliers. This is achieved with both projection methods , except in the NN-application of the phase in the TE mode, where the error is around 15 times higher than for the other parameters. Overall we can see that the approach with GP regression enables the desired results of an accuracy below 1 % in all four parameters.

Table 3-1: Parameters of the GP regression and NN approach as a comparison for their accuracy relative to the complexity of the surrogate models, for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set and a frequency of 1 Hz for the phase ϕ and the resistivity ρ of TE and TM. Abbreviations f.t.t.b: *LS* = length scale, *LSB* = length scale boundaries, *Nu* = smoothness factor, *#Ro* = number of restarts of optimizer, *Optim.* = optimizer, *#BF* = number of basis function, *ME TD* = global model error training data, *ME VD* = global model error validation data, *Max. LE* = maximum local error [%], *Activ.* = activation function, *Loss* = loss function, *Q2/R2*: Q2/R2 calculator, *#L* = number of hidden layers, *#NL* = number of neurons per layers, *#E* = number of epochs, *LR* = learning rate, *BS* = batch size. Values marked with ∗ are valid for all input parameters.

ML	Parameters	TE-Phi	TE-Rho	TM-Phi	TM-Rho
GP regression	LS	1*			
	LSB	(1e-5,1e5)*			
	Nu	2.5			
	#RO	0			
	Optim.	L-BFGS-B			
	# BF	6	4	6	4
	ME TD	$8.22 \cdot 10^{-9}$	$1.96 \cdot 10^{-9}$	$4.64 \cdot 10^{-9}$	$3.19 \cdot 10^{-9}$
	ME VD	$9.54 \cdot 10^{-7}$	$9.07 \cdot 10^{-8}$	$3.45 \cdot 10^{-8}$	$1.54 \cdot 10^{-8}$
	Max. LE	0.34	0.21	0.18	0.11
NN	Activ.	Sigmoid			
	Loss	MSE			
	Optim.	Adam			
	Q2/R2	MSE			
	# BF	6	4	6	4
	# L	3	3	1	2
	# NL	253	113	309	87
	# E	46198	41329	10308	39231
	LR	$1.43 \cdot 10^{-4}$	$2.13 \cdot 10^{-4}$	$2.33 \cdot 10^{-4}$	$2.20 \cdot 10^{-4}$
	BS	56	20	7	45
	ME TD	$3.51 \cdot 10^{-8}$	$3.62 \cdot 10^{-8}$	$9.14 \cdot 10^{-7}$	$4.55 \cdot 10^{-8}$
	ME VD	$1.11 \cdot 10^{-7}$	$8.27 \cdot 10^{-8}$	$1.39 \cdot 10^{-6}$	$5.32 \cdot 10^{-8}$
	Max. LE	2.50	0.17	0.11	0.16

Hence, we conclude that both the application of GP regression and NN result in satisfactory accuracy. The local error of the GP regression is consistently below the accepted local error tolerance of 1%, and as it additionally requires significantly lower computational costs as it is easier to train, we decide that GP regression is the desired and more efficient projection method for this application.

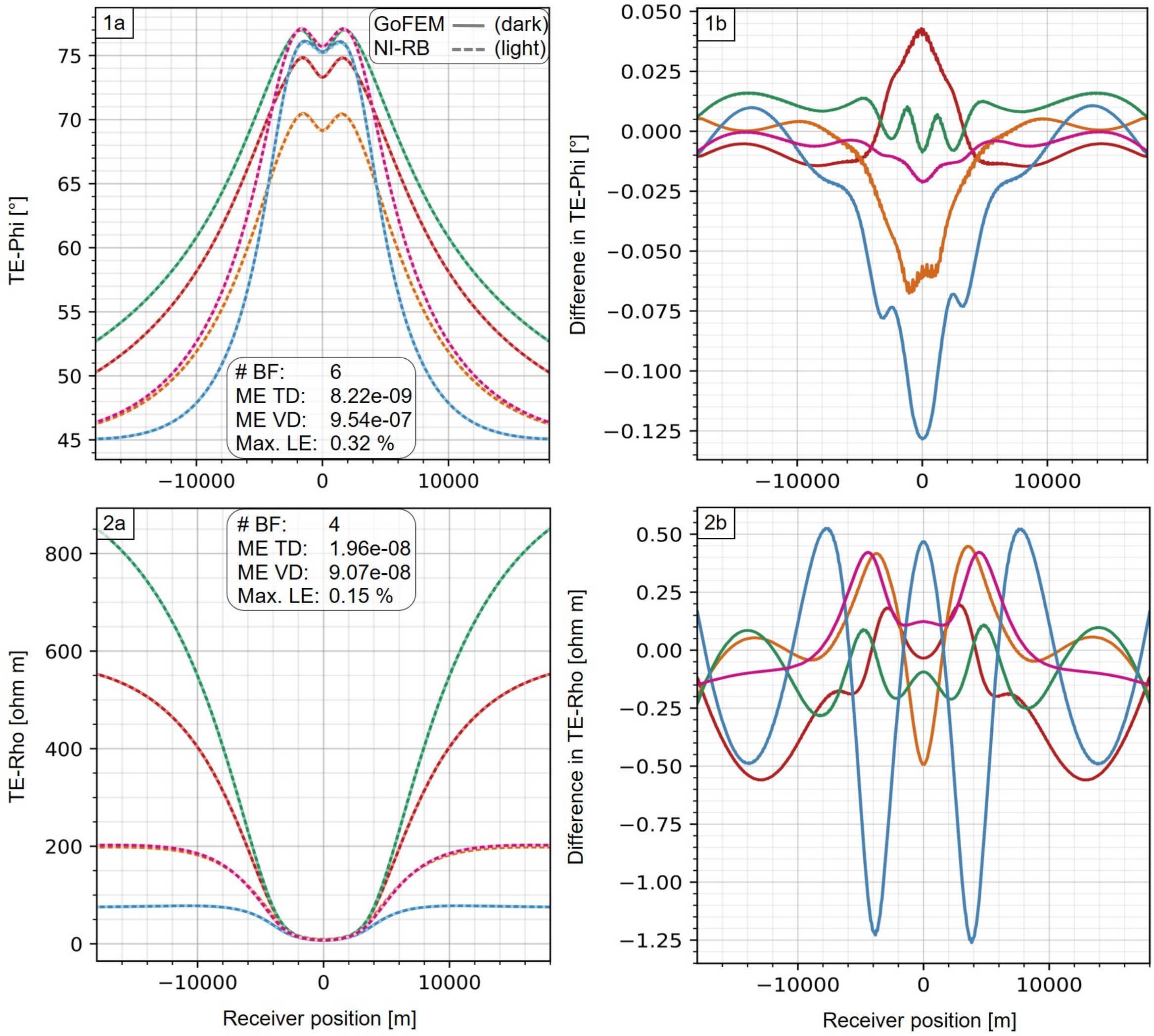

Figure 3-3: Representation of the accuracy of the surrogate model of a MT simulation of the phase and resistivity of the TE mode a high-temperature geothermal system for frequency of $1\ \mathrm{Hz}$ with the projection method GP regression. The left side shows five realizations with different parameter values from the validation data set for the phase ϕ (1a) and the resistivity ρ (2a) of TE-phi. Solid lines represent the GoFEM simulation, the dashed line the NI-RB solutions. *# BF* shows the number of basis functions needed for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set, *ME TD* and *ME VD* demonstrates the global model error of the training data and validation data, respectively, and *Max. LE* shows the maximum local error. The right side shows the difference between the NI-RB solution and the GoFEM simulation of each realization on the left.

Figure 3-3 compares the full solution of GoFEM with the NI-RB simulation with GP regression as ML technique graphically. As an example, we show phase and resistivity of the TE mode for five realizations of the validation data set. A realization in this project represents one simulation of a set of four electrical conductivities in the predefined range (Table 2-1) for each body in the geothermal model (Figure 2-7). The left side shows the high accuracy of the NI-RB method, as no difference is visible to the eye in this scale. The figures on the right side show the difference between the NI-RB and the GoFEM solution. We would like to draw attention to increasing deviations at sections of rapid changes in values of the phase an resistivity around 0 m for the phase and around -6000 and +6000 m for the resistivity. It seems to be more difficult for the NI-RB method to approximate the full solution if it varies rapidly, but without significant impact as the local error still remains below 1 %.

In order to see the performance of the NI-RB method in other environments, we decided to test this technique in another frequency range characterized by less significant value fluctuations. The aim is to test the NI-RB method can detect also very little variations or if they remain overseen. For this test, the phase of the TE mode with a frequency of 1000 Hz is chosen, as its signal is smooth with only some little changes in values (Figure 3-2, red line). Figure 3-4 reveals the results of it. The surrogate model performs very well, with a maximum local error of 0.06 %. We can see that the model requires one basis function less than for the frequency of 1 Hz. Similar to the frequency of 1 Hz, the largest discrepancies happen at abrupt changes (e.g., around 0 m). However, it is to note that there are also increases in differences at both edges of the x-axis, which does not happen for the frequency of 1 Hz. Its reason is the frequency drop close to the borders for some realizations (in Figure 3-4-a the blue and pink line), because quick changes lead to the decrease in accuracy of the NI-RB method visible in Figure 3-4-b. However, the model performance is far below our error tolerance of 1%, leading to the conclusion that the use of the NI-RB method should approximate the full solution within our error tolerance over the entire frequency range of the MT response.

Including both frequency test of the resistivity and phase in the TE and TM mode, we confirm a successful use of the NI-RB method for these two real-valued parameters of both the TE and TM mode.

3-2 Impedance

The second goal of this project is the surrogate model construction for the impedance, representing the complex ratio of the E and H field. It is a variable that is directly obtained from MT measurements, and contains information that are additionally included in the resistivity and phase as mathematically proven in Eq. (2-20). Using the NI-RB method for the impedance is more challenging as for the phase and resistivity, as complex-valued number are not yet implemented in this method, as well as the research of complex-valued MT data in ML techniques is still in its early stage. In the following sections, different approaches of implementing the impedance into the NI-RB method are explained.

3-2-1 Split before the Proper Orthogonal Decomposition

The first approach of using the NI-RB method for the impedance is its split into its real and imaginary part before basis functions are determined via the POD. Both of these parts are

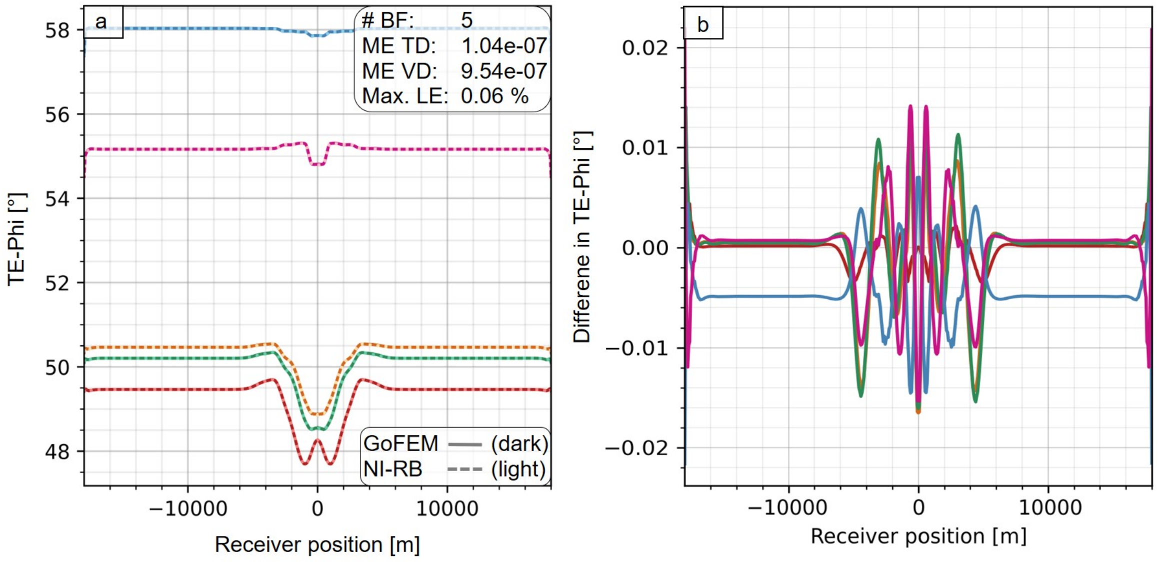

Figure 3-4: Representation of the accuracy of the surrogate model of a MT simulation of the phase in the TE mode a high-temperature geothermal system for frequency of 1000 Hz with the projection method GP regression. The left side shows five realizations with different parameter values from the validation data set for the phase ϕ (a). Solid lines represent the GoFEM simulation, the dashed line the NI-RB solutions. *# BF* shows the number of basis functions needed for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set, *ME TD* and *ME VD* demonstrates the global model error of the training data and validation data, respectively, and *Max. LE* shows the maximum local error. (b) shows the difference between the NI-RB solution and the GoFEM simulation of each realization on the left.

then treated like real-valued parameters, and trained separately in the ML step to calculate the coefficients. A workflow for this approach is shown in Figure 3-5. Similar to the approach for the resistivity and phase described in Section 3-1, we compare the suitability of both projection methods GP regression and NN for this procedure.

Table 3-2 shows all parameter of interest for the construction of the surrogate model for both the GP regression and NN approach, with error threshold of $1 \cdot 10^{-4}$ for the scaled data set. The explanation of the hyperparameters in the NN section can be found in Section 3-1. We would like to highlight specific aspects concluded this table:

- The *number of basis functions* is similarly to the results for the resistivity and phase very low. For the TE and TM mode, four and three basis functions are required, respectively. This lets us assume that the 150 training samples created are not necessary and 50 would have been probably sufficient.

- The *hyperparameters* reveal that the TM mode requires less layers than the real component of the TE mode. This leads to the assumption that the construction of the TM mode is simpler than the TE mode. The *global model errors of the training and validation data set* of the GP regression are two magnitudes higher for the impedance than for the resistivity and phase (Table 3-1). Comparing the global model errors from the GP regression with the NN gives an equal suitability of both projection methods, with a slightly better overall performance of for GP regression.

- The *maximum local error* of both approaches is similar, with the exception of the phase of the TE mode for the NN method. There, the error of 2.5 % exceeds our predefined maximum error tolerance of 1 %.

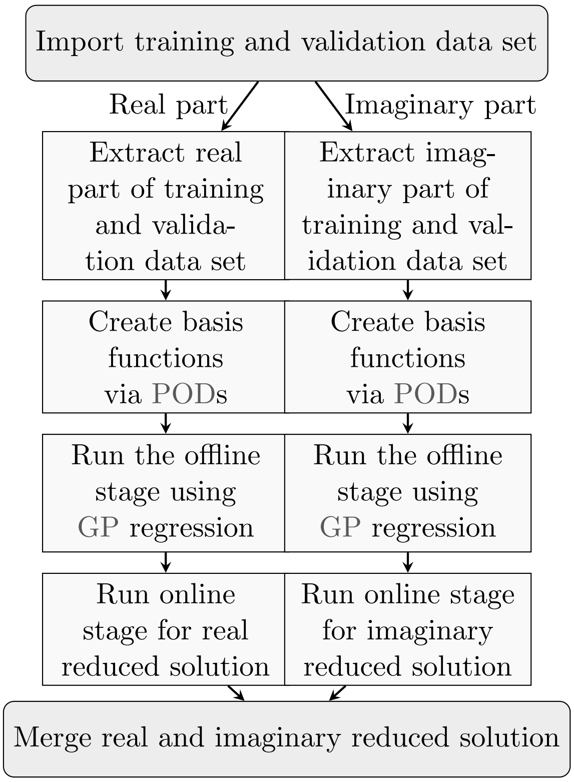

Figure 3-5: Workflow for splitting the impedance in its real and imaginary parts before PODs using GP regression.

Due to the slightly better overall performance, especially in the maximum local error, we decide that the GP regression is the preferred ML strategy. Its lower computational costs caused by less hyperparameters required to be trained, gives an additional advantage, while not loosing but gaining accuracy compared to the NN method.

Figure 3-6 shows visually the performance of the NI-RB method using GP regression as the ML technique. The left side shows the performance of the full solution of GoFEM with solid lines, and of the NI-RB method with dashed lines for five realizations and a frequency of 1 Hz. The full solution is very well approximated by the NI-RB method in both the real and imaginary part, as no difference is visible by eye in the left plots. The right side shows the difference between the NI-RB method and the GoFEM simulations. We can see that one realization (blue line) shows clearly larger deviations than the other four, which can be caused by electrical conductivities that were randomly less frequently trained. However, to prove this assumptions, further investigations are needed. We would also like to point out the regularity of these errors, as they do not include random spikes, but show a smooth behavior. This leads to the conclusion that these discrepancies are mainly caused by the truncated basis functions. Knowing the error source is convenient, as it is easier to be tackled than if the errors were of random nature.

We can conclude an accurate performance of the NI-RB method with the GP regression application for the impedance, when it is split before the generation of the basis functions via the POD. However, we would like to point out the problematic of this approach: creating two models leads to higher inconsistency than if we would only have one model to build. Thus, we follow up in the next section with an approach where the impedance is regarded as a complex-valued parameter throughout the construction of the surrogate model.

Table 3-2: Parameters of the GP regression and NN approach as a comparison for their accuracy relative to the complexity of the surrogate models, for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set and a frequency of 1 Hz. The impedance is split before the POD into its real and imaginary components. Abbreviations f.t.t.b: *LS* = length scale, *LSB* = length scale boundaries, *Nu* = smoothness factor, *#Ro* = number of restarts of optimizer, *Optim.* = optimizer, *#BF* = number of basis function, *ME TD* = global model error training data, *ME VD* = global model error validation data, *Max. LE* = maximum local error [%], *Activ.* = activation function, *Loss* = loss function, *Q2/R2*= Q2/R2 calculator, *#L* = number of hidden layers, *#NL* = number of neurons per layers, *#E* = number of epochs, *LR* = learning rate, *BS* = batch size. Values marked with $*$ are valid for all input parameters.

ML	Parameters	TE-Z real	TE-Z imag	TM-Z real	TM-Z imag
GP regression	LS	1*			
	LSB	(1e-5,1e5)*			
	Nu	2.5			
	#RO	0			
	Optim.	L-BFGS-B			
	# BF	4	4	3	3
	ME TD	$6.85 \cdot 10^{-8}$	$1.57 \cdot 10^{-7}$	$2.51 \cdot 10^{-7}$	$3.34 \cdot 10^{-7}$
	ME VD	$3.90 \cdot 10^{-7}$	$4.15 \cdot 10^{-7}$	$1.28 \cdot 10^{-6}$	$1.28 \cdot 10^{-6}$
	Max. LE	0.48	0.50	0.80	0.72
NN	Activ.	Sigmoid			
	Loss	MSE			
	Optim.	Adam			
	Q2/R2	MSE			
	# BF	4	4	3	3
	# L	4	1	2	2
	# NL	238	74	145	82
	# E	424483	33254	42730	24094
	LR	$4.33 \cdot 10^{-4}$	$4.78 \cdot 10^{-4}$	$1.94 \cdot 10^{-4}$	$6.90 \cdot 10^{-3}$
	BS	117	74	12	122
	ME TD	$1.06 \cdot 10^{-7}$	$2.10 \cdot 10^{-7}$	$2.82 \cdot 10^{-7}$	$7.69 \cdot 10^{-7}$
	ME VD	$2.07 \cdot 10^{-7}$	$2.65 \cdot 10^{-7}$	$2.31 \cdot 10^{-7}$	$1.92 \cdot 10^{-6}$
	Max. LE	1.98	0.22	0.26	0.91

3-2-2 Split inside the Projection Method

In the second approach, the impedance is treated as a complex-valued parameter. The basis functions are calculated for the complex-valued values in its entirety, and also the data sets of the impedance simulations are read into the projection method as complex values. For this approach, the NI-RB method needs to be adapted to complex-valued variables. Therefore, we exchange packages and functions with ones that are suitable for complex numbers. Since better suitable python installations are found for the approach with NNs than with GP regression, we decide to use NNs for the implementation of the complex-valued impedance. Both the structure and some functions of the high-level toolbox *complexPyTorch* developed by

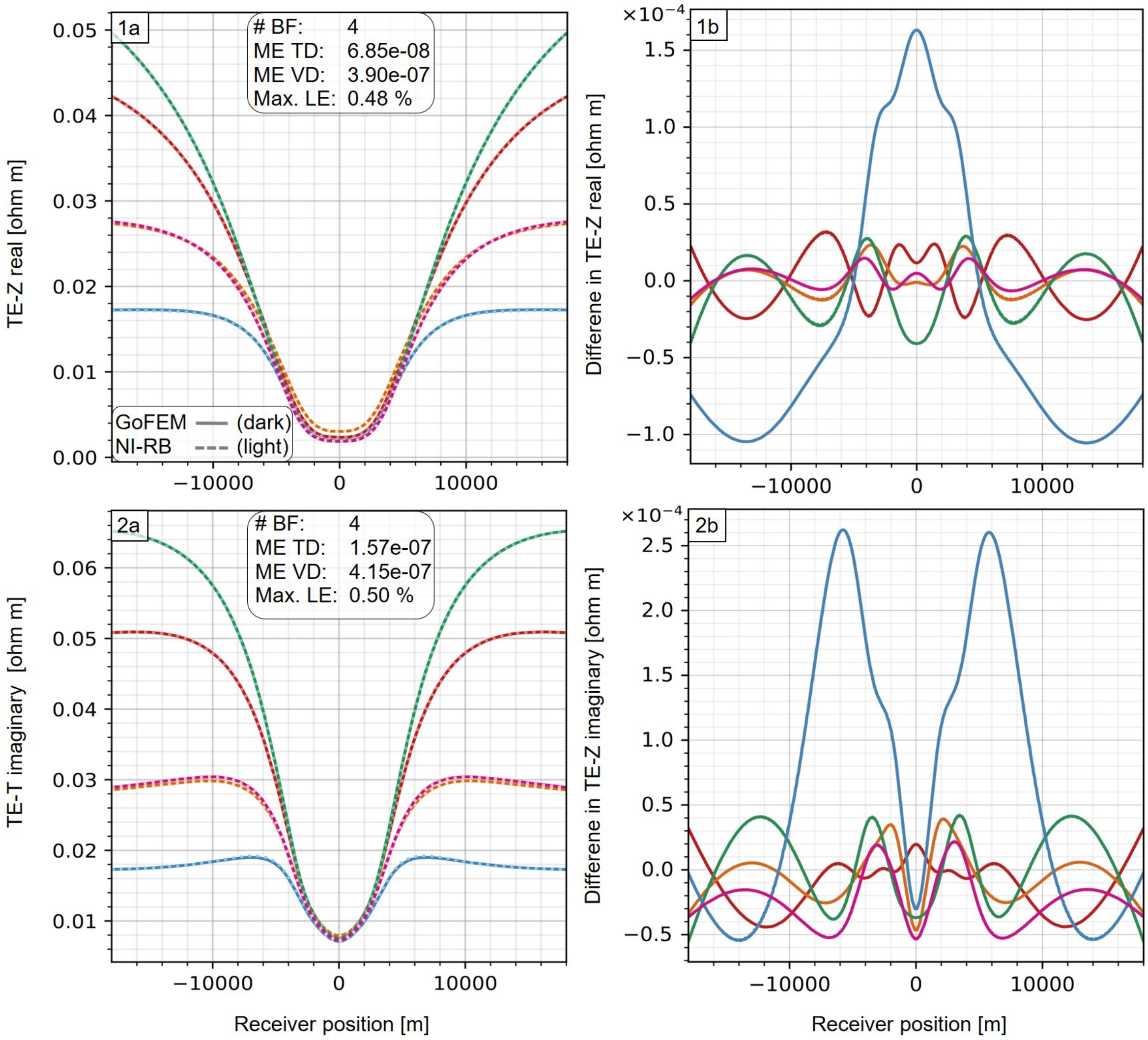

Figure 3-6: Representation of the accuracy of the surrogate model of a MT simulation of the impedance of a high-temperature geothermal system for frequency of $1\ \mathrm{Hz}$ with the projection method GP regression, where the impedance is split before the POD. The left side shows five realizations with different parameter values from the validation data set for the real (1a) and imaginary part (2a) of the impedance *Z*. Solid lines represent the GoFEM simulation, the dashed line the NI-RB solutions. *# BF* shows the number of basis functions needed for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set, *ME TD* and *ME VD* demonstrates the global model error of the training data and validation data, respectively, and *Max. LE* shows the maximum local error. The right side shows the difference between the NI-RB solution and the GoFEM simulation of each realization on the left.

Matthès et al. (2021) are utilized in addition to the library PyTorch (Paszke et al., 2019) for the employment of the CVNNs. The following aspects are tested for this specific application:

- Adjustments of the initial scaling of the training- and validation data set,
- Implementation of an initialization method, including the definition of layers and testing for a suitable activation function,
- Adaption of matrix multiplications,
- Testing for a suitable loss functions,
- Testing for a suitable optimizer,
- Testing for suitable global model error calculators for the training- and validation data set.

It is to note that there are only activation functions available in *PyTorch* and its additional packages which calculate the output separately for the real and imaginary component. Therefore, the impedance is not fully handled as a complex-valued parameter, which is leads to loss of model consistency. We would like to emphasize this splitting, as also in this approach the impedance is, consequently, not treated as a complex-valued parameter in its entirety, as originally expected. After the use of the activation function for each neuron separately, the output of the real and imaginary part is merged together for the training and testing. A workflow for this procedure can be found in Figure 3-8.

In Table 3-3 we can see the architecture, hyperparameters, as well as global model errors and local errors for the following three different cases: (1) the imaginary component is set to zero, (2) the real part is set to zero, and (3) both components are unequal zero. We would like to point out several aspects:

- The *architecture* (including the activation and loss function, the optimizer and the calculations for the Q2 and R2) is for all three tests the same, as we choose them for the implementation of complex-valued impedance, and for comparison reasons.
- The *number of basis functions* is equal, because we split the impedance after the POD.
- The *hyperparameters* are kept constant for all three tests, for comparison reasons.
- The overall *global model error* is smaller for tests (1) and (2), with test (2) performing the best. It is to note that all global model errors are too high.
- The *maximum local errors* are only in two cases below our threshold for error tolerance of 1 %: for the real component when the imaginary part is set to zero (1) and for the imaginary component when the real part is set to zero (2). In the other cases it clearly exceeds our tolerance, or even goes to infinity. This explains the overall *global model error* performing below our expectations, as this value combines both the imaginary and real components.

Table 3-3: Parameters of the NN approach for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set and a frequency of 1 Hz in the case of splitting Z in real and imaginary part inside the activation function of the NN. The second and third columns from the left represent the case where the imaginary and the real component is set to zero, respectively, while in the fourth column both components are unequal zero. Abbreviations f.t.t.b: *Activ.* = activation function, *Loss* = loss function, *Optim.* = optimizer, *Q2/R2*= Q2/R2 calculator, *#BF* = number of basis function, *ME TD* = global model error training data, *ME VD* = global model error validation data, *Max. LE* = Maximum Local Error [%], *#L* = number of hidden layers, *#NL* = number of neurons per layers, *#E* = number of epochs, *LR* = learning rate, *BS* = batch size.

Parameters	Imaginary = 0	Real = 0	None = 0
Activ.	Tanh		
Loss	MSE		
Optim.	Adam		
Q2/R2	MSE		
# BF	4		
# L	3		
# NL	150		
# E	40000		
# LR	$1.94 \cdot 10^{-4}$		
# BS	150		
ME TD	$2.23 \cdot 10^{-2}$	$4.50 \cdot 10^{-3}$	$2.23 \cdot 10^{-1}$
ME VD	$2.92 \cdot 10^{-2}$	$4.85 \cdot 10^{-3}$	$3.23 \cdot 10^{-1}$
Max. LE real	0.13	∞	56.32
Max. LE imag.	∞	0.19	87.42

Looking at Figure 3-7, we can see the results from Table 3-3 graphically. In the first column, where the imaginary part is set to zero (1), the real component is perfectly fit without difference visible by eye. However, the zero line of the imaginary part of the GoFEM is not approximated by the surrogate model at all, as neither the zero values nor the horizontal line are visualized. In the second column (2) the same happens vice versa. The third column, where both components are non-zero, the surrogate model can reconstruct parts of the full solutions shape, especially for the real component. However, its accuracy is insufficient, especially when moving away from the center on the x-axis towards its edges. It is to note that zero-values are generally difficult to approximate for gradient-descent-based optimizers, which explains the inaccurate performance for components being equal to zero (Chuang and Barba, 2022). We assume that the problems for both non-zero components (3) occur due to the lack of python packages being suitable for the implementation of complex-valued parameters specifically in the NI-RB method. It is challenging for this particular application, as we link the basis functions from the POD with the simulation results before the ML step, and also afterwards with the obtained model. This coupling might be one of the reason for the inaccuracy visualized in Figure 3-7-3a and -3b. The fact that it works very well if one component is set to zero shows us that the use of NI-RB for complex-valued parameters is theoretically possible.

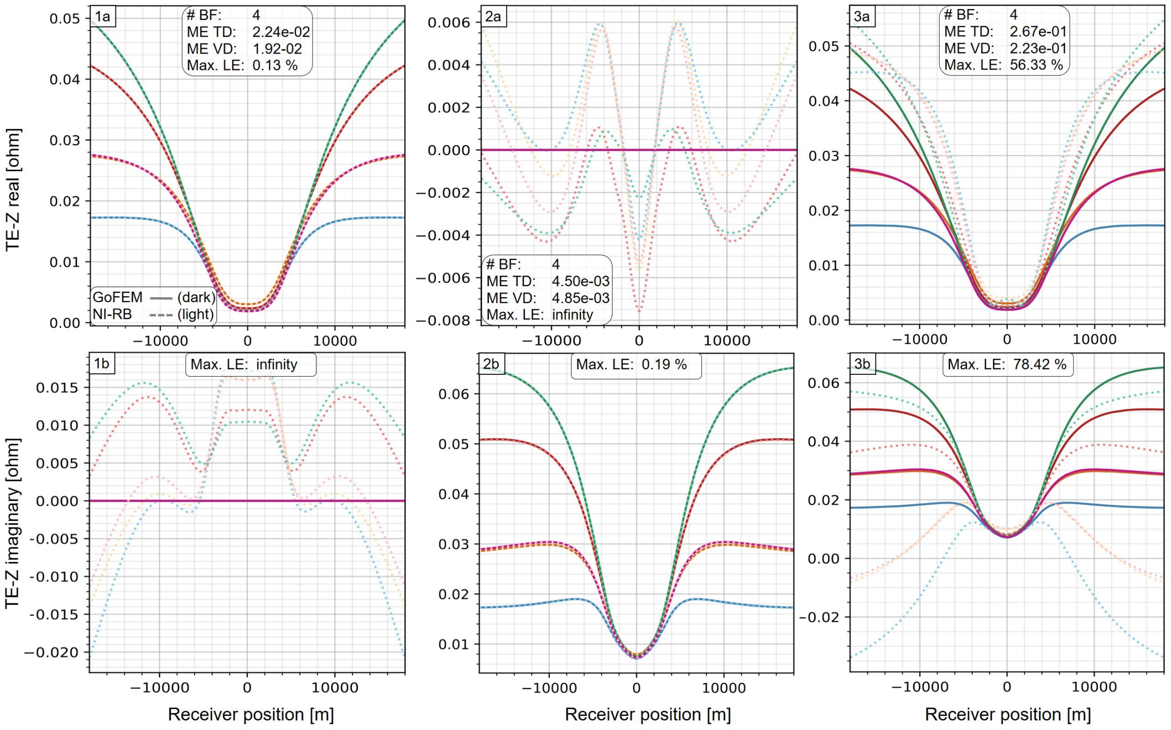

Figure 3-7: Representation of the accuracy of the surrogate model of the impedance in the TE mode of a high-temperature geothermal system for frequency of 1 Hz with the projection method of NN. The impedance is split into its real and imaginary component inside the activation function of the NN. The upper figures show the real, and the lower ones the imaginary part of the impedance, respectively. The first (1) and second (2) columns represent the case where the imaginary or real part is set to zero, respectively. For the third column (3), both the real and imaginary part are unequal zero. Each figure includes five realizations with different parameter values from the validation data set. Solid lines represent the GoFEM simulation, the dashed line the NI-RB solutions. *# BF* shows the number of basis functions needed for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set, *ME TD* and *ME VD* demonstrates the global model error of the training data and validation data, respectively, and *Max. LE* shows the maximum local error.

Table 3-4: Parameters of the NN approach for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set and a frequency of 1 Hz in the case of splitting Z in real and imaginary part inside the activation function of the NN. This table shows tests of different NN architectures, where either the activation or loss function, the optimizer, the Q2 /R2 calculations or the scaling is varied. Abbreviations f.t.t.b: *Activ.* = activation function, *Loss* = loss function, *Optim.* = optimizer, *Q2/R2*= Q2/R2 calculator, *Scaling* = Scaling of the training and validation data set, *#BF* = number of basis function, *#L* = number of hidden layers, *#NL* = number of neurons per layers, *#E* = number of epochs, *LR* = learning rate, *BS* = batch size, *ME TD* = global model error training data, *ME VD* = global model error validation data, *Max. LE* = Maximum Local Error [%], *c* = complex-valued.

Parameters	Base	Activ.	Loss	Optim.	Q2/R2	Scaling
Activ.	Tanh	Tanh c	Tanh	Tanh	Tanh	Tanh
Loss	MSE	MSE	MSE c	MSE	MSE	MSE
Optim.	Adam	Adam	Adam	L-BFGS-B	Adam	Adam
Q2/R2	MSE	MSE	MSE	MSE	MSE c	MSE
Scaling	0-1	0-1	0-1	0-1	0-1	0-$\frac{1}{20}$
# BF	4					
# L	2					
# NL	50					
# E	20000					
# LR	$1.94 \cdot 10^{-4}$					
# BS	140					
ME TD	$2.62 \cdot 10^{-4}$	$3.74 \cdot 10^{-4}$	$2.62 \cdot 10^{-2}$	$7.70 \cdot 10^{-3}$	$2.62 \cdot 10^{-4}$	$6.68 \cdot 10^{-5}$
ME VD	$2.77 \cdot 10^{-4}$	$3.78 \cdot 10^{-4}$	$2.77 \cdot 10^{-4}$	$1.16 \cdot 10^{-2}$	$2.77 \cdot 10^{-4}$	$6.62 \cdot 10^{-5}$
Max. LE real	∞					
Max. LE imag.	∞					

With the current model set-up we can see that receiving complex-valued components is working, and some training is happening. However, the model does not reach the desired level of accuracy. This lets us assume that incorrect calculations happen during the surrogate model construction, where imaginary and real components are combined. This coupling can cause problems when both components are non-zero. For this reason, several components of the offline stage have been carefully tested, such as: (1) varying the suitability of the activation function and its specific use for complex values, (2) the optimizer, (3) the loss function in its calculations and its behavior over epochs, (4) the calculation of the global model error defined via R2/Q2, and (5) the pre-scaling of the training- and validation data set before constructing the surrogate model. In the following, we present these five approaches, and the issues they bring along. We keep one version with specific hyperparameters for the NN as a base, and then vary only one of them at a time to investigate the influence of this hyperparameter. The parameters of the base, and of the varying approaches are displayed in Table 3-4. To show variations in the scaling compared to the approach shown in Table 3-3, we introduce here a scaling for the complex values together. In the previous approach, the real and imaginary components are scaled separately, while here we scale the complex numbers in their entirety. To keep the consistency during the testing, we apply this scaling method to all six variations. No tuning of the hyperparameters of the NN is performed during any of these tests. It is to note that *tanh c* and *MSE c* (*c* standing for *complex*) represent the same calculation method as *tanh* and *MSE* but in the complex domain instead of the real one. These calculations are then done for the real and imaginary part separately, combining them to complex-valued values afterwards. The *L-BFGS-B* optimizer has the advantage of low costs, which is of advantage for large data sets (Zhu et al., 1997). It is generally well-working for smooth functions (Henao, 2014), as present in our case. *Adam* on the other side allows a better control of the step size for the update of the weighting coefficients (Zhang, 2018). The scaling for the training and validation data set is divided by 20 in order to have a matrix between -1 and 1 entering the NN step. This matrix results from the matrix multiplication of the training data set and the basis functions, and is, without this additional scaling by 20, between -20 and 20 in our data set. Having it below -1 and 1 might help the activation function to improve the training. For these reasons, the specific variations are tested.

From Table 3-4 we can conclude that none of the tests show the desired global accuracy, as we aim for a global model error of 10^{-7} for the scaled data set. The best result show a global model error of around 10^{-5}, which is obtained when adjusting the scaling. The optimizer *L-BFGS-B* leads to the largest global model error of 10^{-2}. The local error goes to infinity for all tests. Figure 3-9 shows the loss function, and the real and imaginary components of the impedance, for all six introduced tests. We can see that all surrogate models cannot approximate the GoFEM solutions. In all cases the real component is slightly better adapted to the true solution than the imaginary part, but both are far from our expectations. Looking at the loss functions (Figure 3-9-a) we can see their different behavior depending on the parameter changed. Overall, the loss functions decay very quickly, which leads to the assumption that the surrogate models are not trained, and the first model guess is accepted and given as an output. In test 4, where the optimizer *L-BFGS-B* is applied, the loss function does not decay at all, with the training data loss function fluctuates randomly and the validation data loss functions is constant. In the scaling test (6), we can see some fluctuation in the loss function, which means that some testing has been done. Besides the loss function test (3) displayed in Figure 3-9-3, also the complex MSE loss definition after (Bassey et al., 2021) has been tested. Due to strong fluctuations in the loss function and not meaningful results, the results are not visualized in this thesis.

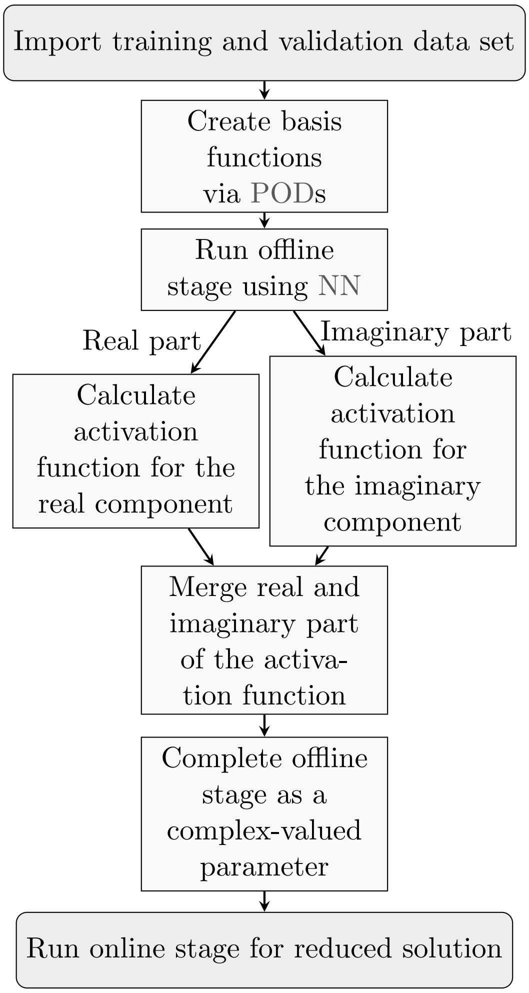

Figure 3-8: Workflow for splitting Z in real and imaginary part in the activation function of the NN.

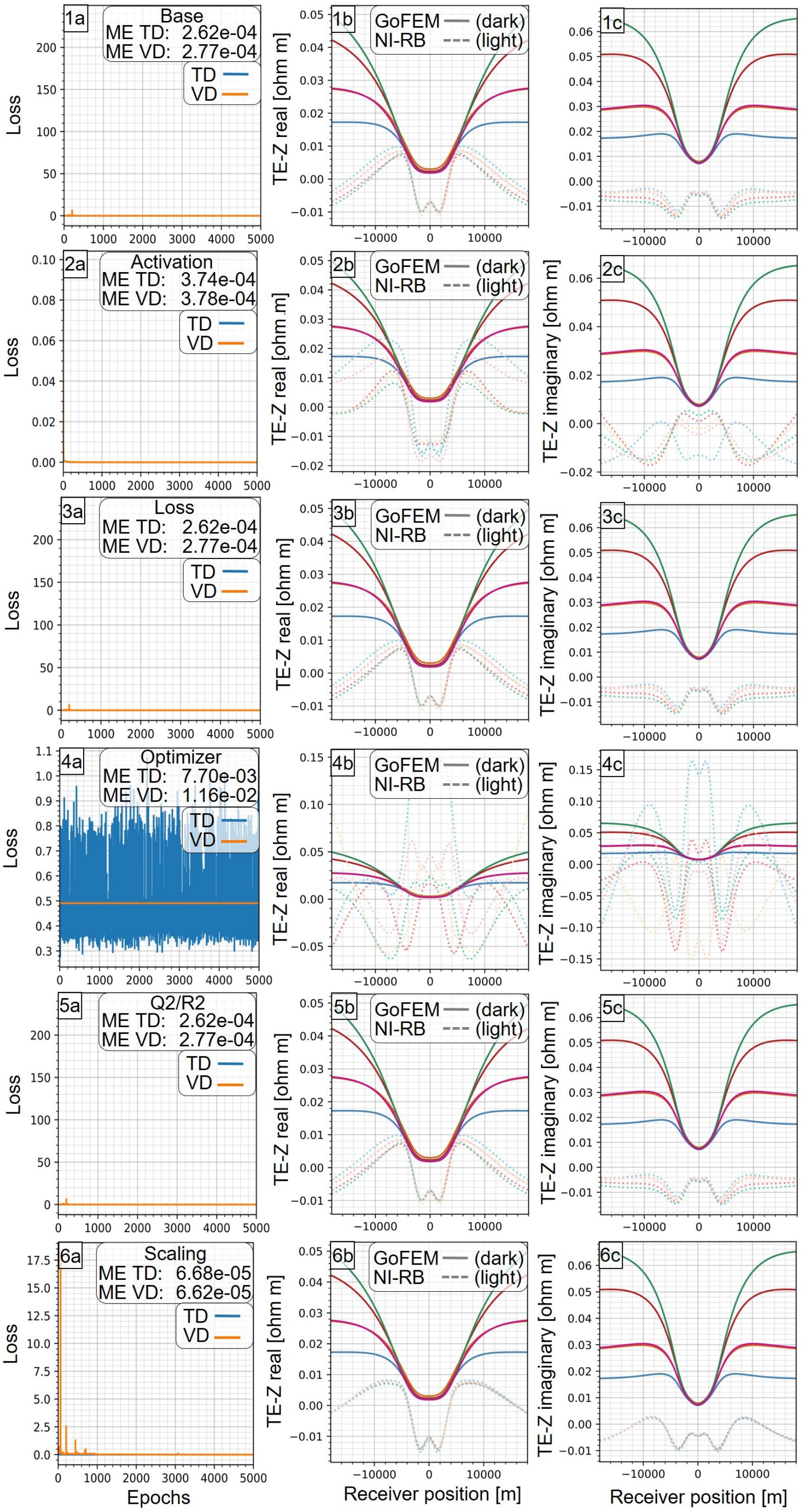

Figure 3-9: (Caption next page.)

Figure 3-9: (Previous page.) Representation of the accuracy of the surrogate model of a MT simulation of the impedance of a high-temperature geothermal system for 1 Hz, testing different architecture of the NN. The left side (-a) shows the loss function over the first 5000 epochs. The figures in the middle and at the right represent five realizations with different parameter values from the validation data set for the real (-b) and imaginary part (-c) of the impedance *Z*, respectively. Solid lines represent the GoFEM simulation, the dashed line the NI-RB solutions. All tests require four basis functions needed for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set, *ME TD* and *ME VD* demonstrates the global model error of the training data and validation data, respectively. The maximum local error goes to infinity in all figures. From top to bottom, we vary specific parameters in the NN step, following Table 3-4: base construction (1), activation function (2), loss function (3), optimizer (4), Q2/R2 calculations (5), and scaling (6).

We can conclude that the error source of the malfunctioning of the NN in the NI-RB method has not been found after this analysis. We assume, that in many cases the NN does not train the reduced model and assumes its first guess to be correct. We therefore cannot show any result of desired accuracy. The extensive testing of different parameters of the NN has not shown any path that could improve the model construction enough to decrease the discrepancy between the surrogate model and GoFEM sufficiently. For this reason, we decide for an alternative approach with better control over the construction of the surrogate model, where the GP regression is used. This is because the use of NNs has two disadvantages in the NI-RB method, which make analysis difficult: (1) NNs are computationally more expensive than the GP regression, making quick checks in code adaptions very time intensive. Each simulation takes minimum 30 minutes, while all stages GP regression are in the range of a few seconds. (2) NN give us less control of its behavior than GP regression due to the specific hyperparameters that can be adjusted, making error detection more challenging. For these reasons, we change our approach of implementing the complex-valued impedance using the NN, and we switch to the GP regression in the next section. We aim to find the error source potentially also affecting the current approach by splitting the impedance into its real and imaginary components after the POD, thus having complex basis functions.

3-2-3 Split after the Proper Orthogonal Decomposition

In order to detect potential errors sources in the previous approach using the NN, we take a step back and replace the NN by GP regression. This projection method allows a faster construction of surrogate models due to less hyperparameters to be trained, which is of advantage when improvements or retrogressions are desired to be detected quickly. However, our approach differs from the one described in Section 3-2-1, where we determined basis functions separately for the real and imaginary component. In this section, we calculate the basis functions for the complex-valued parameter in its entirety, thus we have one complex-valued set of basis functions, compared to two real-valued sets in Section 3-2-1. We split the real and imaginary part only after the POD, and then run the offline and online separately for each component. After the online stage, we merge the reduced solution together to obtain the impedance as complex parameter its entirety. This improves our model consistency, as only the coefficients are determined separately for each component of the complex values. Figure 3-10 shows the procedure of this approach graphically.

Table 3-5 reveals the model architecture as well as the errors divided for the real and imaginary component. The second and third columns represent the case of setting the imaginary part (1) and the real part to zero (2), respectively. In the fourth column, both components are unequal zero (3). The architecture is kept constant to enable direct comparison. We would like to draw attention to the following aspects:

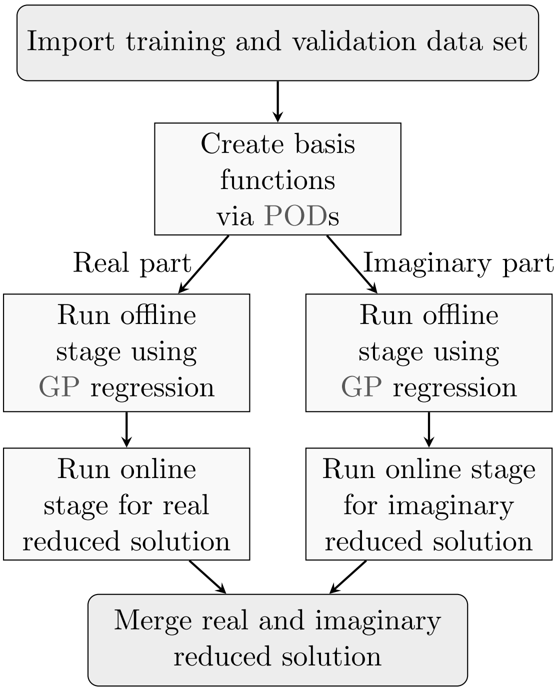

Figure 3-10: Workflow for splitting Z in real and imaginary part after PODs for the projection method of GP regression.

- In the case the imaginary part is set to zero (1), we obtain a *global model error of the real component* in a range which is not promising a result of our desired accuracy. This is also visible in the *maximum local error of the real part*, showing us some coherence but out of our threshold. The *global model error of the imaginary component* is zero, which represents an accurate fitting of the scaled data, while its *maximum local error* goes to infinity for the unscaled data set. This behavior is more easily understandable when visualized graphically, displayed in Figure 3-11.
- When the real part is set to zero (2), the *global model errors of the imaginary part* is in a similar range than the real part if the imaginary part is zero. Thus, we also do not expect a decent fit. The *maximum local error of the imaginary part* is even higher than the one of the real component if the imaginary part is zero.
- When both components are nonzero (3), we obtain the same values for all errors as for the real component in test (1) and for the imaginary component in test (2). Thus, neither the *global model error* nor the *maximum local error* is in our desired range.

Looking at the results graphically in Figure 3-11, we can see that overall the surrogate model does not fit the true solution from GoFEM in any of the three tests. Compared to the approach using NNs described in Section 3-2-2, the components that are nonzero in test (1) and (2) do not fit the true solution well. However, the parts that are set to zero are approached better than in the application of NNs. The surrogate model does not fit the GoFEM solution well in its values, but in its shape, as it is modelled as a horizontal straight line (highlighted with the orange box in Figure 3-11-1b and -2a). This is a better approximation than the wavy behavior when using NNs, as obtained in Section 3-2-2. Test (3) combines both non-zero components of test (1) and (2), thus not satisfying the predefined error tolerance.

Table 3-5: Parameters of the GP regression with an error threshold of $1 \cdot 10^{-4}$ for the scaled data set and a frequency of 1 Hz in case of splitting the impedance into its real and imaginary component after the POD. The second and third columns on the right describe, that first the imaginary and then the real component is set to zero, while in the fourth column both components are unequal zero. Abbreviations f.t.t.b: *LS* = length scale, *LSB* = length scale boundaries, *Nu* = smoothness factor, *#Ro* = number of restarts of optimizer, *Optim.* = optimizer, *#BF* = number of basis function, *ME TD* = global model error training data, *ME VD* = global model error validation data, *Max. LE* = maximum local error [%]. Values marked with $*$ are valid for all input parameters.

Parameters	Imaginary = 0	Real = 0	None = 0
LS	1*		
LSB	(1e-3,1e3)*		
Nu	2.5		
#RO	0		
Optim.	L-BFGS-B		
# BF	4	4	4
ME TD real	$3.25 \cdot 10^{-2}$	0	$3.25 \cdot 10^{-2}$
ME VD real	$3.48 \cdot 10^{-2}$	0	$3.48 \cdot 10^{-2}$
Max. LE real	32.56	∞	32.56
ME TD imag	0	$5.63 \cdot 10^{-2}$	$5.63 \cdot 10^{-2}$
ME VD imag	0	$6.11 \cdot 10^{-2}$	$6.11 \cdot 10^{-2}$
Max. LE imag	∞	57.3	57.3

We conclude from this approach that there must still be problems in the combination of the complex-valued basis functions with the model obtained from the GP regressions, since the approach is working for real-valued parameters (Section 3-2-1). However, this approach can help to solve the problems in accuracy when using NNs, as we have a higher transparency in the offline stage when using the GP regression. It is therefore of importance to be solved in the near future. This coupling problem does not seem to be solved until further extensive investigation is done. This investigation is not feasible within the time frame of this thesis and therefore is aimed to be addressed outside this research. In order to remove any issues related to the splitting, the last approach presented in this thesis is tested. It describes the transformation of the complex-valued impedance to real values, and is described in the next section.

3-2-4 Transformation from Complex to Real Values

In applications of complex-valued parameters, an approach can be used which transforms complex values to a real-valued matrix. We assume to have a matrix $\mathbf{W}$ of $m \cdot n$ elements in the complex domain. $\mathbf{W}$ can be split into $\mathbf{W}_r$ and $\mathbf{W}_i$, representing its real and imaginary values, respectively. We can then transform this complex-valued matrix

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_r & -\mathbf{W}_i \\ \mathbf{W}_i & \mathbf{W}_r \end{bmatrix}, \tag{3-1}$$

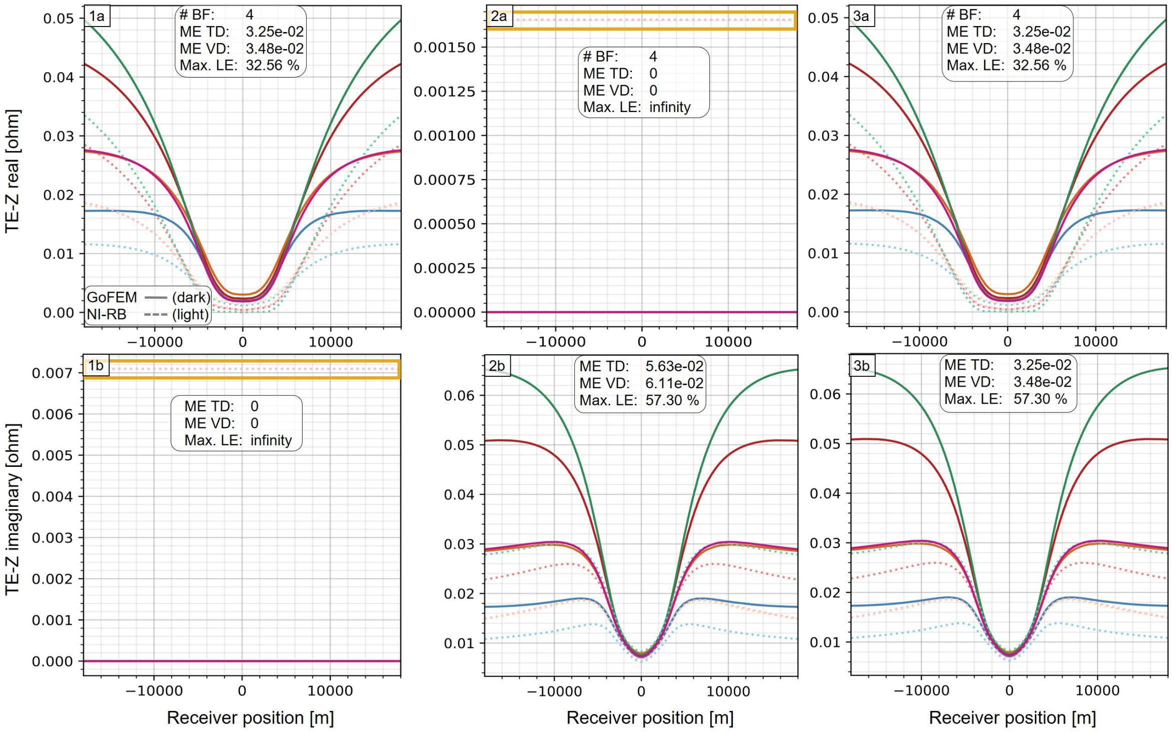

Figure 3-11: Representation of the accuracy of the surrogate model of the impedance in the TE mode of a high-temperature geothermal system for frequency of 1 Hz with the projection method of GP regression, with the split of real and imaginary parts after the POD. The upper figures show the real, and the lower ones the imaginary part of the impedance, respectively. The first (1) and second (2) columns represent the case where the imaginary or real part is set to zero, respectively. For the third column (3), both the real and imaginary part are unequal zero. Each figure includes five realizations with different parameter values from the validation data set. Solid lines represent the GoFEM simulation, the dashed line the NI-RB solutions. The orange box in Figure 1b and 2a highlights the NI-RB solution. *# BF* shows the number of basis functions needed for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set, *ME TD* and *ME VD* demonstrates the global model error of the training data and validation data, respectively, and *Max. LE* shows the maximum local error.

where the matrix on the right consists of $2m \cdot 2n$ elements in the real domain (Barwick, 2016). We apply this approach on the complex-valued MT simulations of the impedance in the NI-RB method. Figure 3-12 shows the workflow for this approach. The POD is performed on the complex-valued training data, as well as the matrix multiplication of the training data with the basis functions. Before the product of the matrix multiplication enters the GP regression, we transform it to a real-valued matrix following Barwick (2016). After the prediction of the coefficients in the online stage, we transform these values back to the complex domain. The reduced solution is thus evaluated from basis functions and coefficients both being part of the complex domain.

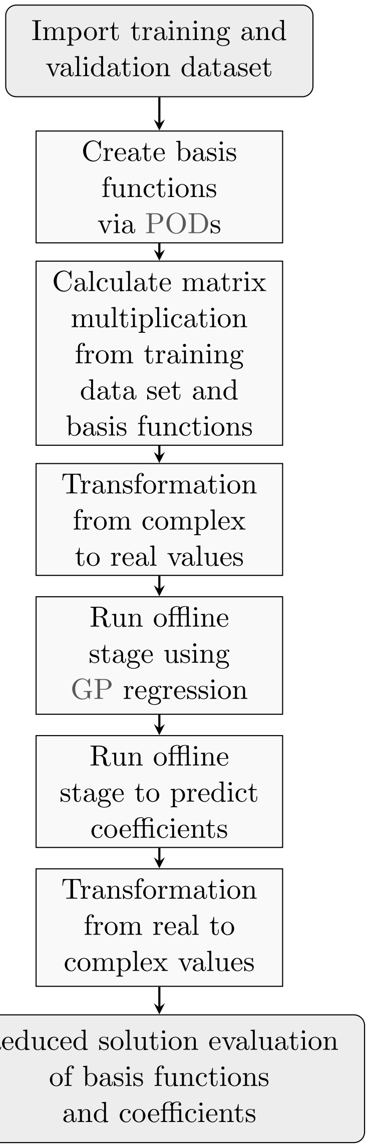

Figure 3-12: Workflow for the transformation of the complex-valued impedance to real values after the POD.

The advantage of this approach is that we obtain one reduced model for both the real and imaginary part together. This keeps our model consistency higher than in the previous approaches. When we split the impedance in real and imaginary components before and after the POD (Section 3-2-1 and 3-2-3), we loose some of the model consistency by having two model for the impedance. Also in Section 3-2-2 where we did not split the components consciously, the complex-valued neural network (CVNN) packages used in this approach splits the real and the imaginary part internally in the activation function. Therefore, even if that approach had delivered accurate results, we also would have lost model consistency with this procedure. With the approach of transforming the complex values to real ones, we include the coupling between both components of the impedance, require only one model and thus maintain a high model consistency. This is beneficial and could help being independent of CVNN libraries being correctly used for the NI-RB method.

For this approach, we use the GP regression as projection method, as it allows faster results due to less hyperparameters needed to be trained. We apply a length scale of one, and boundaries of $(1 \cdot 10^{-5}, 1 \cdot 10^{5})$ for the length scale of all input parameters, with a smoothness factor of 2.5. The number of restarts of the optimizer *L-BFGS-B* is zero. We obtain four basis functions from the POD. Figure 3-13 shows the results of this approach. We receive a decent result for both real and imaginary components, which maximum local error is 2.65 % and 2.86 %, respectively, exceeding slightly our desired threshold of 1 %. The right side of Figure 3-13 shows the differences between the solution of the surrogate model and the true solution of GoFEM. It is visible, that similar to the approaches in Sections 3-1 and 3-2-1 the largest deviations can be found at quick changes in values. This happens mostly at the receiver position around 1 m. During the set-up of this approach, we noticed that the conjugate of the coefficients needs to be taken when evaluating the surrogate model with the basis functions in the online stage to obtain these results. In the previous approaches of the impedance, the conjugate of the basis functions is taken instead. This shows a potential error source that might be the reason for some of the issues and inaccuracies encountered in Sections 3-2-2 and 3-2-3. Further testing is required to see the influence of changing the code after these findings.

The results displayed in Figure 3-13 show a better outcome than obtained in Section 3-2-2 and 3-2-3. However, the accuracy of the obtained results are still not in our desired tolerance for the maximum local error of 1 %, as both the real and imaginary component are above 2.5 %. Additionally, the internal score of the model is with 0.74 not ideal, as it should approximate 1 for decent calculations. It is also to note that the global model error of the validation data set is by two orders of magnitude smaller than the one of the training data set. Both are too large, but additionally it is unexpected that the validation data gives a better score than the training data. These facts lead to the assumption, that the calculations of the coefficients are not yet ideal. This could be caused by the dependencies between the real and imaginary part which is not present in the approach of Section 3-2-1, where both components are split before the POD. After the transformation of complex to real values they are both represented in one matrix, thus influencing each other. This coupling might arise problems in the the GP regression. It would be of high priority to test a different projection method as, e.g., NNs, because it can tackle a wider range of applications. In such a test, we could see if a different projection method improves the calculations of the coefficients, and thus also the global model error (and the local error as a consequence).

3-3 Computational Costs

One of the main advantages of using the NI-RB method is the strong decrease of computational costs. This is especially the case when numerous parameter sets are desired to be tested using the same surrogate model, as the costs of the online stage are much lower than of the offline stage, particularly once the model is loaded.

Table 3-6 reveals insights in the computational costs of each important step in the offline and online stage, for the surrogate model construction of the phase, the resistivity and the impedance of the NI-RB method. For comparison reasons, the costs for creating one single simulation in GoFEM is included, which contains the phase, resistivity and impedance automatically together. It is to note that for one surrogate model, the offline stage including POD and training does not need to be repeated, and as well the model only need to be loaded once, if all parameter studies are done in one go. When only considering repetitive steps of the NI-RB method, the prediction and the evaluation of the model, it provides an acceleration of 10'000 times for the impedance. This improvement comparing one forward simulation of the NI-RB method with respect to the true solution of GoFEM is called *speed-up.* The difference to the sum of the phase or resistivity, which can represent the impedance (Eq. (2-20)), and GoFEM is of similar magnitude. It needs to be considered, that the costs for the GoFEM simulation in this projects are determined for 49 frequencies. The cost for one frequency is determined by dividing the total costs of one GoFEM run by 49.

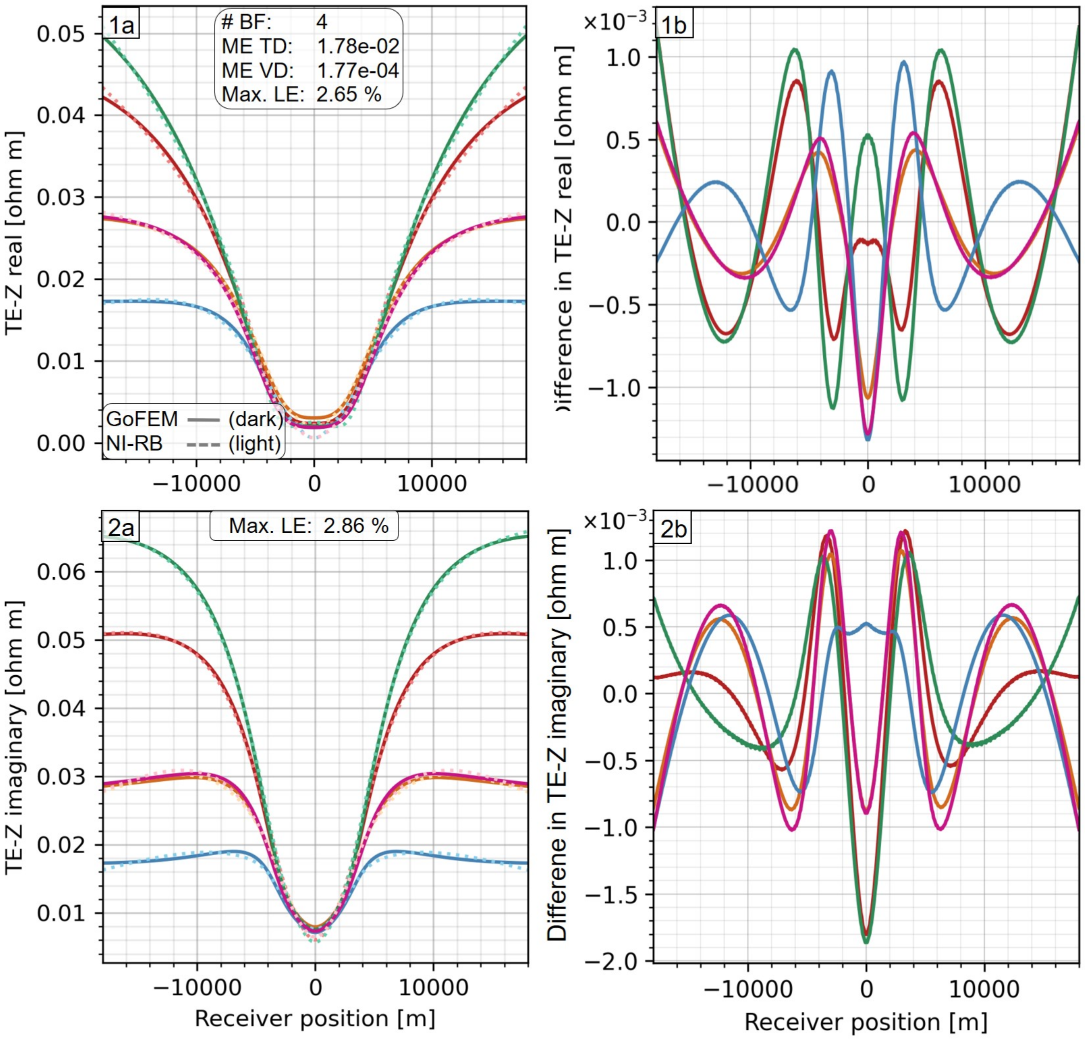

Figure 3-13: Representation of the accuracy of the surrogate model of a MT simulation of the impedance of a high-temperature geothermal system for frequency of $1\ \mathrm{Hz}$ with the projection method GP regression, where the complex-valued impedance is transformed to real values after the POD. The left side shows five realizations with different parameter values from the validation data set for the real (1a) and imaginary part (2a) of the impedance *Z*. Solid lines represent the GoFEM simulation, the dashed line the NI-RB solutions. *# BF* shows the number of basis functions needed for an error threshold of $1 \cdot 10^{-4}$ for the scaled data set, *ME TD* and *ME VD* demonstrates the global model error of the training data and validation data, respectively, and *Max. LE* shows the maximum local error. The right side shows the difference between the NI-RB solution and the GoFEM simulation of each realization on the left.

We should consider that the difference between the overall costs of the NI-RB method and GoFEM increases significantly with the number of simulations performed per surrogate model. This is because the costs of GoFEM grow linearly with the number of simulations calculated, while the ones of the NI-RB method reach a plateau once the model is constructed. Therefore, the speed-up is a frequently used factor to describe the enhancement in terms of computational costs, considering the constructed model is evaluated multiple times. Overall, we can conclude that the use of the NI-RB method is of large advantage with regards to computational costs.

Table 3-6: Computational costs of each important step in the NI-RB method for the MT response. This table includes the comparison of the computational costs of the phase resistivity, the impedance of the split approach before the POD (Section 3-2-1), and GoFEM, for a single simulation. All values are in ms, measured with a computer of $8\,\mathrm{RAM}$, a frequency of the processor of $2.6\,\mathrm{GHz}$ with 2 cores. Values marked with $*$ are obtained by division or multiplication by/with the entire range of 49 frequencies.

	Step	Phase [ms]	Resistivity [ms]	Impedance [ms]
Offline Stage	Forward Sim. GoFEM	$6.74 \cdot 10^{3*}$		
	POD	10.1	10.1	17.8
	GP regression	387	363	253
Online Stage	Load Model	161	1.73	3.48
	Prediction	0.204	0.203	0.54
	Evaluation	0.00975	0.00924	0.0324
Total NI-RB: Predict + Evaluate	1 freq.	0.213	0.212	0.572
	49 freq.	10.47*	10.40*	28.12*
Total GoFEM	1 freq.	$6.74 \cdot 10^{3*}$		
	49 freq.	$3.30 \cdot 10^{5}$		

Chapter 4

Discussion

Let us remind ourselves to the research question posed in the introduction in Chapter 1:

> *How can we apply the NI-RB method to the real-valued phase and resistivity, as well as the complex-valued impedance of the MT responses, in order to stay below the desired maximum local error of 1 %?*

After conducting extensive research within the scope of this project, we can now address and answer this initial research question. The NI-RB method can be used for modelling resistivity and phase with only little adjustments, which are mostly related to the structure of the GoFEM simulations. Scaling and testing the hyperparameters help to improve the precision. The desired accuracy with a local error below 1 % is achieved with both NNs and the GP regression as projection methods, with the GP regression significantly reducing the computational costs due to fewer hyperparameters requiring training. As part of this project, the impedance can be accurately modelled with the NI-RB method when its real and imaginary component are treated separately for the construction of the basis functions as well as for the determination of the coefficients. Two completely separated models are created, to which new parameters can be applied in the online stage. Merging these two solutions results in the final, complex-valued approximation of the impedance. A second approach that gives feasible results for the impedance is the transformation of the complex-valued impedance into a real-valued matrix. However, it causes a maximum local error of 2.5 %, which exceeds our predefined tolerance. In this thesis, this procedure has been tested with the GP regression as the projection method, resulting in a model showing difficulties in its calculations. As these could arise from the dependency of the real and imaginary components, better achievements could eventually be obtained by using NNs. However, to validate this statement further analyses are required. The other two approaches with the application of the NN where the split is performed inside the activation function, as well as the one with the split after the POD could not result in favorable outcomes during this project.

Despite the challenges that have occurred along this thesis, the research shows several advantages and positive outcomes. The NI-RB method proves to be successful for the resistivity and phase, as well as for the impedance when the data is split before the POD. This approach allows the accurate construction of surrogate models for these parameters of the MT response, while achieving the desired error threshold of below 1%. It is important to notice, that the problems faced in this project concerning the impedance are only implementation problems. In theory, both NNs and GP regressions are approaches with the potential to work effectively with complex-valued data (Hirose, 2012). Especially the last approach of transforming complex values of the impedance into real ones looks promising, as the application of NNs instead of the GP regression might help to obtain the desired accuracy. NNs have the advantage to tackle the potential non-linear behavior caused by the dependency between the real and imaginary components (Degen et al., 2022). Another benefit arising from this procedure is that we could find an error source concerning the evaluation of the reduced solution, which might help to improve the other two currently malfunctioning, approaches. Further analyses are required to see the impact of the code correction. For the successful approaches of the resistivity and phase, as well as the impedance with the initial split and its transformation to real values, the difference between the true solution of GoFEM and the approximated solution of the NI-RB method exhibits smooth and consistent patterns. This leads to the assumption that the majority of the inaccuracies included in the NI-RB method arises from the truncation of the basis functions. Knowing this probable error source is helpful for interpreting the methods accuracy, and also to improve it further at a later stage, if required. This smooth pattern proves the reliable estimations achieved by the NI-RB method for the parameters of interest. Furthermore, the NI-RB method allows a significant advantage in terms of computational efficiency in comparison to GoFEM. It enables a speed-up of 10'000, when purely considering repetitive steps as the prediction of the coefficients and the evaluation of the model. This is the speed-up for the impedance, while it is of similar magnitude for the phase and resistivity. It is to take into account that this factor excludes the time required for the forward simulations needed for the surrogate model construction, which is very cost-intensive. However, these simulations only need to be calculated once for an indefinite number of models. When including the offline stage, the computational benefit of the NI-RB method increases with the growing number of simulations desired from it, as the higher costs of the offline stage decrease relatively. These estimations of the computational costs show overall the efficiency of the NI-RB method, reducing the computational costs drastically.

During this thesis, the comparison to the intrusive RB method has been made. Even though the NI-RB method is used in this project, the RB method also could have been applied. The relation between the electrical conductivity and the impedance is not highly non-linear (Equation 2-20). The RB method tends to have difficulties with highly non-linear problems, as it then lacks in accuracy and convergence (Chen et al., 2021). However, the problem tackled here is not strongly non-linear, thus it could have been tested with the RB method. On the other side, the NI-RB method allows a wider application range, as simulation results of diverse forward solvers can directly be used in the NI-RB method (Degen et al., 2022). For the intrusive RB method, the method itself needs to be implemented in the forward solver, making its use more complex. Additionally, non-linear and hyperbolic problems can be approached with the NI-RB method, in case the complexity of the MT simulations increases, leading to a higher flexibility.

Based on the current research, there are several key aspects of this research which can potentially have a positive impact on the MT society. First, our methodology offers a significant advantage compared to previous studies of using RB methods in EM. Unlike traditional approaches focusing on training the potential, we directly construct the surrogate models for the phase, resistivity and impedance. Thus, we train directly on parameters of interest for MT data interpretation, allowing results to be close to the application. By focusing on the impedance using the NI-RB method, we enhance the applicability of this approach to actual field measurements. As the NI-RB method accelerates simulations by nearly 10'000 times for the impedance once the surrogate model is constructed, it opens possibilities for 3D applications. Manassero et al. (2020) demonstrated that RB methods can be applied to resistivity and phase, in order to enable 3D MT applications. In this research we show that the NI-RB method gives the chance to create such 3D models for the impedance. Additionally, the NI-RB method contributes to probabilistic approaches, such as Markov chain Monte Carlo (MCMC) simulations (Wellmann and Caumon, 2018). By creating different responses from one surrogate model by inserting variations in electrical conductivities, we can use use these responses in such a probabilistic approach, testing the fit of the field data to any conductivity combination. Then, e.g., a MCMC can be run to explore the parameter space and find the best fitting simulation (Wellmann and Caumon, 2018). The NI-RB method can create these thousands of simulations required for such an approach in an efficient way, e.g. for the impedance.

Nevertheless, we also need to consider the multiple challenges that occurred during this research. Splitting the impedance in real and imaginary components especially before, but also after the POD, is undesired, as it decreases the model consistency. However, we would like to emphasize that the real and imaginary parts would also be treated separately for the application using NNs with the current available python packages and modules for the NI-RB method. Although this approach seems to come closest to treating the impedance as complex-valued, it introduces inconsistency by applying the activation function to each component of the complex-valued input separately, combining them afterwards. This inconsistency arises at each neuron in the NN. The advantage of this procedure is that the basis functions are determined for the complex-valued parameter in its entirety. However, this is also fulfilled in the approach of splitting the impedance after the POD with the GP regression as a projection method, as explained in Section 3-2-3. Thus, as it maintains the model consistency similarly to the approach of the NN. However, also the split after the POD seems problematic, as the basis functions are calculated for the complex values in their entirety, but the training happens on each part separately. This seems to be the error source, leading to surrogate models of lacking accuracy. The approach of transforming the complex values to real ones manages to keep the model consistency high, as only one surrogate model is constructed for both the real and imaginary component together. This could be a promising solution to enable a construction of a surrogate model without relying on CVNN libraries and its implementation into the NI-RB method. From all tested approaches, it promises the highest model consistency and would therefore be the desired procedure. However, to achieve a desired accuracy below 1 %, as it is now above 2.5 %, we would like to test the use of NNs instead of the GP regression, as the dependency of the real and imaginary components might cause problems in the current projection method. This needs to be tested in the near future.

Additionally to the challenges faced during this thesis, we would like to reveal the limitations encountered during this project. We introduce them chronologically to the approach of the NI-RB method. First, as we use MT, we underlie all assumptions caused by this geophysical method. They include the plane wave assumption, the location of the source for EM fields, the magnetic permeability and electric permittivity to be constant, and the neglect of displacement functions. More details can be found in Chapter 2-2-3. Secondly, the model created in this project includes the second limitation, as it has a fixed geometry. This does not represent reality, because geothermal systems have different shapes and varying depths. Thus, testing different shapes of the bodies at varying depth, would be of great interest. The third limitation focuses on the simulated parameters. The results of the simulations from GoFEM, as well as the constructed model from the NI-RB method, show the apparent resistivity. The apparent resistivity represents the resistivity of the subsurface, assuming it is a homogeneous half-space. However, in reality, the underground is often heterogeneous, and our specific model aims to investigate the MT response in regions of varying physical properties. Therefore, we would need to invert the apparent to the true resistivity for results representing the reality. This inversion process could also benefit from the application of the NI-RB method. However, it is worth noting that the apparent resistivity still holds relevance, especially if the goal of the NI-RB method is to be compared with field measurements, where the apparent resistivity is indirectly measured. Another limitation is the focus on one frequency in each model. In the current state, a single model is needed to be constructed for each frequency, which might be of difficulties when desiring a dense frequency spacing from milli- to kilohertz, which is typical for MT. However, this could be adapted when creating a loop over around 50 frequencies, which then automatically creates a model for each frequency. As the models do not represent large-sized files, this can be a solution. Another point is that the simulations, as they are now, are defined for a fixed time and thus not of transient nature. It would be interesting to see changes of the reservoir over time. Time-variant implementations already exist in the NI-RB method (Degen et al., 2023), thus that would be clearly an interesting option. The last limitation is about the use of NI-RB methods for the impedance. As already mentioned, a model construction based on the complex-valued impedance in its entirety is currently not possible, as there is a lack of python modules and packages which enable the correct calculations for this specific application. However, taking into account the error source of the conjugate found in the transformation approach, we might be able to improve our results. To the current state of the project, we require two surrogate models for the impedance, which reduces the model consistency. The approach of transforming the complex to real values allows having only one model and therefore a higher model consistency than any other procedure attempted in this thesis, however its accuracy needs to be further increased.

Chapter 5

Outlook and Conclusion

As we approach the conclusion of this project, even more questions and ideas have arose than were present to the begin of this research. Hence, we would like to present possible future directions for this project, based on the findings and limitations discussed in Section 4. Of course, we would first highlight the importance of finalizing the last three approaches of the impedance, in which the split is done inside the NN (Section 3-2-2), after the POD (Section 3-2-3) and where the impedance is transformed into a real-valued matrix (Section 3-2-4). As we have found a potential error source in the evaluation of the surrogate model during the approach of Section 3-2-4 that could additionally improve the results of the previous two approaches, further analysis is required. However, it needs to be considered if these approaches are of interest compared to the transformation, as the latter allows a higher model consistency. Furthermore, we would like to test the application of NNs instead of the currently used GP regression during the transformation, to see if that increases the accuracy to be above 99 %. As this approach might be of advantage compared to the others due to its high model consistency throughout the process, this would be of priority. As a consequence, it would then be very interesting to perform a consistency analysis. The aim would be to test the consistency between the real and imaginary part of the impedance, depending on the location of the split. This would help to better understand the importance of treating impedance as a complex-valued parameter in its entirety, of splitting the impedance after defining the basis functions, as well as the effect of the transformation to real values. Zorin et al. (2020) shows how this consistency test can be performed for MT applications. Additionally, the model's geometry could be changed to detect variations in the accuracy of the NI-RB method with increasing model complexity. Altering the layout of the reservoir can allow the application of a bayesian model selection. There, different models of varying geometries are evaluated based on their suitability to the measured field data. Thus we could have several geometry models, and run sets of electrical conductivities on each of them. By comparing these simulations with actual measured data, we can be guided to the geometry explaining the field data optimally. Another approach of implementing variations in the model layout is the geometry parametrization. It would be advantageous to include the geometry in the surrogate model as a parameter, such as the electrical conductivity is used in this research. It would make the surrogate model more flexible to test various depths of the geothermal reservoir, as well as different shapes of it. This would also be of interest for to compare simulations with MT field data, as then the geometry could be tested continuously. This would be different from the approach of bayesian model selection, which would test geometries discretely with a specific surrogate model for each predefined geometry. Additionally, one could test the computational efficiency of NI-RB on a 3D model to quantify the computational costs of the construction and evaluation of surrogate models in a 3D context. This approach could then be compared to performances of the RB method, which is nowadays used for the phase and resistivity (Manassero et al., 2020). A final suggestion for the continuation of this project is to test the influence of the four bodies defined in the geothermal system in this project on the MT response. As creating simulations is of low computational costs particularly after the construction of the surrogate model, a global SA can be performed in an efficient manner. It would reveal details about the body with the largest and smallest impact on the MT response. More information about SAs can be found in the appendix B-4.

In conclusion, based on the conducted research, we can confirm the successful use of the NI-RB method for the real-valued resistivity and phase of the MT response, meeting the desired error threshold of below 1%. The implementation of the complex-valued impedance encounters certain challenges and limitations. The most promising approach is the transformation of the complex-valued impedance into a real-valued matrix as it enables a high model consistency. However, it currently does not meet our desired accuracy by a small margin. This can eventually be solved by replacing the GP regression with NNs. Additionally, this approach has revealed a error source occurring during the online stage. Implementing this gained knowledge in the approach of the NN (Section 3-2-2) and of the split after the POD (Section 3-2-3) might help to decrease their current issues, as currently their results are not suitable for our application due to low accuracy.

Through our research, we have faced several unsuccessful approaches, which, nevertheless, have provided us valuable learning experiences. These attempts have allowed insights into areas that require further testing and modifications. By doing so, along with good teamwork and a fresh mindset, we can encounter solutions that could not be discovered until this stage. Then, through the process of research we should always keep the words of Manson (2016) in our mind, stating that

> *True happiness occurs only when you find the*
> *problems you enjoy having and enjoy solving.*

Bibliography

Abdi, H. (2007). Singular value decomposition (svd) and generalized singular value decomposition. *Encyclopedia of measurement and statistics*, 907:912.

Agemar, T., Weber, J., and Schulz, R. (2014). Deep geothermal energy production in germany. *Energies*, 7(7):4397–4416.

Agyemang, V. O. (2022). Groundwater exploration by magnetotelluric method within the birimian rocks of mankessim, ghana. *Applied Water Science*, 12(3):26.

Ahmed, N., Assadi, M., Ahmed, A. A., and Banihabib, R. (2023). Optimal design, operational controls, and data-driven machine learning in sustainable borehole heat exchanger coupled heat pumps: Key implementation challenges and advancement opportunities. *Energy for Sustainable Development*, 74:231–257.

Ahn, S., Couture, S. V., Cuzzocrea, A., Dam, K., Grasso, G. M., Leung, C. K., McCormick, K. L., and Wodi, B. H. (2019). A fuzzy logic based machine learning tool for supporting big data business analytics in complex artificial intelligence environments. In *2019 IEEE international conference on fuzzy systems (FUZZ-IEEE)*, pages 1–6. IEEE.

Alhamid, M. I., Daud, Y., Surachman, A., Sugiyono, A., Aditya, H., Mahlia, T., et al. (2016). Potential of geothermal energy for electricity generation in indonesia: A review. *Renewable and Sustainable Energy Reviews*, 53:733–740.

Aretouyap, Z., Nouck, P. N., and Nouayou, R. (2016). A discussion of major geophysical methods used for geothermal exploration in africa. *Renewable and Sustainable Energy Reviews*, 58:775–781.

Asaue, H., Kubo, T., Yoshinaga, T., and Koike, K. (2012). Application of magnetotelluric (mt) resistivity to imaging of regional three-dimensional geologic structures and groundwater systems. *Natural resources research*, 21:383–393.

Asher, M. J., Croke, B. F., Jakeman, A. J., and Peeters, L. J. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resources Research*, 51(8):5957–5973.

Azeez, K. A. and Harinarayana, T. (2007). Magnetotelluric evidence of potential geothermal resource in puga, ladakh, nw himalaya. *Current science*, pages 323–329.

Azeez, K. A., Kumar, T. S., Basava, S., Harinarayana, T., and Dayal, A. (2011). Hydrocarbon prospects across narmada–tapti rift in deccan trap, central india: Inferences from integrated interpretation of magnetotelluric and geochemical prospecting studies. *Marine and Petroleum Geology*, 28(5):1073–1082.

Baba, K., Utada, H., Goto, T.-n., Kasaya, T., Shimizu, H., and Tada, N. (2010). Electrical conductivity imaging of the philippine sea upper mantle using seafloor magnetotelluric data. *Physics of the Earth and Planetary Interiors*, 183(1-2):44–62.

Bahadori, A., Zendehboudi, S., Zahedi, G., et al. (2013). A review of geothermal energy resources in australia: current status and prospects. *Renewable and Sustainable Energy Reviews*, 21(0):29–34.

Bai, Z., Dewilde, P. M., and Freund, R. W. (2005). Reduced-order modeling. *Handbook of numerical analysis*, 13:825–895.

Baillieux, P., Schill, E., Abdelfettah, Y., and Dezayes, C. (2014). Possible natural fluid pathways from gravity pseudo-tomography in the geothermal fields of Northern Alsace (Upper Rhine Graben). *Geothermal Energy*, 2(1):1–14.

Barbier, E. (2002). Geothermal energy technology and current status: An overview. *Renewable and Sustainable Energy Reviews*, 6:3–65.

Barwick, C. (2016). Complex matrices. *Massachusetts Institute of Technology*, Lecture 29.

Bassey, J., Qian, L., and Li, X. (2021). A survey of complex-valued neural networks. *arXiv preprint arXiv:2101.12249*.

Basu, P. K. and Dhasmana, H. (2022). Electromagnetic theory fundamentals. In *Electromagnetic Theory*, pages 1–10. Springer.

Bathe, K.-J. (2007). Finite element method. *Wiley encyclopedia of computer science and engineering*, pages 1–12.

Becken, M., Ritter, O., and Burkhardt, H. (2008). Mode separation of magnetotelluric responses in three-dimensional environments. *Geophysical Journal International*, 172(1):67–86.

Bell, J. (2022). What is machine learning? *Machine Learning and the City: Applications in Architecture and Urban Design*, pages 207–216.

Benner, P. and Hess, M. (2017). Reduced basis approximations for maxwell's equations in dispersive media. *Model Reduction of Parametrized Systems*, pages 107–119.

Berdichevsky, M., Dmitriev, V., and Pozdnjakova, E. (1998). On two-dimensional interpretation of magnetotelluric soundings. *Geophysical Journal International*, 133(3):585–606.

Bertrand, E., Caldwell, T., Hill, G., Bennie, S., and Soengkono, S. (2013). Magnetotelluric imaging of the ohaaki geothermal system, new zealand: Implications for locating basement permeability. *Journal of Volcanology and Geothermal Research*, 268:36–45.

Bianchi, C. and Meloni, A. (2007). Natural and man-made terrestrial electromagnetic noise: an outlook. *Annals of geophysics*.

Bibby, H., Caldwell, T., and Brown, C. (2005). Determinable and non-determinable parameters of galvanic distortion in magnetotellurics. *Geophysical Journal International*, 163(3):915–930.

Bignall, G., Dorj, P., Batkhishig, B., and Tsuchiya, N. (2005). Geothermal resources and development in mongolia: country update. In *Proceedings of the World Geothermal Congress*.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

Brown, G. (2012). The inaccessible earth: an integrated view to its structure and composition. *Springer Science & Business Media*.

Cai, S., Mao, Z., Wang, Z., Yin, M., and Karniadakis, G. E. (2021). Physics-informed neural networks (pinns) for fluid mechanics: A review. *Acta Mechanica Sinica*, 37(12):1727–1738.

Cao, L. (2022). Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)*, 55(3):1–38.

Caulk, R. A., Ghazanfari, E., Perdrial, J. N., and Perdrial, N. (2016). Experimental investigation of fracture aperture and permeability change within enhanced geothermal systems. *Geothermics*, 62:12–21.

Chatterjee, A. (2000). An introduction to the proper orthogonal decomposition. *Current science*, pages 808–817.

Chave, A. D. and Jones, A. G. (2012). *The Magnetotelluric Method: Theory and practice*. Cambridge University Press, Cambridge.

Chave, A. D. and Smith, J. T. (1994). On electric and magnetic galvanic distortion tensor decompositions. *Journal of Geophysical Research: Solid Earth*, 99(B3):4669–4682.

Chen, W., Wang, Q., Hesthaven, J. S., and Zhang, C. (2021). Physics-informed machine learning for reduced-order modeling of nonlinear problems. *Journal of computational physics*, 446:110666.

Chowdhary, K. (2020a). *Fundamentals of artificial intelligence*. Springer.

Chowdhary, K. (2020b). Natural language processing. *Fundamentals of artificial intelligence*, pages 603–649.

Chuang, P.-Y. and Barba, L. A. (2022). Experience report of physics-informed neural networks in fluid simulations: pitfalls and frustration. *arXiv preprint arXiv:2205.14249*.

Degen, D. and Cacace, M. (2021). Effects of transient processes for thermal simulations of the Central European Basin. *Geoscientific Model Development*, 14(3):1699–1719.

Degen, D., Cacace, M., and Wellmann, F. (2022). 3d multi-physics uncertainty quantification using physics-based machine learning. *Scientific Reports*, 12(1):17491.

Degen, D., Caviedes Voullième, D., Buiter, S., Hendriks Franssen, H.-J., Vereecken, H., González-Nicolás, A., and Wellmann, F. (2023). Perspectives of physics-based machine learning for geoscientific applications governed by partial differential equations. *Geoscientific Model Development Discussions*, 2023:1–50.

Degen, D., Spooner, C., Scheck-Wenderoth, M., and Cacace, M. (2021a). How biased are our models? - A case study of the alpine region. *Geoscientific Model Development*, 14(11):7133–7153.

Degen, D., Veroy, K., Freymark, J., Scheck-Wenderoth, M., Poulet, T., and Wellmann, F. (2021b). Global sensitivity analysis to optimize basin-scale conductive model calibration–a case study from the upper rhine graben. *Geothermics*, 95:102143.

Deichmann, N. and Giardini, D. (2009). Earthquakes induced by the stimulation of an enhanced geothermal system below basel (switzerland). *Seismological Research Letters*, 80(5):784–798.

Delory, G., Grimm, R., Nielsen, T., and Farrell, W. (2007). Prospecting for subsurface liquid water using magnetotellurics on mars. In *Seventh International Conference on Mars*, volume 1353, page 3293.

Efendiev, Y., Galvis, J., and Wu, X.-H. (2011). Multiscale finite element methods for high-contrast problems using local spectral basis functions. *Journal of Computational Physics*, 230(4):937–955.

El Naqa, I. and Murphy, M. J. (2015). *What is machine learning?* Springer.

Elders, W. and Moore, J. (2016). Geology of geothermal resources. *Geothermal Power Generation*, pages 7–32.

Ellsworth, W. L., Giardini, D., Townend, J., Ge, S., and Shimamoto, T. (2019). Triggering of the pohang, korea, earthquake (m w 5.5) by enhanced geothermal system stimulation. *Seismological Research Letters*, 90(5):1844–1858.

English, J. M., English, K. L., Dunphy, R. B., Blake, S., Walsh, J., Raine, R., Vafeas, N. A., and Salgado, P. R. (2023). An overview of deep geothermal energy and its potential on the island of ireland. *First Break*, 41(2):33–43.

Farghali, M., Osman, A. I., Mohamed, I. M., Chen, Z., Chen, L., Ihara, I., Yap, P.-S., and Rooney, D. W. (2023). Strategies to save energy in the context of the energy crisis: a review. *Environmental Chemistry Letters*, pages 1–37.

Farquharson, N., Schubert, A., and Steiner, U. (2016). Geothermal energy in munich (and beyond) a geothermal city case study. *GRC Transactions*, 40:189–196.

Fetzer, J. H. and Fetzer, J. H. (1990). *What is Artificial Intelligence?* Springer.

Gao, J., Zhang, H., Zhang, S., Chen, X., Cheng, Z., Jia, X., Li, S., Fu, L., Gao, L., and Xin, H. (2018). Three-dimensional magnetotelluric imaging of the geothermal system beneath the gonghe basin, northeast tibetan plateau. *Geothermics*, 76:15–25.

Geuzaine, C. and Remacle, J.-F. (2009). Gmsh: A 3-d finite element mesh generator with built-in pre-and post-processing facilities. *International journal for numerical methods in engineering*, 79(11):1309–1331.

Gómez-Treviño, E., Muñiz, Y., Cuellar, M., and Calderón-Moctezuma, A. (2018). Invariant te and tm magnetotelluric impedances: application to the bc87 dataset. *Earth, Planets and Space*, 70(1):1–14.

Grayver, A. V. and Kolev, T. V. (2015). Large-scale 3D geoelectromagnetic modeling using parallel adaptive high-order finite element method. *Geophysics*, 80(6):E277–E291.

Grimm, R., Nguyen, T., Persyn, S., Phillips, M., Stillman, D., Taylor, T., Delory, G., Turin, P., Espley, J., Gruesbeck, J., et al. (2021). A magnetotelluric instrument for probing the interiors of europa and other worlds. *Advances in Space Research*, 68(4):2022–2037.

Guo, M. and Hesthaven, J. S. (2018). Reduced order modeling for nonlinear structural analysis using gaussian process regression. *Computer methods in applied mechanics and engineering*, 341:807–826.

Gupta, N. et al. (2013). Artificial neural network. *Network and Complex Systems*, 3(1):24–28.

Hamet, P. and Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69:S36–S40.

Harinarayana, T., Azeez, K. A., Murthy, D., Veeraswamy, K., Rao, S. E., Manoj, C., and Naganjaneyulu, K. (2006). Exploration of geothermal structure in puga geothermal field, ladakh himalayas, india by magnetotelluric studies. *Journal of Applied Geophysics*, 58(4):280–295.

Hayward, K. J. and Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, 17(2):209–233.

Heise, W., Caldwell, T., Bertrand, E., Hill, G., Bennie, S., and Palmer, N. (2016). Imaging the deep source of the rotorua and waimangu geothermal fields, taupo volcanic zone, new zealand. *Journal of Volcanology and Geothermal Research*, 314:39–48.

Henao, W. (2014). An l-bfgs-b-ns optimizer for non-smooth functions. *Master's thesis.*

Hermance, J. and Peltier, W. (1970). Magnetotelluric fields of a line current. *Journal of Geophysical Research*, 75(17):3351–3356.

Hess, M. W. and Benner, P. (2013). Fast evaluation of time–harmonic maxwell's equations using the reduced basis method. *IEEE Transactions on Microwave Theory and Techniques*, 61(6):2265–2274.

Hesthaven, J. S. and Ubbiali, S. (2018). Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78.

Hill, G. J., Bibby, H. M., Ogawa, Y., Wallin, E. L., Bennie, S. L., Caldwell, T. G., Keys, H., Bertrand, E. A., and Heise, W. (2015). Structure of the tongariro volcanic system: Insights from magnetotelluric imaging. *Earth and Planetary Science Letters*, 432:115–125.

Hirose, A. (2012). *Complex-valued neural networks*, volume 400. Springer Science & Business Media.

Hoversten, G. M., Myer, D., Key, K., Alumbaugh, D., Hermann, O., and Hobbet, R. (2015). Field test of subbasalt hydrocarbon exploration with marine controlled source electromagnetic and magnetotelluric data. *Geophysical prospecting*, 63(5):1284–1310.

Hübert, J., Whaler, K., and Fisseha, S. (2018). The electrical structure of the central main ethiopian rift as imaged by magnetotellurics: implications for magma storage and pathways. *Journal of Geophysical Research: Solid Earth*, 123(7):6019–6032.

Huntington, D. and Lyrintzis, C. (1998). Improvements to and limitations of latin hypercube sampling. *Probabilistic engineering mechanics*, 13(4):245–253.

Imamura, N. and Schultz, A. (2020). Quality estimation of magnetotelluric impedance tensors using neural networks. *The Leading Edge*, 39(10):702–710.

Ishizu, K., Ogawa, Y., Mogi, T., Yamaya, Y., and Uchida, T. (2021). Ability of the magnetotelluric method to image a deep conductor: Exploration of a supercritical geothermal system. *Geothermics*, 96:102205.

Jolie, E., Scott, S., Faulds, J., Chambefort, I., Axelsson, G., Gutiérrez-Negrín, L. C., Regenspurg, S., Ziegler, M., Ayling, B., Richter, A., et al. (2021). Geological controls on geothermal resources for power generation. *Nature Reviews Earth & Environment*, 2(5):324–339.

Kabeyi, M. J. B. (2019). Geothermal electricity generation, challenges, opportunities and recommendations. *International Journal of Advances in Scientific Research and Engineering (ijasre)*, 5(8):53–95.

Kana, J. D., Djongyang, N., Raïdandi, D., Nouck, P. N., and Dadjé, A. (2015). A review of geophysical methods for geothermal exploration. *Renewable and Sustainable Energy Reviews*, 44:87–95.

Käufl, J. S., Grayver, A. V., Comeau, M. J., Kuvshinov, A. V., Becken, M., Kamm, J., Batmagnai, E., and Demberel, S. (2020). Magnetotelluric multiscale 3-d inversion reveals crustal and upper mantle structure beneath the hangai and gobi-altai region in mongolia. *Geophysical Journal International*, 221(2):1002–1028.

Kharazmi, E., Zhang, Z., and Karniadakis, G. E. (2019). Variational physics-informed neural networks for solving partial differential equations. *arXiv preprint arXiv:1912.00873.*

Kilic, F. C. (2016). Geothermal energy in turkey. *Energy & Environment*, 27(3-4):360–376.

Krishnamoorthy, C., Chen, W., and Rajeev, S. (2018). *Artificial intelligence and expert systems for engineers.* CRC press.

Kulasekara, H. and Seynulabdeen, V. (2019). A review of geothermal energy for future power generation. In *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, pages 223–228. IEEE.

Kumar, L., Hossain, M. S., Assad, M. E. H., and Manoo, M. U. (2022). Technological advancements and challenges of geothermal energy systems: A comprehensive review. *Energies*, 15(23):9058.

Kuvshinov, A., Grayver, A., and Samrock, F. (2019). Deep Electromagnetic Studies of the Earth. *ETH Zürich*, Institure.

Lang, I. (2022). Coscine – fair im datenmeer. *Zenodo.*

Larson, M. G. and Bengzon, F. (2013). *The finite element method: theory, implementation, and applications*, volume 10. Springer Science & Business Media.

Lary, D. J. (2010). Artificial intelligence in geoscience and remote sensing. In *Geoscience and Remote Sensing New Achievements.* IntechOpen.

Lattimer, B., Hodges, J., and Lattimer, A. (2020). Using machine learning in physics-based simulation of fire. *Fire Safety Journal*, 114:102991.

Lindsay, A. D., Gaston, D. R., Permann, C. J., Miller, J. M., Andrš, D., Slaughter, A. E., Kong, F., Hansel, J., Carlsen, R. W., Icenhour, C., Harbour, L., Giudicelli, G. L., Stogner, R. H., German, P., Badger, J., Biswas, S., Chapuis, L., Green, C., Hales, J., Hu, T., Jiang, W., Jung, Y. S., Matthews, C., Miao, Y., Novak, A., Peterson, J. W., Prince, Z. M., Rovinelli, A., Schunert, S., Schwen, D., Spencer, B. W., Veeraraghavan, S., Recuero, A., Yushu, D., Wang, Y., Wilkins, A., and Wong, C. (2022). 2.0 - MOOSE: Enabling massively parallel multiphysics simulation. *SoftwareX*, 20:101202.

Liu, Y., Starzyk, J. A., and Zhu, Z. (2008). Optimized approximation algorithm in neural networks without overfitting. *IEEE transactions on neural networks*, 19(6):983–995.

Lopez-Pacheco, M. and Yu, W. (2022). Complex valued deep neural networks for nonlinear system modeling. *Neural Processing Letters*, pages 1–22.

Lu, S.-M. (2018). A global review of enhanced geothermal system (egs). *Renewable and Sustainable Energy Reviews*, 81:2902–2921.

Luhmann, A. J., Tutolo, B. M., Bagley, B. C., Mildner, D. F., Seyfried Jr, W. E., and Saar, M. O. (2017). Permeability, porosity, and mineral surface area changes in basalt cores induced by reactive transport of co 2-rich brine. *Water Resources Research*, 53(3):1908–1927.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9:381–386.

Manassero, M. C., Afonso, J. C., Zyserman, F., Zlotnik, S., and Fomin, I. (2020). A reduced order approach for probabilistic inversions of 3-d magnetotelluric data i: general formulation. *Geophysical Journal International*, 223(3):1837–1863.

Manson, M. (2016). *The Subtle Art of Not Giving a F* ck: A Counterintuitive Approach to Living a Good Life*. Macmillan Publishers Aus.

Mansoori, I., Oskooi, B., Pedersen, L. B., and Javaheri, R. (2016). Three-dimensional modelling of magnetotelluric data to image sehqanat hydrocarbon reservoir in southwestern iran. *Geophysical Prospecting*, 64(3):753–766.

Manteufel, R. (2000). Evaluating the convergence of latin hypercube sampling. In *41st Structures, Structural Dynamics, and Materials Conference and Exhibit*, page 1636.

Martí, A. (2014). The role of electrical anisotropy in magnetotelluric responses: From modelling and dimensionality analysis to inversion and interpretation. *Surveys in Geophysics*, 35:179–218.

Matthès, M. W., Bromberg, Y., de Rosny, J., and Popoff, S. M. (2021). Learning and avoiding disorder in multimode fibers. *Physical Review X*, 11(2):021060.

Mazzarini, F. and Isola, I. (2010). Monogenetic vent self-similar clustering in extending continental crust: Examples from the east african rift system. *Geosphere*, 6(5):567–582.

McCarthy, J. et al. (2007). What is artificial intelligence. *Stanford University*.

McKay, M. D., Beckman, R. J., and Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61.

Miller, C. A., Barretto, J., Stagpoole, V., Caratori-Tontini, F., Brakenrig, T., and Bertrand, E. (2022). The integrated history of repeated caldera formation and infill at the okataina volcanic centre: Insights from 3d gravity and magnetic models. *Journal of Volcanology and Geothermal Research*, 427:107555.

Mohd Adnan, M., Sarkheyli, A., Mohd Zain, A., and Haron, H. (2015). Fuzzy logic for modeling machining process: a review. *Artificial Intelligence Review*, 43:345–379.

Mohtasham, J. (2015). Renewable energies. *Energy Procedia*, 74:1289–1297.

Munoz, G. (2014). Exploring for geothermal resources with electromagnetic methods. *Surveys in geophysics*, 35(1):101–122.

Munoz, G., Bauer, K., Moeck, I., Schulze, A., and Ritter, O. (2010). Exploring the groß schönebeck (germany) geothermal site using a statistical joint interpretation of magnetotelluric and seismic tomography models. *Geothermics*, 39(1):35–45.

Murphy, R. R. (2019). *Introduction to AI robotics*. MIT press.

Newman, G. A., Gasperikova, E., Hoversten, G. M., and Wannamaker, P. E. (2008). Three-dimensional magnetotelluric characterization of the coso geothermal field. *Geothermics*, 37(4):369–399.

Nti, I. K., Adekoya, A. F., Weyori, B. A., and Nyarko-Boateng, O. (2022). Applications of artificial intelligence in engineering and manufacturing: a systematic review. *Journal of Intelligent Manufacturing*, 33(6):1581–1601.

Nunes Rodrigues, A. C., Santos Pereira, A., Sousa Mendes, R. M., Araújo, A. G., Santos Couceiro, M., and Figueiredo, A. J. (2020). Using artificial intelligence for pattern recognition in a sports context. *Sensors*, 20(11):3040.

Olasolo, P., Juárez, M., Morales, M., Liarte, I., et al. (2016). Enhanced geothermal systems (egs): A review. *Renewable and Sustainable Energy Reviews*, 56:133–144.

Olsson, A., Sandberg, G., and Dahlblom, O. (2003). On latin hypercube sampling for structural reliability analysis. *Structural safety*, 25(1):47–68.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

Peacock, J. R., Mangan, M. T., McPhee, D., and Wannamaker, P. E. (2016). Three-dimensional electrical resistivity model of the hydrothermal system in long valley caldera, california, from magnetotellurics. *Geophysical Research Letters*, 43(15):7953–7962.

Peacock, J. R., Thiel, S., Heinson, G. S., and Reid, P. (2013). Time-lapse magnetotelluric monitoring of an enhanced geothermal system. *Geophysics*, 78(3):B121–B130.

Peacock, J. R., Thiel, S., Reid, P., and Heinson, G. (2012). Magnetotelluric monitoring of a fluid injection: Example from an enhanced geothermal system. *Geophysical Research Letters*, 39(18).

Peng, X. B., Abbeel, P., Levine, S., and Van de Panne, M. (2018). Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14.

Peterson, J. W., Lindsay, A. D., and Kong, F. (2018). Overview of the incompressible navier–stokes simulation capabilities in the moose framework. *Advances in Engineering Software*, 119:68–92.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2017). Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*.

Reinecker, J., Hochschild, T., Kraml, M., Löschan, G., and Kreuter, H. (2019). Experiences and challenges in geothermal exploration in the upper rhine graben. In *Proceedings of European Geothermal Congress*, pages 11–12.

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., and Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling and Software*, 114(January):29–39.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270.

Samrock, F., Grayver, A., Dambly, M. L. T., Müller, M. R., and Saar, M. (2023). Geophysically guided well siting at the aluto-langano geothermal reservoir. *Geophysics*, 88(5):1–43.

Samrock, F., Grayver, A. V., Bachmann, O., Karakas, Ö., and Saar, M. O. (2021). Integrated magnetotelluric and petrological analysis of felsic magma reservoirs: Insights from ethiopian rift volcanoes. *Earth and Planetary Science Letters*, 559:116765.

Samrock, F., Grayver, A. V., Eysteinsson, H., and Saar, M. O. (2018). Magnetotelluric image of transcrustal magmatic system beneath the tulu moye geothermal prospect in the ethiopian rift. *Geophysical Research Letters*, 45(23):12–847.

Sazli, M. H. (2006). A brief review of feed-forward neural networks. *Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering*, 50(01).

Schulz, E., Speekenbrink, M., and Krause, A. (2018). A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, 85:1–16.

Selway, K., O'Donnell, J. P., and Özaydin, S. (2019). Upper mantle melt distribution from petrologically constrained magnetotellurics. *Geochemistry, Geophysics, Geosystems*, 20(7):3328–3346.

Shinde, P. P. and Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*, pages 1–6. IEEE.

Siniscalchi, A., Romano, G., Tripaldi, S., et al. (2021). Magnetotellurics: An overview. *Encyclopedia of Geology, 2nd edition*, 1:698–705.

Slob, E., Draganov, D., and Wapenaar, K. (2007). Interferometric electromagnetic green's functions representations using propagation invariants. *Geophysical Journal International*, 169(1):60–80.

Song, X., Li, G., Huang, Z., Shi, Y., Wang, G., Song, G., and Xu, F. (2022). Review of high-temperature geothermal drilling and exploitation technologies. *Gondwana Research*.

Spichak, V., Geiermann, J., Zakharova, O., Calcagno, P., Genter, A., and Schill, E. (2015). Estimating deep temperatures in the soultz-sous-forêts geothermal area (france) from magnetotelluric data. *Near Surface Geophysics*, 13(4):397–408.

Stern, M., Hexner, D., Rocks, J. W., and Liu, A. J. (2021). Supervised learning in physical networks: From machine learning to learning machines. *Physical Review X*, 11(2):021045.

Swischuk, R., Mainini, L., Peherstorfer, B., and Willcox, K. (2019). Projection-based model reduction: Formulations for physics-based machine learning. *Computers & Fluids*, 179:704–717.

Taillefer, A., Soliva, R., Guillou-Frottier, L., Le Goff, E., Martin, G., and Seranne, M. (2017). Fault-related controls on upward hydrothermal flow: an integrated geological study of the têt fault system, eastern pyrénées (france). *Geofluids*, 2017.

Tseng, K. H., Ogawa, Y., Nurhasan, Tank, S. B., Ujihara, N., Honkura, Y., Terada, A., Usui, Y., and Kanda, W. (2020). Anatomy of active volcanic edifice at the Kusatsu–Shirane volcano, Japan, by magnetotellurics: hydrothermal implications for volcanic unrests. *Earth, Planets and Space*, 72(1):1–11.

Ussher, G., Harvey, C., Johnstone, R., and Anderson, E. (2000). Understanding the resistivities observed in geothermal systems. In *proceedings world geothermal congress*, pages 1915–1920. Kyushu Japan.

Vadyala, S. R., Betgeri, S. N., Matthews, J. C., and Matthews, E. (2022). A review of physics-based machine learning in civil engineering. *Results in Engineering*, 13:100316.

Vallianatos, F. et al. (2002). A note on the topographic distortion of magnetotelluric impedance. *Annals of Geophysics*.

Vijayakumar, S. and Sheshadri, K. (2019). Applications of artificial intelligence in academic libraries. *International Journal of Computer Sciences and Engineering*, 7(1):136–140.

Vozoff, K. (1991). The magnetotelluric method. In *Electromagnetic Methods in Applied Geophysics: Volume 2, Application, Parts A and B*. Society of Exploration Geophysicists.

Wainwright, H. M., Finsterle, S., Jung, Y., Zhou, Q., and Birkholzer, J. T. (2014). Making sense of global sensitivity analyses. *Computers and Geosciences*, 65:84–94.

Wang, J., Li, Y., Gao, R. X., and Zhang, F. (2022a). Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems*, 63:381–391.

Wang, Q., Hesthaven, J. S., and Ray, D. (2019). Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem. *Journal of computational physics*, 384:289–307.

Wang, S., Yu, X., and Perdikaris, P. (2022b). When and why pinns fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768.

Wellmann, F. and Caumon, G. (2018). 3-d structural geological models: Concepts, methods, and uncertainties. In *Advances in geophysics*, volume 59, pages 1–121. Elsevier.

Willcox, K. E., Ghattas, O., and Heimbach, P. (2021). The imperative of physics-based modeling and inverse theory in computational science. *Nature Computational Science*, 1(3):166–168.

Wu, Y.-c. and Feng, J.-w. (2018). Development and application of artificial neural network. *Wireless Personal Communications*, 102:1645–1656.

Zarrouk, S. J. and McLean, K. (2019). Geothermal well test analysis: fundamentals, applications and advanced techniques. *Academic Press*.

Zhang, L. X., Pang, M. Y., Han, J., Li, Y., and Wang, C. B. (2019). Geothermal power in china: Development and performance evaluation. *Renewable and Sustainable Energy Reviews*, 116:109431.

Zhang, Z. (2018). Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pages 1–2. Ieee.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560.

Zorin, N., Aleksanova, E., Shimizu, H., and Yakovlev, D. (2020). Validity of the dispersion relations in magnetotellurics: Part i—theory. *Earth, Planets and Space*, 72:1–18.

List of Figures

List of Tables

Appendix A

Data Management

The python scripts, data and surrogate models are currently being stored for the long term in Coscine (Lang, 2022) for the research group, committee members, and for the public on demand. This cloud allows files storage and data sharing. GoFEM is publicly accessible via GitHub (Grayver and Kolev, 2015). The shared Python scripts are sorted depending on the specific topic, following the structure of the thesis. Supplements to the the phase, resistivity and impedance for each approach are being stored.

The link to the Coscine library is:

https://coscine.rwth-aachen.de/p/magnetotellurics/

Appendix B

Theoretical Background

This appendix gives a more extensive theoretical background on subjects related to this thesis. It covers information about AI and ML, as well as about the RB method and physics-based models. Furthermore, it introduces the theory of SA as a potential future direction arising from this project.

B-1 Introduction to Artificial Intelligence and Machine Learning

This section aims to give a brief introduction to AI and its various branches. Since in this project the branch of ML is applied, this is the main focus.

B-1-1 Background of Artificial Intelligence

AI is a field of science that aims to copy human intelligence into computer systems (Fetzer and Fetzer, 1990). Its development began in the time after World War II, when researchers such as the English mathematician Alan Turing started exploring the concept of intelligent machines (McCarthy et al., 2007). Over time, the number of researchers involved in AI increased, leading to an accelerated development of AI. This is caused by its wide field of study with many varying applications, including medicine (Hamet and Tremblay, 2017), finances (Cao, 2022), engineering (Nti et al., 2022), criminology (Hayward and Maas, 2021) and geosciences (Lary, 2010).

There are various branches of AI (Vijayakumar and Sheshadri, 2019; Mohd Adnan et al., 2015; Wu and Feng, 2018) and here we introduce six of them:

- **Pattern Recognition** includes the use of AI to identify and classify patterns in diverse environments (Nunes Rodrigues et al., 2020). This process takes information from sensor signals and compares it to patterns known from prior knowledge and previous teaching processes. (Vijayakumar and Sheshadri, 2019). Additionally, mathematical and statistical algorithms are applied to make optimal predictions using uncertainty and probability (Bishop and Nasrabadi, 2006).

- **Expert Systems** plays a role when specific expertise in a particular domain is asked (Krishnamoorthy et al., 2018). In order to find solutions of a problem, a system combines several elements, like prior knowledge, interference mechanisms between parameters and user interface (Krishnamoorthy et al., 2018). These gathered information are then used to offer advice, solutions and decision making to the particular field it is applied to (Vijayakumar and Sheshadri, 2019).
- **Natural Language Processing** aims to build a connection between the human language and computational machines (Chowdhary, 2020b). Its goal is to analyse and understand the natural language, and to generate texts and speeches, motivated by theory. A famous example nowadays is ChatGPT.
- **Robotics** represents the branch of AI which includes the control over robots about how to react to an environment, how to solve problems and how to learn from past experiences (Murphy, 2019; Krishnamoorthy et al., 2018; Vijayakumar and Sheshadri, 2019). These robots use information from sensors to react in specific ways, based on their prior knowledge (Krishnamoorthy et al., 2018).
- **Fuzzy Logic** includes *fuzzy sets* that generalize a mathematically classical set by defining a degree with a real number between 0 and 1 (Chowdhary, 2020a). This approach allows to work with uncertainties, which can be beneficial for its accuracy and efficiency in managing large and complex data sets (Ahn et al., 2019).
- **ML** can be defined by the ability of computers to learn without explicit programming, as stated by the American pioneer Arthur Samuel in 1959 (Vijayakumar and Sheshadri, 2019). Its goal is to autonomously learn from data and to improve its performance over time (Chowdhary, 2020a). ML is broadly divided into three primary sub-branches, depending on the type of signal or response to the learning system: (1) supervised learning, (2) unsupervised learning, and (3) reinforcement learning.

The NI-RB method can be assigned to the category of ML, and therefore, this branch is further explored in the following section.

B-1-2 Overview about Machine Learning

ML can be defined as a computational process using input data to produce a specific output without the need for explicit programming at each step (El Naqa and Murphy, 2015). Its algorithm works based on iterations, where the output is continuously adjusted and improved, so called *trained*, through code self-configuration, making the algorithm somewhat *soft coded* (El Naqa and Murphy, 2015). The training processes is the main advantage of ML, as it allows the machine to work autonomously as soon as the algorithm understood how to produce a desired outcome (Mahesh, 2020). The algorithm can vary depending on the individual problem and the desired result. Due to this wide range of algorithms available, ML has numerous applications across various domains, including computer vision, prediction, semantic analysis, natural language processing and information retrieval (Shinde and Shah, 2018). Figure B-1 shows possible algorithms used in ML.

Chapter B-1-1 highlighted three primary ML algorithms: supervised, unsupervised, and reinforcement learning. One of these specific algorithms is chosen depending on the individual problem and the data set. The following paragraphs provide an overview of these three main learning mechanisms.

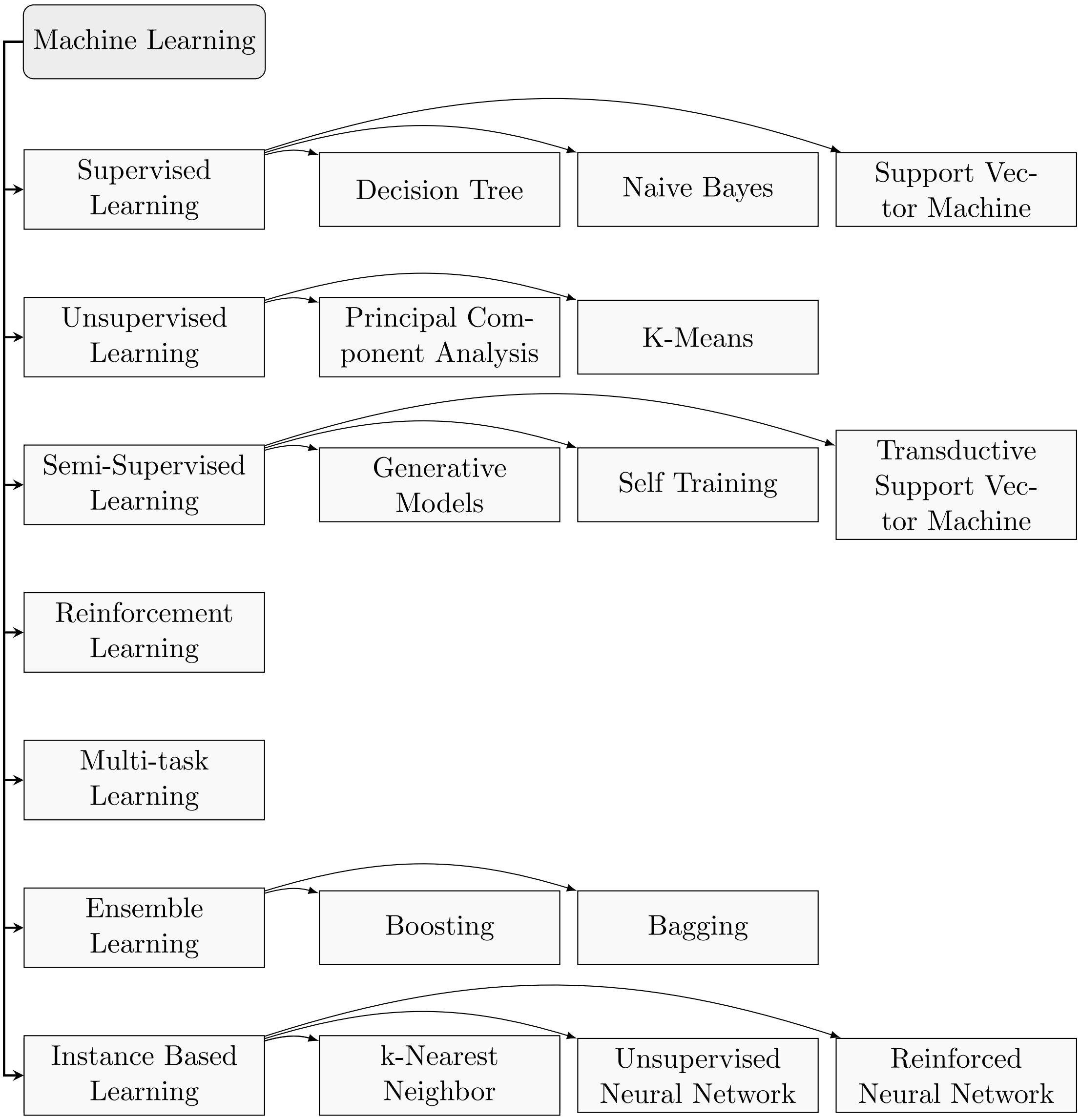

Figure B-1: Possible algorithms in ML (modified after Mahesh (2020)).

Supervised Learning describes the processing of labelled training data (Bell, 2022). This technique needs a large set of training data, which is manually categorized in a way the result is desired. Therefore, the computer learns based on predefined input-output pairs to find a function to correlate these pairs (Mahesh, 2020). After the learning process of encountering multiple examples part of the training data set, the system should be able to categorize new data itself, which can be tested using a validation data set. Figure B-2 shows a typical workflow for supervised learning.

Unsupervised Learning works with unlabeled data sets, in contrast to supervised learning. Instead of inserting pre-labeled input data, the system tries to categorize itself (Mahesh, 2020). Therefore, the output is initially undefined (Bell, 2022). The system can be guided in desired directions by defining DoF and initial conditions (El Naqa and Murphy, 2015). Based on previous experiences, it defines characteristics of new data sets (Mahesh, 2020). Figure B-3 shows a typical workflow for unsupervised learning.

Reinforcement Learning aims to maximize a cumulative reward by a sequence of actions (Mahesh, 2020). In its behavior it combines the fields of Markov decision processes with supervised ML (El Naqa and Murphy, 2015). Feedback is included in the learning process in form of rewards or penalties depending of the current environment state (El Naqa and Murphy, 2015). This interaction is visualized in Figure B-4.

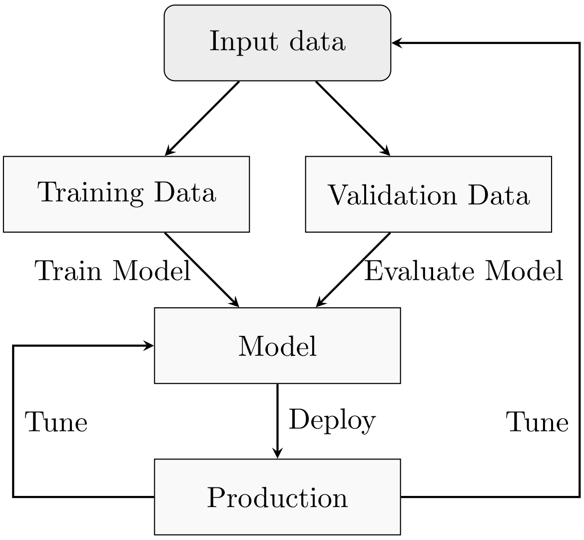

Figure B-2: Workflow for supervised learning (modified after Mahesh (2020)).

B-1-3 Data-Driven Machine Learning

Compared to the later described physics-based modelling, data-driven ML is a method not necessarily relying on physical laws. This ML technique requires large data sets in order to predict outcomes from given inputs (Lattimer et al., 2020). It is able to extract important parts of a signal as critical information and can predict a systems behavior by correlating parameters with each other (Wang et al., 2022a). It uses samples of known variables to give insights of their relationship. The focus of data-driven ML lies on obtaining a high accuracy and to minimize the error between the prediction of the ML and the actual observation included in a test data set (Wang et al., 2022a). Figure B-5 shows a possible workflow for the data-driven ML method.

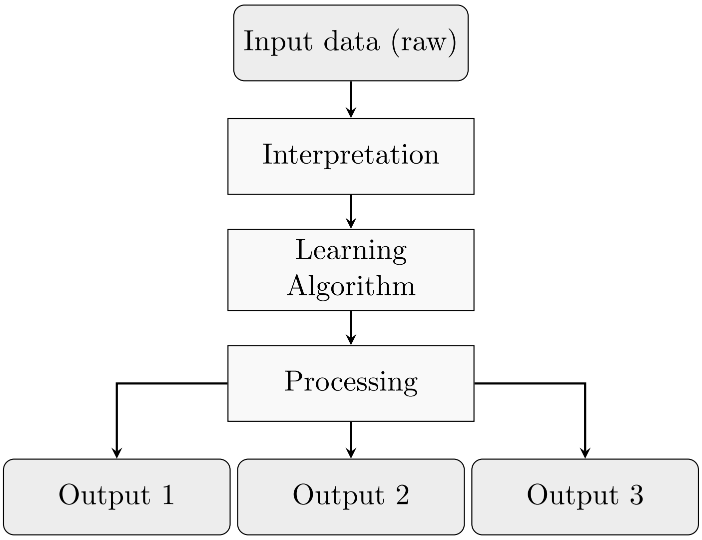

Figure B-3: Workflow for unsupervised learning (modified after Mahesh (2020)).

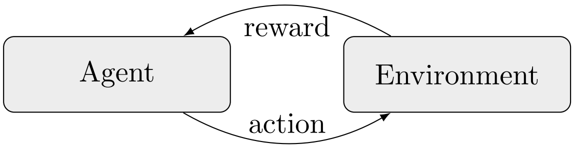

Figure B-4: Workflow for reinforcement learning (modified after Mahesh (2020)).

This method enables the handling of data sets with high uncertainties and low complexity, and the analysis of the complex relationship between input and output (Wang et al., 2022a). The presence of the black box allows a non-intrusive and highly flexible approach of potentially large accuracy, if a large data set is given (Willcox et al., 2021). It often highly correlates with samples of known variables (Wang et al., 2022a). The black box furthermore increases the modelling speed while decreasing the computational costs, and enables the construction of a model with low memory (Wang et al., 2022a). On the other side it requires large data sets in order to allow a high accuracy, which is often difficult in geophysical applications (Degen et al., 2022). Its functioning is also not rigorously proven, as the existence of the black box makes physical interpretation and the understanding of the inner logic of the model construction difficult (Degen et al., 2022; Wang et al., 2022a). This is also caused by the fact that ML treats physics simply as data, which makes it difficult to use in UQ (Degen et al., 2022).

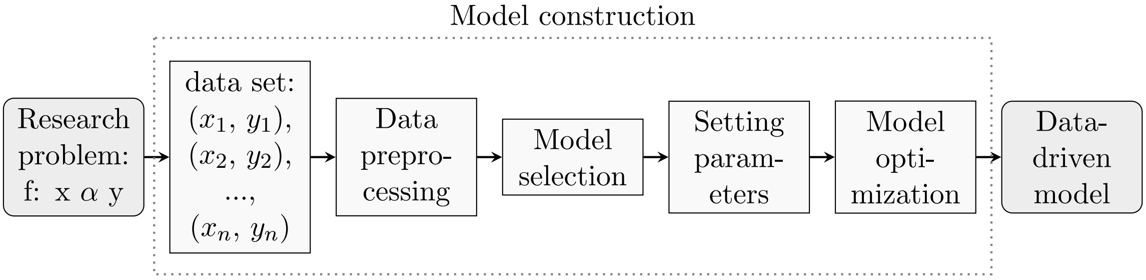

Figure B-5: Workflow for the data-driven ML method (modified after Wang et al. (2022a)).

B-2 Reduced Basis Method

The intrusive RB method is a type of physics-based reduced order modelling. It is a projection-based model order technique with the aim to reduce the DoF of a full-order model while maintaining the input-output relationship (Degen et al., 2022). It consist of an offline and online stage like the NI-RB method, where the focus of the offline stage is to construct a reduced order model which is then solved in the online stage (Hesthaven and Ubbiali, 2018).

The approach of modelling by approximations and therefore simplifications of original models, while maintaining crucial processes and other important aspects of the model is called *reduced order modeling (ROM)* (Bai et al., 2005). It is relevant to enable an efficient and numerically stable computation, by reducing the state-space dimension. This leads to a decrease of computational costs while loosing accuracy in a controlled way (Hesthaven and Ubbiali, 2018). One class of ROMs are projection based surrogate models. In this specific class are low-dimensional approximations derived by projecting the original model onto a low-dimensional subspace (Swischuk et al., 2019) spanned by a set of basis functions. Physical parameters like material properties, initial conditions and boundary condition (BC) can be defined in the projection framework, aiming to preserve the underlying physics (Swischuk et al., 2019; Degen et al., 2022). Possible methods to obtain a projection-based surrogate model are, besides others, PODs, Karhunen-Loève expansion or Krylov subset methods (Asher et al., 2015).

The model of the RB method is constructed by basis functions and weights during the offline stage. This basis functions can be calculated by various algorithms, such as PODs or the greedy algorithm (Hesthaven and Ubbiali, 2018; Chen et al., 2021). The basis functions are calculated from snapshots, which are the solutions given by high-fidelity models of different inputs (Swischuk et al., 2019). During the online stage, the solution is searched as a linear combination of the previously defined basis functions via projection (Hesthaven and Ubbiali, 2018). In the RB method, this projection is done via the Galerkin-projection (Hesthaven and Ubbiali, 2018). The full-order model is therefore projected onto a reduced space, spanned by the defined basis functions (Chen et al., 2021). The final solution is defined for each new parameter location with reduced coefficients (Swischuk et al., 2019; Wang et al., 2019).

When having non-linear or hyperbolic problems, the RB method causes some problems, as it tends to have problems with convergence and accuracy (Chen et al., 2021; Wang et al., 2019). Therefore, the intrusive approach is generally only applicable to a narrow range of geoscientific problems, as they are often of hyperbolic or nonlinear nature (Degen et al., 2022). For such applications, the computational gain obtained by applying this method is limited with respect to the full-order model (Wang et al., 2019).

B-3 Physics-Based Models

Models that are based on PDEs and therefore represent laws of physics and nature are called *physics-based models* (Degen et al., 2022; Willcox et al., 2021). We can incorporate these laws into models by specifying physical processes, parameters, the correlation between physical variables or BCs through PDEs, which constrain the simulations to a specific physical system (Wang et al., 2022a; Willcox et al., 2021). Typical steps in construction of physics-based models are shown in Figure B-6. The obtained model can be used to estimate uncertainties while having a high approximation of the Earth's system dynamics (Degen et al., 2022). Since many properties of the Earth are better represented by a range of values rather than a single value, parameters are frequently defined over such ranges. Therefore, parameters like initial conditions, BCs, sources, material properties and the model geometry often vary (Willcox et al., 2021). The model uses these physical parameters and BCs under the implementation of the hypothesis set to solve the target variable (Figure B-6) (Wang et al., 2022a). This method allows the analysis of complex relationships between input and output, while ensuring physical laws. This increases the interpretability and the understanding of the model, and also allows the user to control the model construction more than when a black box is present, which is typical for ML. On the other side, high-fidelity simulations are needed, which increase the computational costs drastically and therefore limit also the response rate in data analysis (Wang et al., 2022a).

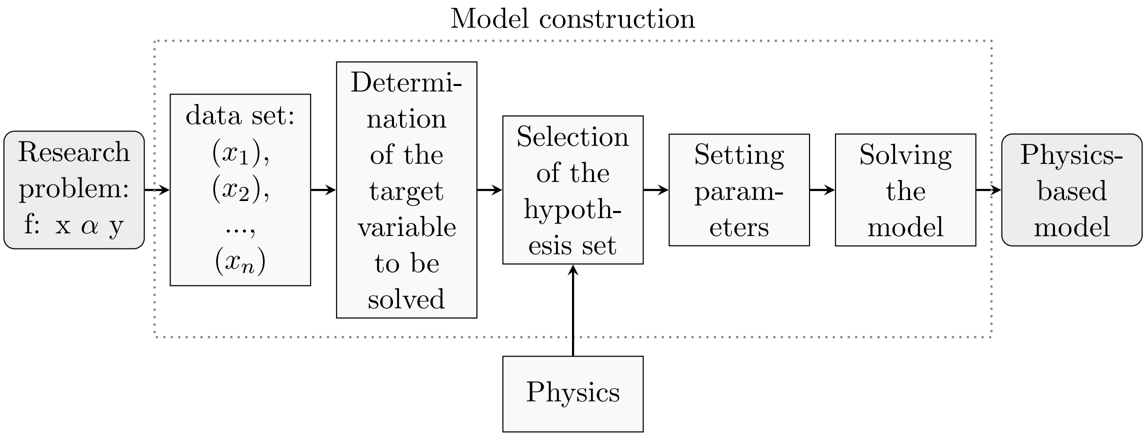

Figure B-6: Workflow for physics-based models (modified after Wang et al. (2022a)).

B-4 Sensitivity Analysis

SA is used in modelling and predictions, in order to find the model parameters with the most significant effect on the model output, and the ones which can be neglected compared to the others (Saltelli et al., 2019). Its goal is to reduce complexity when it is not of advantage and to prioritize relevant parameters, which information can later be applied in UQs. It overall allows model optimization and calibration, enabled by parameter analysis, to determine if a model parameter has a sufficiently high impact on the modelling outcome to be estimated from it (Degen et al., 2021b). A typical sub-categorization of SA is the division into local and global SA (Degen et al., 2021b; Wainwright et al., 2014), which are explained in more detail in the following paragraphs.

The **local SA** is based on the assumption that the parameters used in the process of modelling are independent of each other (Degen et al., 2021b). Its goal is to estimate the influence of the model parameters on the model output in the vicinity of the initial parameters, which is also referred to as the *first-order index.* In the example of modelling the temperature distribution in the subsurface as described in Degen et al. (2021b), we define a function u = f(x), with x = ($\mathrm{x_1}$, ..., $\mathrm{x_n}$) representing thermal conductivities present in the subsurface. When performing a local SA at x^0, we evaluate for $\left(\frac{\partial u}{\partial x_k}\right)_{x=x^0}$. Thus, we do not include parameter correlations, assuming that a change in the model response is caused from a single changed parameter with a pre-defined variance of, e.g., 1 %. The local SA then calculates the parameter influence with respect to an user-defined reference parameter set (Degen et al., 2021a). Mathematically, the local SA index is calculated for each model parameter x = ($\mathrm{x_1}$, ..., $\mathrm{x_n}$) as the scaled partial derivative of the function u = f(x) with respect to $\mathrm{x_n}$ (Wainwright et al., 2014). Assuming each parameter $\mathrm{x_n}$ is described by a set of $\{x_n \mid n = 1, \ldots, k\}$, we can state that the local SA index for $\mathrm{x_n}$ is

$$S_n^{\text{local}} = \frac{\tau_{x,n}}{\tau_y} \frac{\partial y}{\partial x_n}\bigg|_{x_n;} = \frac{\tau_{x,n} f\left(x_1*, \ldots, x_n * + \Delta x_n, \ldots, x_k*\right) - f\left(x_1*, \ldots, x_k*\right)}{\Delta x_y}, \tag{B-1}$$

with the parameter-scaling factor $\tau_{x,n}$ representing the standard deviation or parameter uncertainty, the output-scaling factor τ_y describing an significant change in the model output, and the increment Δx_n by which each parameter n is increased from the reference parameter values x_n* (Wainwright et al., 2014). Overall, i = (k+1) simulations are required in the local SA, which sums up k simulations plus the reference parameter set for measuring changes of an increment of Δx_n of k parameters.

The general aim is to fit the mentioned reference values x_n* obtained from the model as closely as possible to the observed values x_n (Degen et al., 2021b). Due to measurement and model errors this is difficult to achieve, resulting in uncertainties of the local SA. Additionally, a local SA can lead to am overestimation of the influence of model parameters (Degen and Cacace, 2021). Therefore, no local but a global SA should applied as a future step of this project, under the consideration of a method reducing computational costs.

The **global SA** aims to estimate the entire parameter domain and also includes parameter correlations (Degen et al., 2021b). It therefore allows more detailed results than the local SA, but with much larger computational costs as several thousands of forward simulations are usually required. This fact makes the application of global SAs infeasible, especially for complex, time-consuming simulations as geothermal systems might demand (Degen et al., 2021b). Several approaches have been developed to counteract these computational costs, such as increasing the model assumptions or introducing response surface fittings (Wainwright et al., 2014). In this project, a global SA is feasible despite the complexity of the high-temperature geothermal system, due to the application of the NI-RB method to reduce computational costs of forward simulations by the use of surrogate models. In contrast to the local SA, the global one does not only reveal the first-order index to describe the influence of single model parameters, but also the *total-order* index, defining the correlation in between the parameters (Degen et al., 2021a).

In a general approach, the global SA tries to minimize the constraint of the reference x^0, which can be done with, e.g., the Sobol SA, a variance-based method (Degen et al., 2021b; Wainwright et al., 2014). Instead of setting $x = x^0$ as done in the local SA, it evaluates over the entire model space f(x), and calculates the ratio between the partial and the total variance as the *Sobol Sensitivity Index* (Degen et al., 2021b). It therefore provides a quantification of the relative relevance of each parameter to the uncertainty in the outputs (Wainwright et al., 2014). A sampling strategy applicable for this project can the Saltelli sampler, in order to reduce the amount of forward evaluations required (Degen and Cacace, 2021; Saltelli, 2002; Saltelli et al., 2010).

Mathematically speaking, the first-order sensitivity index S_i is defined as

$$S_i = \vec{V}\left[\vec{E}\left[U|X_i\right]\right]/\vec{V}[U], \tag{B-2}$$

with the mean $\vec{E}[\bullet]$, the variance $\vec{V}[\bullet]$, the parameters X_i and the model response U (Wainwright et al., 2014). $\vec{E}\left[U|X_i\right]$ describes the mean of U conditioned on purely X_i. The total-order sensitivity index S_{ti} is described as

$$S_{ti} = 1 - \vec{V}\left[\vec{E}\left[U|X_{-i}\right]\right]/\vec{V}[U], \tag{B-3}$$

with $\vec{E}\left[U|X_{-i}\right]$ including the influence of all parameters considered in the model, also taking into account interactions between the parameters themselves (Wainwright et al., 2014).

In the case of a square-integrable function with orthogonal members, it can be decomposed to

$$\int u^2(x)\mathrm{d}x - u_0^2 = \sum_{s=1}^{n} \sum_{i_1<\cdots<i_s}^{n} \int u_{i_1\cdots i_s}^2 \ \mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_s}, \tag{B-4}$$

with $1 \leq i_1 < \cdots < i_s \leq n, u_0$ is a constant, and u_i a function (Degen et al., 2021b). Rewriting Eq. (B-4) to

$$\int f^2(x)\mathrm{d}x - f_0^2 = \sum_i \int f_i^2\left(x_i\right) + \sum_{i<j} \int f_{ij}^2\left(x_i, x_j\right) + \ldots + \int f_{12\cdots n}^2\left(x_1, x_2, \cdots, x_n\right), \tag{B-5}$$

allows the determination of the first-order sensitivity index (or partial variance) S_i and the total-order sensitivity index (or total variance) S_{ti} as (Degen et al., 2021b):

$$S_{ti} = \int f^2 \ \mathrm{d}x - f_0^2, \quad S_{i_1\cdots i_s} = \int f_{i_1\cdots i_s}^2 \ \mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_s}. \tag{B-6}$$

As mentioned before, is the final Sobol Sensitivity Index the ratio between S_i and S_{ti}, including also the parameter interaction. In contrast to the local SA, the global SA does not require a reference parameter set. The determination of a reference set can include errors due to uncertainties, and is dependent on the user (Degen et al., 2021b). Thus, it includes the whole predefined parameter range and therefore allows a larger variations of these parameters (Degen et al., 2021b).

Appendix C

Values of the Electrical Conductivity used in the Simulations

Table C-1: Electrical Conductivity Parameters for the training data set of the high-temperature geothermal model obtained from GoFEM [S m^{-1}]. The values are generated with LHS, and vary as the following ranges [S m^{-1}]: surrounding rock: 0.001-0.02, clay-cap: 0.5-2, upflow zone: 0.001-0.5, source: 0.2-0.5.

Surrounding rock	Clay cap	Upflow zone	Heat source
0.00179	0.873558	0.071222	0.256628
0.005235	0.560586	0.001433	0.391123
0.001078	1.264702	0.006403	0.322269
0.013436	1.753511	0.001323	0.289193
0.005151	1.564487	0.00679	0.246411
0.00313	1.309952	0.040495	0.359855
0.008206	0.763456	0.028488	0.390951
0.002566	0.542313	0.060559	0.2499
0.003655	1.216053	0.012348	0.231985
0.003128	1.066297	0.008	0.204038
0.012851	0.750166	0.005552	0.265686
0.01253	0.534214	0.005745	0.351058
0.015502	1.409274	0.02085	0.468874
0.00121	1.914293	0.222892	0.276564
0.002349	0.946691	0.117568	0.447279
0.00333	0.673321	0.332476	0.431199
0.002117	0.885385	0.278004	0.409824
0.003224	0.855413	0.007897	0.253117
0.003753	0.527395	0.140436	0.246007
0.003954	0.900106	0.003097	0.407118

Continued on next page

Table C-1 – Continued from previous page

Surrounding rock	Clay cap	Upflow zone	Heat source
0.001731	0.722139	0.05082	0.278837
0.017673	1.288451	0.016719	0.484789
0.004999	1.798917	0.101476	0.487309
0.006636	1.829513	0.160087	0.45652
0.00106	1.118019	0.036737	0.388589
0.006724	0.99576	0.468405	0.440956
0.001689	0.904728	0.010948	0.295766
0.00993	0.899959	0.017166	0.292051
0.00235	1.356113	0.001921	0.205656
0.001713	1.059372	0.142431	0.245485
0.013549	0.668964	0.0318	0.210659
0.002634	0.644685	0.001256	0.323062
0.001309	0.790041	0.009564	0.219862
0.001363	1.670176	0.003395	0.433138
0.001233	1.883392	0.021013	0.239492
0.001284	0.627056	0.039119	0.236584
0.005974	0.835088	0.022829	0.374072
0.00927	1.112136	0.004353	0.203078
0.004796	1.467472	0.050159	0.249955
0.001258	0.907519	0.011765	0.356231
0.006471	0.740549	0.007459	0.243922
0.00434	1.990779	0.141312	0.302067
0.002291	1.023374	0.003688	0.254578
0.006095	0.842842	0.00161	0.47761
0.009648	0.707107	0.323625	0.245416
0.005856	1.95432	0.075907	0.303918
0.00139	0.747425	0.001156	0.270614
0.001631	0.687771	0.205172	0.495439
0.002032	1.242575	0.001422	0.256138
0.004701	0.681444	0.023794	0.223928
0.01865	0.593231	0.160014	0.407471
0.001506	1.936341	0.001981	0.218523
0.008389	1.972465	0.001749	0.335122
0.008731	0.80478	0.015085	0.438462
0.006218	1.265757	0.001823	0.34132
0.001839	0.92445	0.002065	0.457617
0.002687	0.571701	0.036009	0.226681
0.001072	0.899171	0.020161	0.309539
0.016217	0.50698	0.124796	0.296582
0.002911	1.337928	0.00254	0.205574
0.017565	0.858565	0.001309	0.29298
0.003282	1.701334	0.007774	0.337175
0.002157	1.208597	0.004929	0.213248

Continued on next page

Table C-1 – Continued from previous page

Surrounding rock	Clay cap	Upflow zone	Heat source
0.002481	1.277509	0.222897	0.214555
0.003091	0.782773	0.067035	0.358414
0.002115	1.815038	0.351583	0.354062
0.002582	1.654811	0.005817	0.371794
0.007592	0.915945	0.037532	0.259286
0.001952	1.685683	0.013322	0.376364
0.005406	0.675175	0.052281	0.443852
0.001139	0.720298	0.003257	0.311435
0.011781	0.866537	0.196845	0.284167
0.002337	1.536875	0.056798	0.280716
0.007745	1.453973	0.015723	0.209376
0.017919	0.761368	0.001678	0.277308
0.002853	0.66281	0.030509	0.251487
0.00297	1.565547	0.097329	0.230873
0.001209	0.933033	0.079118	0.285908
0.003153	0.941696	0.048124	0.201841
0.006343	1.717131	0.001064	0.233711
0.016544	1.081725	0.054493	0.351906
0.001094	1.313425	0.042499	0.313343
0.013822	1.032876	0.028083	0.40499
0.015896	1.164734	0.002338	0.492422
0.003554	0.577009	0.017804	0.235143
0.007902	0.604299	0.008103	0.327033
0.007009	0.540862	0.026944	0.415007
0.0053	1.427344	0.00632	0.36061
0.003928	0.656712	0.009968	0.36728
0.012019	0.511687	0.093379	0.331053
0.015582	0.754364	0.29789	0.397635
0.007442	0.638754	0.166783	0.468937
0.010876	1.781797	0.14729	0.282436
0.018281	0.694157	0.016388	0.24692
0.012761	0.561231	0.005137	0.257707
0.00103	1.197479	0.003538	0.315263
0.014971	1.143402	0.450803	0.206834
0.003626	0.615572	0.02585	0.28766
0.005195	0.502316	0.002243	0.41248
0.005092	0.700602	0.398114	0.221209
0.015273	1.624505	0.002998	0.232288
0.006735	1.175548	0.002876	0.339241
0.00101	0.650671	0.004176	0.321094
0.002742	1.639588	0.0592	0.238034
0.017218	0.535887	0.001901	0.4305
0.001664	1.848899	0.119731	0.24096

Continued on next page

Table C-1 – Continued from previous page

Surrounding rock	Clay cap	Upflow zone	Heat source
0.003484	1.301342	0.00276	0.390415
0.00574	0.874583	0.009176	0.383325
0.010041	0.97716	0.046171	0.222565
0.002384	0.582367	0.274201	0.27394
0.014101	1.122462	0.252396	0.217192
0.001599	1.481098	0.034547	0.380991
0.004608	0.530957	0.004537	0.268966
0.002797	1.733074	0.337315	0.385674
0.006602	1.231144	0.01039	0.365043
0.002432	1.071773	0.2858	0.294775
0.007151	1.325619	0.002152	0.435792
0.001336	1.551145	0.021902	0.319139
0.009457	1.186463	0.381957	0.279007
0.010451	1.494849	0.005355	0.253028
0.011096	0.550953	0.001205	0.449308
0.008559	1.061914	0.008446	0.498475
0.012261	0.587774	0.085953	0.427878
0.005516	0.545884	0.101446	0.291196
0.002073	0.812252	0.012263	0.425272
0.001914	0.526072	0.013886	0.343411
0.008907	1.219819	0.489749	0.486443
0.008061	0.713672	0.014473	0.378671
0.019801	0.950439	0.242153	0.260875
0.00303	1.594753	0.432508	0.289423
0.001732	1.765406	0.069871	0.489423
0.01132	1.154019	0.003125	0.267328
0.011548	1.414214	0.006587	0.41755
0.002201	1.042466	0.018557	0.47181
0.004893	0.566442	0.001109	0.208101
0.013018	0.768438	0.044297	0.446571
0.00417	1.918528	0.310491	0.298399
0.013281	0.556068	0.033145	0.422683
0.019026	0.99539	0.001482	0.400072
0.003774	1.508729	0.414955	0.262474
0.001567	0.882703	0.469873	0.395214
0.001802	1.798341	0.366456	0.388037
0.003415	1.375542	0.130075	0.474701
0.002245	1.101905	0.188856	0.345515
0.001418	1.35035	0.181192	0.347633
0.00385	0.609909	0.213851	0.48348
0.001162	0.890899	0.001364	0.325041
0.008224	1.013959	0.061704	0.272272
0.003348	1.401204	0.232326	0.402523

Continued on next page

Table C-1 – Continued from previous page

Surrounding rock	Clay cap	Upflow zone	Heat source
0.004428	1.388313	0.008803	0.248433

Table C-2: Electrical Conductivity Parameters for the validation data set of the high-temperature geothermal model obtained from GoFEM [S m^{-1}]. The values are generated with random sampling, and vary as the following ranges [S m^{-1}]: surrounding rock: 0.001-0.02, clay-cap: 0.5-2, upflow zone: 0.001-0.5, source: 0.2-0.5.

Surrounding rock	Clay cap	Upflow zone	Heat source
0.00179	0.873558	0.071222	0.256628
0.005235	0.560586	0.001433	0.391123
0.001078	1.264702	0.006403	0.322269
0.013436	1.753511	0.001323	0.289193
0.005151	1.564487	0.00679	0.246411
0.00313	1.309952	0.040495	0.359855
0.008206	0.763456	0.028488	0.390951
0.002566	0.542313	0.060559	0.2499
0.003655	1.216053	0.012348	0.231985
0.003128	1.066297	0.008	0.204038
0.012851	0.750166	0.005552	0.265686
0.01253	0.534214	0.005745	0.351058
0.015502	1.409274	0.02085	0.468874
0.00121	1.914293	0.222892	0.276564
0.002349	0.946691	0.117568	0.447279
0.00333	0.673321	0.332476	0.431199
0.002117	0.885385	0.278004	0.409824
0.003224	0.855413	0.007897	0.253117
0.003753	0.527395	0.140436	0.246007
0.003954	0.900106	0.003097	0.407118
0.001731	0.722139	0.05082	0.278837
0.017673	1.288451	0.016719	0.484789
0.004999	1.798917	0.101476	0.487309
0.006636	1.829513	0.160087	0.45652
0.00106	1.118019	0.036737	0.388589
0.006724	0.99576	0.468405	0.440956
0.001689	0.904728	0.010948	0.295766
0.00993	0.899959	0.017166	0.292051
0.00235	1.356113	0.001921	0.205656
0.001713	1.059372	0.142431	0.245485