Deliverable 17.6: Advancements in Monitoring Data Management, Modelling and Visualisation

Work Package 17 MODATS

This project has received funding from the European Union's Horizon 2020 research and innovation programme 2014-2018 under grant agreement N°847593.

Document information

Project Acronym	**EURAD**
Project Title	**European Joint Programme on Radioactive Waste Management**
Project Type	**European Joint Programme (EJP)**
EC grant agreement No.	**847593**
Project starting / end date	**1 June 2019 – 31 May 2024**
Work Package No.	**17**
Work Package Title	**Monitoring Equipment and Data Treatment for Safe Repository Operation and Staged Closure**
Work Package Acronym	**MODATS**
Deliverable No.	**17.6**
Deliverable Title	**Advancements in Monitoring Data Management, Modelling and Visualisation**
Lead Beneficiary	**Galson Sciences Limited (Linked Third Party to VTT)**
Contractual Delivery Date	**31 May 2024**
Actual Delivery Date	**7 May 2024**
Type	**Report**
Dissemination level	**Public**
Authors	**Matt White (Galson Sciences Limited)**

To be cited as:

M.J. White, T. Aihkisalo, J. Bertrand, J. Buchwald, F. Chinesta, J. Cotton, N. Graebling, T.J. Haines, G. Hu, P. Kuusela, A. Laikari, E. Manukyan, D. Muñoz, W. Pfingsten, A. Purhonen, M. Schoenball, A.E. Thomas, J. Verstricht, C. Wetter (2024): Advancements in Monitoring Data Management, Modelling and Visualisation. Deliverable D17.6 of the HORIZON 2020 project EURAD. EC Grant agreement no: 847593.

Disclaimer

Acknowledgement

This document is a deliverable of the European Joint Programme on Radioactive Waste Management (EURAD). EURAD has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847593.

Status of deliverable		
	By	Date
Delivered (Lead Beneficiary)	Matt White (Galson Sciences Limited)	07.05.2024
Verified (WP Leader)	Johan Bertrand (Andra) Edgar Bohner (VTT) Kateryna Fuzik (SSTC NRS)	07.05.2024
Reviewed (Reviewers)	Francis Claret (BRGM)	07.05.2024
Approved (PMO)	Bharti Reddy (Nuclear Waste Services)	15.05.204
Submitted to EC (Coordinator)	Andra (Coordinator)	16.05.2024

Executive Summary

The Monitoring Equipment and Data Treatment for Safe Repository Operation and Staged Closure (MODATS) work package (WP) of the European Joint Programme on Radioactive Waste Management (EURAD) is conducting research, development and demonstration (RD&D) into: monitoring data acquisition, treatment and management; use of monitoring data to enhance system understanding, including development of digital twins; interactions with civil society and other stakeholders; development of monitoring technologies; and development of knowledge regarding repository monitoring.

The focus of MODATS is monitoring during the operational phase of repository programmes to build further confidence in the long-term safety case. In particular, MODATS is focusing on confidence in monitoring data. There are particular challenges associated with monitoring data acquired during repository operations, including challenges associated with its management (including processing and storage), modelling and visualisation. These are most frequently associated with the long timescales envisaged for monitoring programmes and the need to maintain the passive safety of the disposal system. Therefore, MODATS has undertaken research to develop methods through which monitoring data can be processed, stored, modelled and visualised. This has included preliminary work on the development of digital twins.

The work on data management, modelling and visualisation was based on six test cases:

- ALC1605 Test Case (undertaken by Amvalor, Andra and ENSAM): The ALC1605 test case used thermal monitoring data to develop methods for combined physics-based and data-driven-modelling of repository evolution.
- FE Experiment Test Case 1 (Nagra): Nagra's FE experiment test case used data from the FE experiment to progress approaches to quality control, data processing, and data storage.
- FE Experiment Test Case 2 (PSI): PSI's FE experiment test case used temperature and relative humidity sensor data to develop machine learning-assisted physics-based and data-driven 3D heat transport modelling.
- FE Experiment Test Case 3 (UFZ): UFZ's FE experiment test case focused on the development of interactive visualisation systems, and visualisation tools for the comparison of simulation and observation data.
- POPLU and Prototype Repository Experiments Test Case (VTT): The POPLU and Prototype Repository experiments test case developed tools and approaches for managing monitoring data, and provided illustrations of how these tools and approaches could be applied during repository monitoring.
- PRACLAY Experiment Test Case (EURIDICE): The PRACLAY test case developed a structured and uniform approach to data processing and storage of monitoring data.

The conclusions from the work are summarised below.

Data Processing

- The work in MODATS has allowed the identification and categorisation of monitoring data anomalies, description of the characteristics that define each type of anomaly, and identification of options for their processing.
- MODATS has developed a range of tools for undertaking monitoring data processing and these have been made available online.
- Data processing should be undertaken with reference to user requirements; processing does not have just one end goal; data sets after processing will differ according to the user requesting the data.
- Currently, data processing requires expert user checking of outcomes.
- Data processing should not remove data from the database but flag data so that it can be identified and used as appropriate.

Data Storage

- Monitoring data should be stored with the user in mind, rather than being stored from the perspective of data acquisition.
- Repository monitoring will be undertaken over long periods, so flexibility is required in monitoring databases, and regular upgrades to software and hardware should be planned.

Modelling and Visualisation

- The work in MODATS has demonstrated the potential for hybrid models that combine physics-based and data-driven-modelling to provide a basis for analysis of monitoring data during repository operation.
- The work in MODATS has also demonstrated benefits of communication through visualisation, including benefits to expert users and to university students. Communication to non-expert actors using virtual tours should include provision of contextual information, task-based activities and independent control of the tour to maximise learning.

Digital Twins

- A repository digital twin is a virtual model of part of a repository that is updated automatically to address specific objectives.
- A digital twin of a repository cannot replicate the full reality of the underground (i.e. including all processes relevant to post-closure safety), as the repository system is subject to irreducible uncertainty. Instead, digital twins should focus on specific aspects of a repository (for example a twin of the underground infrastructure) defined by the objectives of the digital twin project.
- Repository digital twins are not “one size fits all”, but come in different forms depending on the objective for which they are developed.
- In MODATS, algorithms have been developed for surrogate models and for implementing physics-informed machine learning, representing prototypes of repository digital twins.
- Future digital twins might help in global parameter sensitivity analysis and related parameter uncertainty quantification, and the digital twins produced in MODATS have provided an illustration of how such digital twins might be developed.
- A comprehensive review of the potential applications of digital twins in repository programmes is required to identify the ways in which digital twins can practically enhance repository programmes.
- Further use of the MODATS Reference Experiments, including development of their digital twins and their underlying surrogate models is required to establish the manner in which digital twins can be applied during repository monitoring.

Workflow for Data Handling from Acquisition to Decision Support

The work in MODATS has highlighted that monitoring programmes (e.g., monitoring programmes associated with URL experiments) typically follow a bespoke data management process. The developments in data management made during MODATS have allowed the elaboration of an overall data management process for using monitoring data to provide enhanced system understanding (Figure E.1). This data management process is considered as a common approach that can be applied in all programmes. Development of this process, and application of it during repository monitoring, is expected to improve data management in repository monitoring programmes and thereby improve the reliability of the data produced. The various test cases and integrated discussion of advancement made in data management, modelling and visualisation provided in this report illustrate the application of this process.

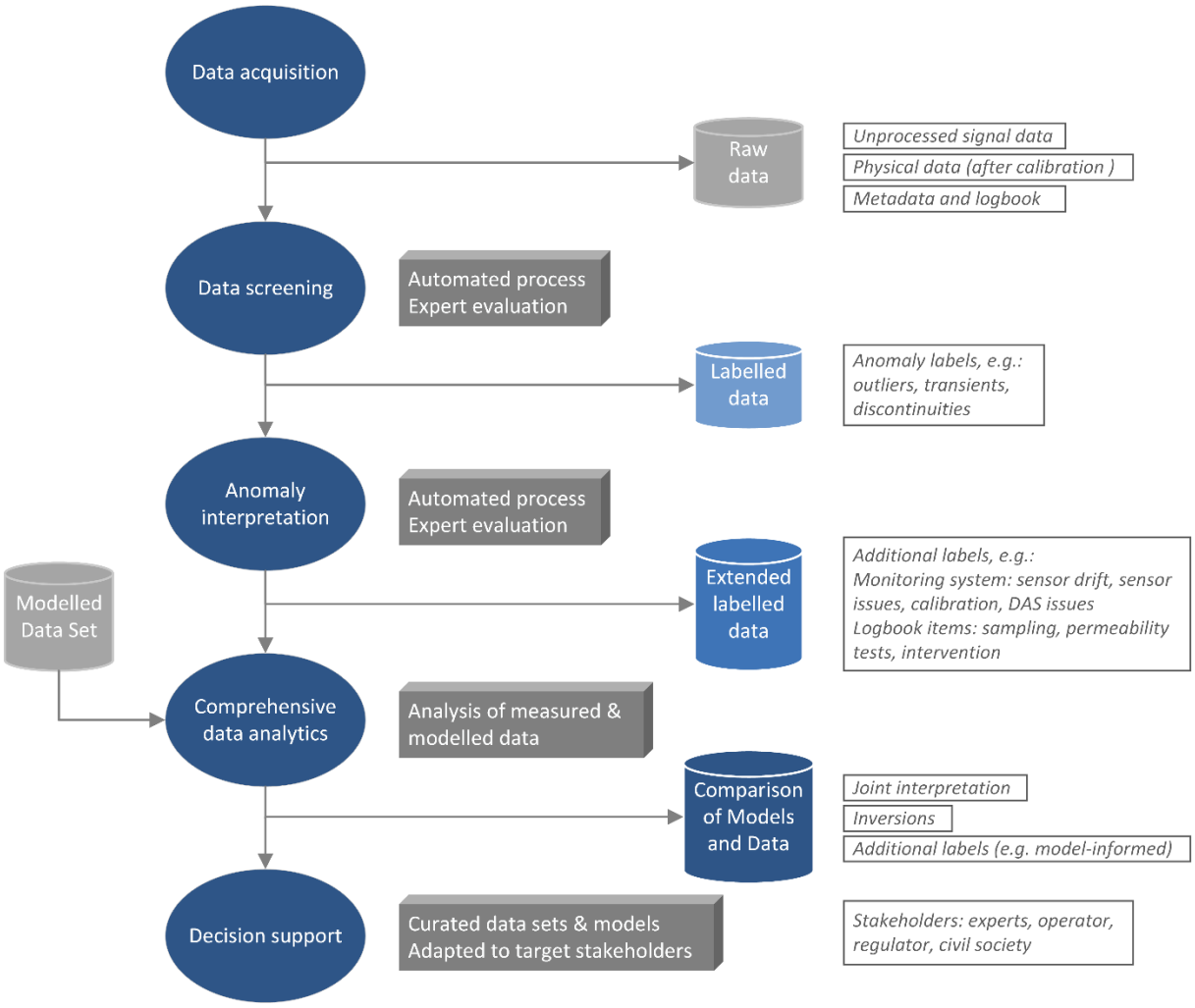


Figure E.1 - Proposed workflow for data handling from acquisition to decision support.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANN:	Artificial neural network
API:	Application programming interface
BIS:	Borehole information system
CG:	Connecting Gallery
CNN:	Convolutional neural networks
COx:	Callovo-Oxfordian
CPU:	Central processing unit
CSV:	Comma-separated value
DDM:	Data-driven modelling or model
DDML:	Data-driven machine learning
DECOVALEX:	Development of coupled models and their validation against experiments
DT:	Decision tree
DTS:	Distributed temperature sensing
EBS:	Engineered barrier system
EDZ:	Excavation damage zone
ETN:	European Thematic Network
EURAD:	European Joint Programme on Radioactive Waste Management
FE:	Full-Scale Emplacement
FEIS:	FE Information System
FO:	Fibre optic
GAS:	Mechanistic understanding of gas transport in clayey materials
GBM:	Granulated bentonite material
GSIS:	GeoScientific information system
GSL:	Galson Sciences Limited
GUI:	Graphical user interface
HADES:	High Activity Disposal Experimental Site
HITEC:	Influence of temperature on clay-based material behaviour
HLW:	High-level waste
HRL:	Hard Rock Laboratory
IAEA:	International Atomic Energy Agency
JSON:	JavaScript Object Notation
kNN:	k-nearest neighbour
LAN:	Local area measurement network
LIMS:	Laboratory information management system
LSTM:	Long short-term memory

MAE:	Mean absolute error
ML:	Machine learning
MODATS:	Monitoring equipment and data treatment for safe repository operation and staged closure
MSE:	Mean squared error
NaN:	Not-a-number
NAS:	Network-attached storage
NEA:	Nuclear Energy Agency
OTDR:	Optical time domain reflectometry
OECD:	Organisation for Economic Co-operation and Development
OFS:	Optical fibre system
OGS:	OpenGeoSys
PBM:	Physics-based modelling
PC:	Principal component
PCA:	Principal Component Analysis
PG:	PRACLAY Gallery
PIML:	Physics-informed machine learning
POD:	Proper orthogonal decomposition
POPLU:	Posiva Plug
PR:	Polynomial regression
PRACLAY:	Preliminary demonstration test for clay disposal
RF:	Random forest
RD&D:	Research, development and demonstration
RepMet:	Radioactive Waste Repository Metadata Management
RNN:	Recurrent neural network
SQL:	Structured query language
SUS:	System usability scale
SVM:	Support vector machine
THM:	Thermal, hydraulic and mechanical
THMC:	Thermal, hydraulic, mechanical and chemical
UMUX:	Usability metric for user experience
URL:	Underground research laboratory
VEIS:	Virtual Experiment Information System
VisCoSiR:	Tool for the Visual Comparison of Simulation Results
VTK:	Visualization Toolkit (acronym used for both software and file format)
WMO:	Waste management organisation

WP:	Work package
ZFC:	Connected fractured zone (acronym derived from French language term)
ZFD:	Diffuse fractured zone (acronym derived from French language term)

1. Introduction

Author: Matt White (Galson Sciences)

1.1 Background

Repository monitoring is motivated by the desire to inform decision making, strengthen understanding of system behaviour and to build further confidence in the repository safety case [1]. The Monitoring Equipment and Data Treatment for Safe Repository Operation and Staged Closure (MODATS) work package (WP) of the European Joint Programme on Radioactive Waste Management (EURAD) is conducting research, development and demonstration (RD&D) into: monitoring data acquisition, treatment and management; use of monitoring data to enhance system understanding, including development of digital twins; interactions with civil society and other stakeholders; development of monitoring technologies; and development of knowledge regarding repository monitoring. MODATS is building on previous international collaborative RD&D activities, including a European Thematic Network (ETN) [2], and the MoDeRn [3] and Modern2020 [1] projects. The context to repository monitoring is captured in the MODATS state-of-the-art report [4], and a comprehensive overview of MODATS is provided in the WP synthesis [5].

The RD&D in MODATS is supported by existing information and data from underground research laboratory (URL) experiments, including five reference experiments, referred to as the MODATS Reference Experiments:

- ALC1605: A demonstration of the French reference disposal concept for high-level waste (HLW), which studies the impact of thermal loading on the metallic liner and the rock, and acts as a step in the qualification of the monitoring system design for the HLW concept. It is led by Andra in the Bure URL, France.
- The Full-Scale Emplacement (FE) experiment: The FE experiment investigates thermal, hydraulic and mechanical (THM) coupled processes at full scale, in repository-like conditions, to validate existing models. It also aims to verify the technical feasibility of constructing a disposal tunnel using standard industrial equipment. It is led by Nagra in the Mont Terri URL, Switzerland.
- The Posiva Plug (POPLU) experiment: The POPLU experiment was a full-scale test of a possible design for a disposal tunnel end plug component in the disposal concept for the spent fuel repository in Olkiluoto (Finland) and Forsmark (Sweden). It was led by Posiva and SKB in the ONKALO Facility, Finland.
- The Prototype Repository: The Prototype Repository is a full-scale field experiment in crystalline rock. The experiment aims to simulate conditions that are largely relevant to the Swedish/Finnish KBS-3V disposal concept for spent fuel. It is led by SKB in the Äspö Hard Rock Laboratory (HRL).
- Preliminary Demonstration Test for Clay Disposal (PRACLAY): The PRACLAY experiment is a large-scale experiment designed to study the impact of the heat generated by HLW on the host clay formation. It also looks at how excavation affects the behaviour of the clay. The experiment is conducted by EURIDICE and ONDRAF-NIRAS in the High Activity Disposal Experimental Site (HADES) URL in Mol, Belgium.

The focus of MODATS is monitoring during the operational phase of repository programmes to build further confidence in the long-term safety case. In particular, MODATS is focusing on confidence in monitoring data. There are particular challenges associated with monitoring data acquired during repository operations, including challenges associated with its management (including processing and storage), modelling and visualisation. These are most frequently associated with the long timescales (several decades) envisaged for monitoring programmes and the need to maintain the passive safety of the disposal system. Therefore, MODATS has undertaken research to develop methods through which monitoring data can be processed, stored, modelled and visualised. This has included preliminary work on the development of digital twins.

1.2 Objectives

The key objective of this report is to provide an integrated discussion of the advancements made in monitoring data management (processing and storage), modelling and visualisation in MODATS. The work done in separate activities (test cases) is described and the results and conclusions summarised. Integrated conclusions from the work are drawn. Section 1.3 provides the context to this work and Section 1.4 presents the overall objectives of the work and approach taken.

1.3 Data Management, and Modelling, Visualisation and Digital Twins in the Context of Repository Monitoring

The primary objective of monitoring is to provide information to assist in making decisions on how, when and if to implement various steps in the management of the repository system. With respect to long-term safety, the key purposes of monitoring of repository systems are [6 § 2]:

- To provide information for making management decisions in a stepwise programme of repository construction, operation and closure.
- To strengthen understanding of some aspects of system behaviour used in developing the safety case for the repository and to allow further testing of models predicting those aspects.
- To provide information to give society at large the confidence to take decisions on the major stages of the repository development programme and to strengthen confidence, for as long as society requires, that the repository is having no undesirable impacts on human health and the environment.

A requirement to conduct a programme of monitoring prior to, and during, the construction and operation of a disposal facility and after its closure, if this is part of the safety case, is included in the IAEA Specific Safety Requirements on disposal of radioactive waste [7]. Recommendations and guidance on how to comply with the safety requirements is provided in the IAEA Specific Safety Guide on geological disposal facilities for radioactive waste [8], and in the IAEA Specific Safety Guide on Monitoring and Surveillance of Radioactive Waste Disposal Facilities [9].

It is recognised that repository monitoring programmes will have to respond to the specific national context under which the programme operates; the ETN [2] recognised that there are different approaches to repository monitoring in waste management programmes, depending on the objectives, and specifically “the extent to which monitoring is seen as confirming processes related to evolution of the repository and its long-term safety” [2 § 7.3]. The factors that are typically considered in the objectives of the repository monitoring programme include [2 § 7.3]:

- Waste type and engineered barrier system (EBS) properties and expected performance, which affect the extent to which parameters related to long-term performance can be monitored.
- Implementation strategy, including plans for progression from one step to the next, including periods of observations in (open) underground structures.
- Regulatory regime and requirements.
- Degree of concept flexibility.
- Political and public expectations.

Driven by these factors, tailored approaches are being considered to monitoring in each repository programme. These include intensive monitoring of disposed waste in support of reversibility, a focus on monitoring of the geosphere to demonstrate suitable boundary conditions for long-term safety, and monitoring in representative locations to ensure the passive safety of emplaced waste. All of these approaches involve collection of data on multiple processes over multiple decades. Following acquisition, data will need to be managed effectively and efficiently during the operation of repositories.

Monitoring data management is generally defined here as the processes and procedures that ensure the acquired monitoring data can be used to fulfil the objectives of monitoring programmes. Monitoring data management involves the processing of the data, organisation of the data (including metadata) into

databases with appropriate formats and structures, and supply of data for modelling and visualisation, and so that it can be used to support decision making[1].

1.3.1 Data Processing

Monitoring data are prone to errors owing to, e.g., sensor fouling (accumulation of deposits), calibration drifts, data transfer problems, and configuration errors in the data acquisition system. Data errors, referred to here as anomalies, have to be detected and handled. Invalid data could fall outside the plausible range of values, in which case, it may be easy to identify. Alternatively, they could sit within the plausible range and be difficult to differentiate from valid data [10 § 7.2].

Processing of monitoring data may include combining data acquired at different rates or reducing data (e.g., averaging of data collected over relatively short periods, so that it reflects a value representative of a longer period) so that it is more suitable for its intended use.

Data processing is routinely applied in the URL experiments using a range of bespoke methods. Often, the methods are developed for the specific experiment or for a specific sensor used in an experiment. Such bespoke methods may not be appropriate for repository monitoring data sets. In particular, repository monitoring data need to be analysed in a systematic and comprehensive way that allows for the systematic identification of anomalies related to sensor ageing and malfunction, outside influences, and deviations of the repository system from expected behaviour, and provides data for modelling and visualisation in a quality-assured manner.

Systematic methods and tools are therefore required to process repository monitoring data, so that it is suitable for the purpose for which it was acquired. In particular, methods and tools are needed to collate, clean and validate monitoring data, and visualise and report data, so that the sensor signals can be correctly interpreted. RD&D in MODATS has developed several such data processing tools.

1.3.2 Data Storage

Repository monitoring will be conducted over long periods, potentially stretching into many decades. Owing to the long timescale, and the potentially large quantities of data that could be collected, methods need to be developed to efficiently and effectively store monitoring data to ensure its availability, and allow its effective and efficient use, in the long term.

Data storage includes archiving of both the quantitative outcomes from monitoring (i.e., the parameter values), plus the metadata required in order to interpret the parameter values. Metadata is information that enables data to be managed in a structured manner [11]. This ensures that data quality and confidence in the stored data is maintained, and that data remain suitable for use in the future. Examples of metadata useful in the context of a repository monitoring programmes include information on the installation of the monitoring system and information on activities that might influence the parameter values recorded (e.g. parallel construction activities or periods of electrical outages).

RD&D in MODATS developed data storage methods that will allow monitoring data and metadata to be effectively managed, for use as required throughout the repository programme.

1.3.3 Modelling, Digital Twins and Visualisations

In repository programmes, where considerable amounts of data will be generated, it is necessary to have appropriate tools to reliably analyse monitoring data to support programmatic decision making and the related objectives. It is envisaged that evaluation of system behaviour using monitoring data will be undertaken on both a continuous and a periodic basis [12]. Continuous evaluation might focus on specific parameters and checking to see if parameter evolution is consistent with the evolution assumed

[1] Data management, modelling and visualisation incorporate “data analysis”, which is defined as the process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. In this report, we do not use the term “analysis” but instead use more specific terms.

in the safety case. Periodic evaluation would involve an integrated consideration of the full range of monitoring data, and would be focused on supporting updates to the safety case, including the underpinning safety assessment.

Modelling

Evaluation of the impacts of the coupled effects of mechanical deformation, fluid and gas flow through the repository and thermal loading from the waste is an important aspect of the safety case of a repository. Understanding of these impacts is gained through RD&D, including the use of numerical models capable of simulating coupled thermal, hydraulic, mechanical and chemical (THMC) processes and monitoring data [13]. Most waste management organisations (WMOs) are in the early stages of their repository programmes, and, therefore, the experience of using repository monitoring data in coupled models to support decisions is limited. There is, however, considerable experience of using URL experiment monitoring data in coupled numerical models. In particular, the Development of Coupled Models and their Validation against Experiments (DECOVALEX) project is a long-term international research collaboration for advancing the understanding and modelling of coupled THMC processes in geological systems [14]. Summaries of recent research in DECOVALEX are presented in [13, 14 and 15].

Recently, data science has emerged as an alternative to the approaches adopted in coupled numerical modelling [16]. Data science involves using scientific methods, processes, algorithms and systems to extract or extrapolate information and knowledge from previously collected data (training data) to identify the characteristics of a data set and to predict the future trend. This is referred to as machine learning (ML). A sub-set of ML is data-driven modelling (DDM). Coupled numerical models, referred to as physics-based modelling (PBM) in data science, assume that a physical model describing the behaviour behind processes is available and sufficiently accurate to understand the operation of processes and (in some cases) to predict future behaviour. Originally, data science was suited to data-intensive applications such as image processing and pattern recognition. In recent years, physics-informed ML[2] methods have been developed to an extent that they accelerate numerical simulations and have become directly usable for process-driven areas including application to modelling of THMC processes in repositories.

The spectrum of ML methods applicable to modelling of repository THMC processes is extensive, presenting a diverse array of techniques, for example:

- Linear regression models constitute a fundamental approach for straightforward parameter prediction[3] tasks. These models establish a relationship between input features, such as historical data and geographical information, and predictions [17].
- Time series models, including autoregressive integrated moving average [18] and exponential smoothing [19] are particularly applicable to short-term predictions.
- Deep learning models are currently widely applied in data science applications, particularly recurrent neural networks (RNNs) and convolutional neural networks (CNNs). RNNs, which are proficient in capturing temporal dependencies within time series data, prove effective for short-term predictions [20]. Conversely, long short-term memory (LSTM) networks, a subset of RNNs, Excel in modelling prolonged dependencies within time series data [21, 22]. CNNs, which are adept at addressing spatial dependencies in predictions, find application in predicting parameter values across geographical regions [23, 24]. Additional approaches used for predictions include random forests [25], which amalgamate multiple decision trees to enhance accuracy.
- Support vector machines [26], which are applicable in the presence of non-linear and high-dimension data, present another noteworthy avenue for parameter predictions.

[2] Physics-informed machine learning allows scientists to use prior knowledge to help the training of the algorithm, making it more efficient. This means it will need fewer samples than a pure DDM to train it well or to make the training more accurate.
[3] Prediction is referred to as “forecasting” in data science.

- Gaussian process models [27, 28] provide a means to encapsulate uncertainty and provide probabilistic predictions.
- Hybrid models [29] constitute an intriguing prospect, involving a fusion of ML models with physical models to bolster prediction accuracy.

The potential benefits offered by these approaches are more rapid modelling, which may benefit the continuous evaluation of monitoring data, and greater ability to model parameter evolution across 3D space. A particular opportunity is the use of the ML algorithms in digital twins.

Digital Twins

The Defense Acquisition University definition of digital twin, commonly used in defence, aerospace and related industries, (quoted in [30]) is: "*an integrated multiphysics, multiscale, probabilistic simulation of an as-built system, enabled by Digital Thread, that uses the best available models, sensor information, and input data to mirror and predict activities/performance over the life of its corresponding physical twin*".

There are three important parts in the digital twin of an object [31]:

- A model of the object.
- An evolving set of data relating to the object.
- A means of dynamically updating or adjusting the model in accordance with the data.

Digital twins have most frequently been used to support the production and maintenance of structures, especially engineered structures such as aircraft, bridges, and machinery. However, in recent years, they have begun to be developed in other industries. The coupling of digital twins with data science applications provides significant opportunities for the expansion of the use of digital twins, including in repository programmes. These opportunities could include development of a replica of a disposal system that progressively becomes more detailed and more closely representative of the real-world system as more data is collected and information is derived from the acquired data. Such a digital twin requires feedback from monitoring data to the underpinning models contained within it.

Geological disposal presents unique challenges to the development of digital twins owing to the mix of engineered and natural structures, the presence of spatial and temporal heterogeneities, and the extent to which detailed information can be collected without disturbing the passive safety of the system. There is a need also to develop the methodologies and software tools through which digital twins can be developed and evolve during the operational phase in response to collection of repository monitoring data.

Visualisation

A key feature of some digital twins is the visualisation of the system structure, the data acquired used in modelling, and the modelling results. Visualisation has the potential to add significant value to the analysis of data by allowing the integrated 4D comparison of different data sets against the domain of interest. However, this requires development of approaches for modelling and viewing the system structure and for integrating data that may be stored in different formats. Developing value from the visualisation requires research into the use of visualisations to enhance expert understanding, to communicate to stakeholders and educational purposes.

Modelling, Digital Twins and Visualisations in MODATS

In MODATS, RD&D was undertaken with the objective of developing data-driven methodologies to model monitoring data, developing prototype digital twins and developing visualisation approaches to further understanding, to support communication and for use in education. The MODATS Reference Experiment data sets supported this work.

1.4 Overall Objectives and Approach

The work in MODATS on data management, modelling and visualisation was based on six test cases that each addressed different aspects of data management, modelling and visualisation, as follows:

- ALC1605 Test Case (undertaken by Amvalor, Andra and ENSAM): The ALC1605 test case used thermal monitoring data to develop methods for combined physics-based and data-driven-modelling of repository evolution.
- FE Experiment Test Case 1 (Nagra): Nagra's FE experiment test case used data from the FE experiment to progress approaches to quality control, data processing, and data storage.
- FE Experiment Test Case 2 (PSI): PSI's FE experiment test case focused on data use, and used temperature and relative humidity sensor data to develop ML-assisted physically-based 3D heat transport modelling.
- FE Experiment Test Case 3 (UFZ): UFZ's FE experiment test case focused on the development of interactive visualisation systems, and visualisation tools for the comparison of simulation and observation data.
- POPLU and Prototype Repository Experiments Test Case (VTT): The POPLU and Prototype Repository experiments test case developed tools and approaches for managing monitoring data, and provided illustrations of how these tools and approaches could be applied during repository monitoring.
- PRACLAY Experiment Test Case (EURIDICE): The PRACLAY test case developed a structured and uniform approach to data processing and storage of monitoring data.

The overall aims of the test cases are to provide improvements to the data processing, modelling and visualisation workflow for repository monitoring data. This includes harmonising data management approaches and producing common methods to improve the reliability of the data and transparency in the methods used to manage it. This should improve the confidence in the data for all stakeholders. The test cases have also sought to take the first steps in developments of digital twins based on monitoring data, and to evaluate the potential for digital twins to support more rapid evaluation of monitoring data (e.g. support continuous evaluation), to improve the accuracy of modelling in 3D and to support the detection of monitoring system malfunction.

Within MODATS, each test case was undertaken separately. However, regular meetings were held to comment on development of the test cases, including meetings dedicated to the test cases and meetings focused on the entire WP. The record of the test cases is captured in this document, along with the publication of scientific papers that provide additional details of the work undertaken. Initial drafts of test case descriptions have been reviewed and edited by partners within MODATS. The individuals that have contributed to each section of the document are identified at the start of the relevant section.

1.5 Report Structure

This report is presented in seven sections. Following this introduction (Section 1), separate sections (or sub-sections) are presented for each test case respectively (Sections 2-5). For each test case, a description of the Reference Experiment is provided, the objectives and methods used in the test case are described, the results presented, and conclusions drawn.

Section 6 provides an integrated discussion of repository monitoring data management based on the outcomes from the test case work undertaken in MODATS. The discussion is structured around the topics of data processing, data storage, data use, and digital twins.

Section 7 provides the conclusions from the MODATS work on data management, modelling and visualisation.

2. ALC1605 (Amvalor, Andra and ENSAM)

Authors: David Muñoz and Francisco Chinesta (AMVALOR), Julien Cotton and Johan Bertrand (Andra), Anoop Ebey Thomas (ESI Group)

Editor: Matt White (Galson Sciences)

2.1 Introduction

ML has emerged as a valuable tool for temperature forecasting [32, 33], owing to its adeptness in handling intricate patterns and vast data sets. The spectrum of ML methods applicable to temperature forecasting is extensive, presenting a diverse array of techniques. Work in this test case focused on the most prevalent methods employed in this domain, as introduced in Section 1.3.3. In particular, the hybrid approach combining ML and physics-based models is emphasised, demonstrating its potential for advancements in temperature forecasting.

The test case investigates challenges for complex and instrumented systems (multi-scale, multi-physics, and multi-media), focusing on practical applications of ML methods. It is based on data from temperature monitoring of a HLW cell demonstrator (called ALC1605) implemented at Andra's Bure URL. This endeavour encompasses a blend of numerical simulation, real structure measurements, and data mining techniques. This approach represents a pioneering advancement in the realm of nuclear waste management.

In this work, the term "digital twin" is used to describe the digital replication of physical assets, derived from real-life data collected through various sensors and monitoring technologies while the asset is in operation. This process typically involves creating an analytical, data-driven model (referred to as the twin) to analyse, update, and manage the performance of its physical counterpart. In contrast, the term "hybrid twin" is associated with solutions that involve constructing an additional, complementary virtual model. This supplementary model is inherently physics-based and delineates cause-and-effect relationships.

This section is structured as follows:

- Section 2.2 provides a description of the ALC1605 instrumented heating cell.
- Section 2.3 describes the methodology used to construct the hybrid twin model and its constituent components.
- Section 0 describes the results of applying the model, featuring illustrative examples.
- Section 0 explores the applications of the methodology.
- Section 2.6 presents the conclusions.

2.2 Description of the Reference Experiment: ALC1605

2.2.1 HLW concept

Andra plans to dispose of HLW in the Cigéo geological disposal facility in the Callovo-Oxfordian Clay of the Paris Basin[4]. The design of the facility envisages that HLW will be emplaced in small-diameter tunnels referred to as disposal cells (Figure 2.1). The disposal cells will be lined with a low-carbon steel sleeve to facilitate the emplacement process and retrievability of the waste if so desired in the future. This design formed the basis of the ALC1605 test case.

[4] It is planned that long-lived intermediate-level waste will also be disposed of in the Cigéo facility.

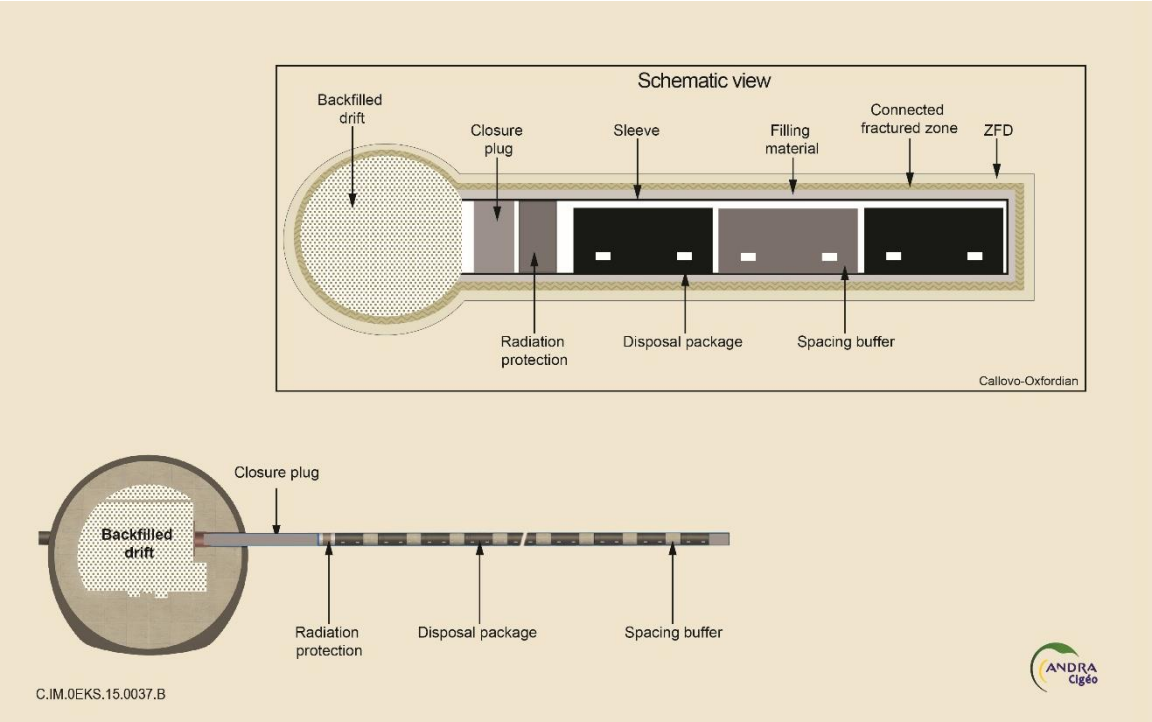


Figure 2.1 - Illustration of the HLW disposal cell design. ZFD is the French acronym for the discrete fracture zone, part of the excavation damage zone.

2.2.2 ALC1605

The ALC1605 demonstrator is designed to be as similar as possible to the HLW reference concept. Table 2.1 illustrates the main elements of the cell, in the reference design and in ALC1605.

Table 2.1 - Synthesis of the reference concept and ALC1605 features.

Component	Reference concept	ALC1605
Excavation technique	Cell totally excavated and then introducing of pipes in one time	
Excavation diameter (mm)	800 – 920	**920**
Pipe reference steel	API 5L X65 MS PSL 2	API 5L X65 MS PSL 2
Excavation orientation	σH	σH
Tilt	Cell axis, up: 2 % +/- 1 % in each point of the cell	
Azimuth	Cell's axis orientation: N155° or N335° (according to σH) +/- 1 %	
Pipe dimensions	Ø outside 660 mm (26'') or 762 mm (30'') Thickness: 25.4 mm (1'')	Ø outside 762 mm (30'') Thickness: 25.4 mm (1'')
Length	80-150 m	80 m
Cell bottom	Closed	First tube equipped with cement injection system

Component	Reference concept	ALC1605
Insert	Mobile insert with concrete station bloc for waste package transport's engine	Note envisaged
Bride	Provisory radioactive tight closer	Note envisaged
Pipe connection type	Clipped	Clipped with cement material proof joint
Surrounding filled materials	Yes	Yes
Water collecting device	Integrated to insert	Note envisaged
Cell closure device	Radiological closure device and bride	Determined during conception phase

The characteristics of ALC1605 are:

- A total length of 28.5 m, including a 6 m cell head and a 19 m "useful" part, with a 1 m overlap area between the liner and the insert.
- Excavated cavity head diameter 791 mm, with insertion of a diameter insert outer 767 mm (an annular space of 12 mm radius) and thickness 21 mm.
- The useful part excavated in diameter 750 mm, with the installation of a liner of diameter outside 700 mm (an annular space of 25 mm radius) and thickness 20 mm.
- A bottom plate, a plate at the top of the liner and a plate at the top of the insert.
- The completion of the heating over a length of 15 m between 10 and 25 m deep.

ALC1605 is instrumented to the extrados of the liner by:

- Two systems for measuring the convergence of optical fibre folders (Shirts No. 7 and 9).
- Three longitudinal lines have been equipped with fibre-optic sensitive cables for the measurement of axial deformations and temperature along the entire length of the cell (Folders 2 to 12).
- A temperature measurement section by PT1000 sensors (Folder n ° 7).
- A section of measurement of the deformations by vibrating rope (Folder n ° 9).
- A cell allowing geochemical measurements (Folder n ° 10).

Figure 2.2 shows the positioning of the instrumentation.

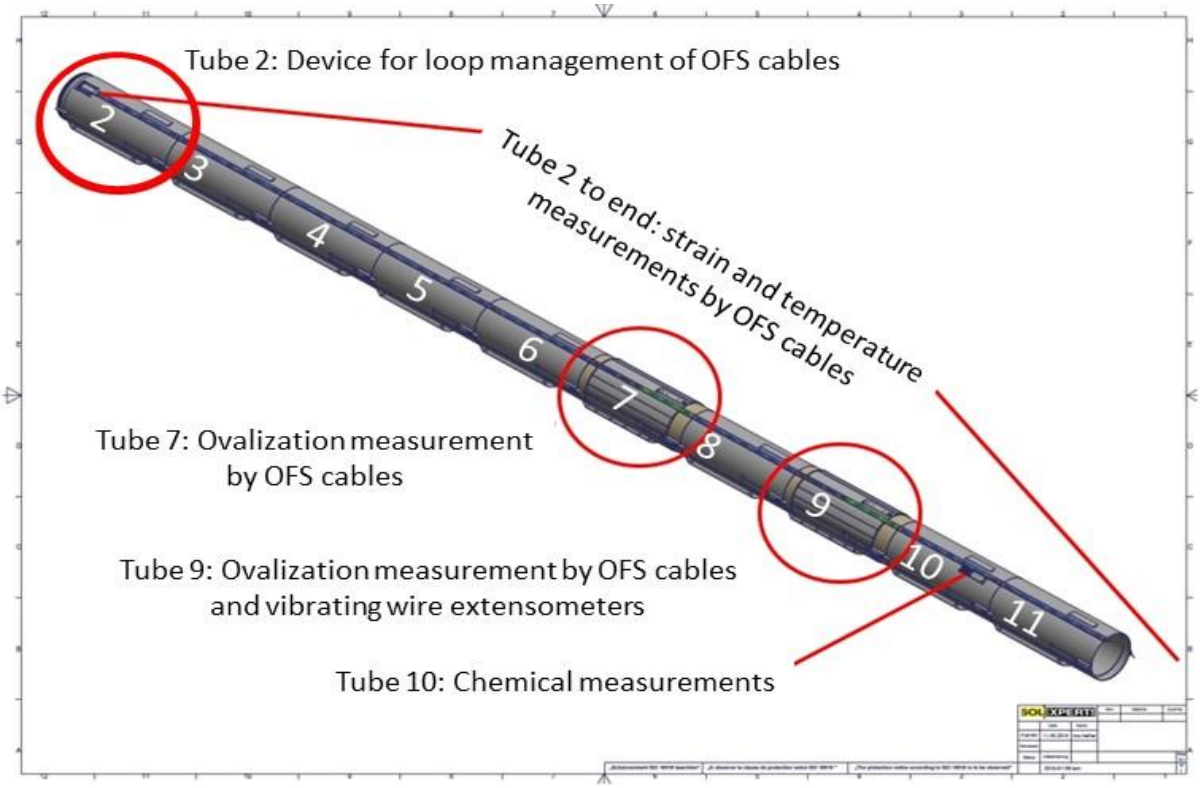


Figure 2.2 - ALC1605 scheme of connections and numbering of individual casing segments (tubes).

2.2.3 Objective of the ALC1605 demonstrators

The ALC1605 demonstrators simulates, as realistically as possible, the construction of a HLW disposal cell and emplacement of waste within it. The main aim of the ALC1605 experiment is to study the behaviour of a HLW cell conforming to the reference concept under heating. The specific objectives can be defined as follows:

- Study the impact of thermal loading on the thermomechanical behaviour of the metallic liner in the presence of filling material (direction and amplitude of ovalization, time to reach the blocking, axial thermal expansion, possible discharge). These data will complement those obtained in a parallel experiment (AHA1605 which was identical except there was no thermal loading) to improve the description of the rock / structure interaction in the modelling of the long-term behaviour of the cell liner in order, in particular, to verify the gap distance for the package handling.
- Studying the impact of thermal loading on the THM behaviour of rock in the near field beyond the excavation damaged area, and in the far-filed (i.e., beyond a few diameters of the tunnel) in the presence of filler material. This includes monitoring the temperature and interstitial pressure fields around the cell at different distances from the access gallery. A comparison with the measurements made during a previous experiment (ALC1604) will identify a potential impact of the filler material on the kinetics and amplitudes of thermal overpressures in the near-filed and far-field. In addition, the experiment investigates the evolution of permeability over time to study healing of the damage zone.
- ALC1605 is a step in the qualification of the monitoring system design for the HLW concept. The main objectives related to qualification were (i) to qualify the measurement technique, (ii) to validate the design and installation method *in situ*, (iii) to compare optical fibre system (OFS) data to reference sensors, and to assess system durability and potential drift, over several years during a thermal experiment.

2.3 Methodology

This section outlines the methodology used to develop hybrid twin models for the ALC1605 instrumented heating cell. First the operational principles of hybrid twin models are described, followed by an examination of its two key components: the PBM and the DDM.

2.3.1 Hybrid Twin Approach

A hybrid twin model aims to improve the accuracy of numerical simulations rooted in physics by integrating real-world measurements. We define "ignorance" as the deviation between the real-world behaviour and the outcomes produced by numerical simulations. Incorporating the ignorance model into these simulations enables us to replicate the actual application behaviour, as depicted in the equation below:

$$Y(X,t,\mu) = A(X,t,\mu) + B(X,t,\mu), \qquad \text{(Equation 2.1)}$$

where Y represents the observed real-world phenomena, while A denotes the physics-based solution, and B symbolises the realm of uncertainty or ignorance. Additionally, X signifies the spatial coordinates, t represents temporal values, and μ encompasses any additional parameters, including material properties, boundary conditions, and other relevant factors.

2.3.2 Physics-Based Model

The PBM plays a critical role in incorporating the fundamental principles of physics into forecasting of the temperature for the ALC1605 experiment, as it interfaces with the hybrid twin model. The PBM uses the finite element method (FEM) to derive the temperature distribution within the system.

The FEM mesh encompasses all components within the tunnel, as well as a substantial portion of the surrounding terrain (Figure 2.3). The mesh comprises 420,120 quadratic elements, each interconnected by a network of 1,701,835 nodes. Nonetheless, during the computation of the thermal problem, we streamline the process by employing a technique known as multi-point constraints [34, 35].

In addition to examining the tunnel's immediate surroundings, it is essential to focus on the tunnel itself, as it serves as a crucial element in the overall setup. Figure 2.4 illustrates the engineered components that constitute the tunnel. This visual representation not only showcases the tunnel's structure but also highlights the placement of the heating elements, which replicate the radioactive waste that will eventually be stored in such tunnels.

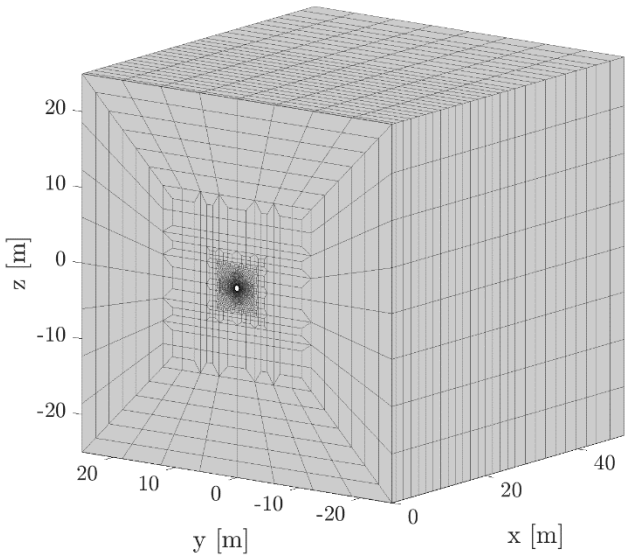


Figure 2.3 - Mesh employed to solve the thermal problem over the whole domain.

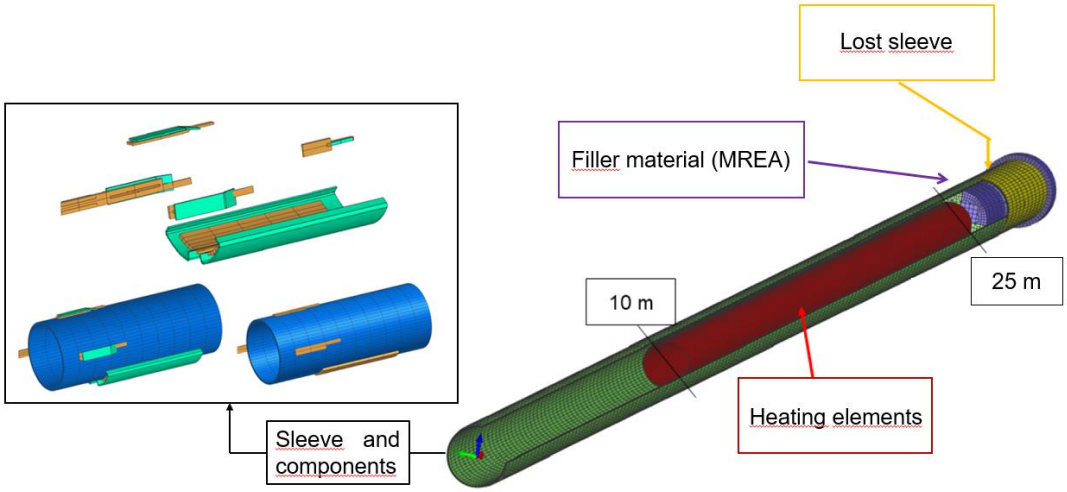


Figure 2.4 - Overview of the engineered components of the tunnel.

When examining the physics underlying the model, the intricate interplay between the surrounding air, the sleeves, and the heating elements is ascertained through a thermal calculation. This calculation employs a fluid/structure coupling framework that characterises the heat transfer fluid as an ensemble of one-dimensional elements. Within these elements, a diffusion-convection problem is solved to capture the heat exchange dynamics. In parallel, the thermal computation for the remainder of the domain adheres to the principles of the heat equation. This domain's thermal behaviour is discretised through a conventional FEM approach. This combined methodology ensures a robust and accurate representation of the complex thermal processes at play in the system, facilitating a comprehensive understanding and effective analysis of the model's behaviour.

When considering the material properties within this domain, it becomes essential to emphasise the distinctive characteristics across various regions. On one hand, the sleeves and their associated components, as illustrated in Figure 2.4, exhibit material properties consistent with steel. In contrast, the surrounding areas of the tunnel showcase diverse material characteristics contingent upon their specific regions. As shown in Figure 2.5, we present a frontal view of the gallery, concentrating on the region adjacent to the tunnel. Within this specific area, we delineate three distinct zones, each governed by unique material properties: the filler material, referred to as MREA, the connected fractured zone abbreviated as ZFC, and the diffuse fractured zone, referred to as ZFD. Extending beyond these defined zones, the material properties maintain consistency throughout the remaining domain, aligning with those attributed to Callovo-Oxfordian (COx) claystone. This differentiation stems from the impact on the adjacent zone during borehole drilling, resulting in modified material properties, particularly a decrease in thermal conductivity. Furthermore, it is crucial to note that the thermal conductivity of these regions is anisotropic, dependent on the rock's stratification layers. Hence, we distinguish between horizontal (λ^h) and vertical (λ^v) thermal conductivity. The main thermal properties assigned to different parts and components of the model is provided in Table 2.2. Turning attention to the thermal properties of the air, adjustments are made based on temperature variations.

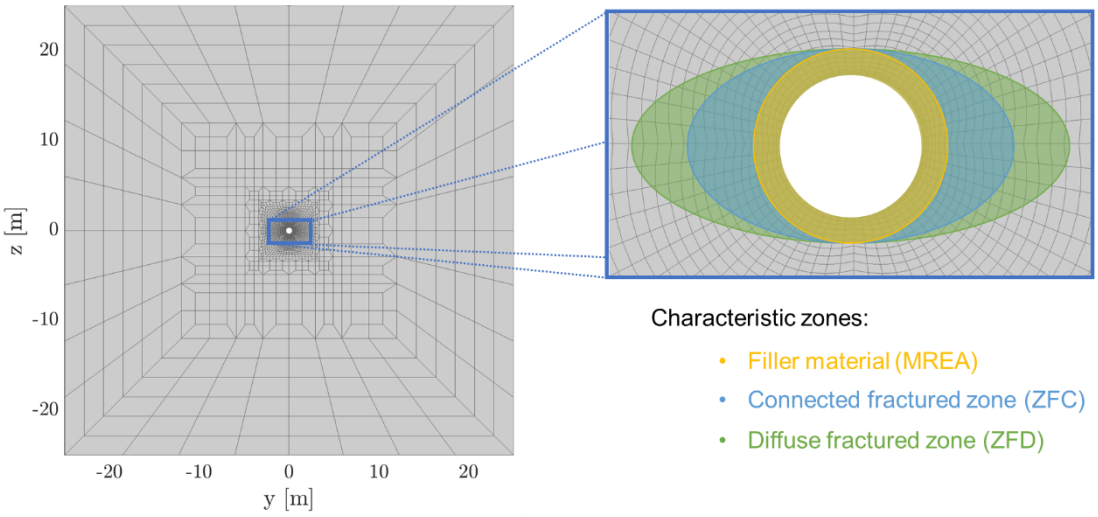


Figure 2.5 - A frontal perspective of the gallery, prominently featuring the area adjacent to the tunnel, while also emphasising the presence of three distinct material property zones.

Table 2.2 - Overview of the main thermal properties of the materials employed to define the model.

	λ ($W.m^{-1}.K^{-1}$)	**Cp ($J.Kg^{-1}.K^{-1}$)**	**ρ ($Kg.m^{-3}$)**
Steel	50.000	460	7800
MREA	0.890	2560	1160
ZFC	λ^h = 1.815 λ^v = 1.175	800	1700
ZFD	λ^h = 1.990 λ^v = 1.290	800	1700
COx	λ^h = 2.050 λ^v = 1.330	800	1700

However, our objective extends beyond the creation of a rigid, physics-based model reliant solely on a fixed set of parameters. We are striving for a more adaptable and versatile solution that takes into account real-time constraints and can accommodate changes in input parameters. To achieve this, our approach first involves the definition of the parameters that we need to explore. By considering the variations and bounds of these parameters, we define the multidimensional space that we intend to explore. Within this space, we conduct analyses and simulations to generate corresponding temperature fields. These temperature fields provide an understanding of the system's behaviour under different parameter settings, allowing us to adapt and optimise our solutions in response to dynamic real-world conditions. In particular, we focus on four parameters, whose bounds are defined as:

1. The power of the heating elements, $P \in [427.5,877.5]$ watts per element.
2. The thermal conductivity of the MREA, $\lambda_{\text{MREA}} \in [0.7120,1.0680]\ W \cdot m^{-1} \cdot K^{-1}$.
3. The horizontal thermal conductivity of the ZFC, $\lambda^h_{\text{ZFC}} \in [0.7120,1.0680]\ W \cdot m^{-1} \cdot K^{-1}$

4. The vertical thermal conductivity of the ZFC, $\lambda^{v}_{\mathrm{ZFC}} \in [0.7120, 1.0680]\ W \cdot m^{-1} \cdot K^{-1}$.

Once the mesh has been constructed and parameterised, temperature data is used to construct a surrogate model for further analysis and prediction. The surrogate model is a mathematical model employed to approximate the temperature field given a new set of parameters. To construct the surrogate model, we employ a systematic approach. First, we organise the collected temperature data into a structured matrix ($\mathbf{A}_i$), for given set of parameters, where the rows represent individual nodes, and the columns correspond to different time steps. This matrix serves as the foundation for our subsequent modeling efforts. The next step in our process involves the decomposition of the temperature field for each unique set of parameters. To accomplish this, we use a method known as CUR decomposition [36]. The CUR decomposition allows us to effectively extract and represent the essential components of the data, facilitating the creation of a compact and insightful surrogate model. In the context of this decomposition, careful consideration has been given to the selection of matrices to maintain the constancy of $\mathbf{C}$ and $\mathbf{R}$ irrespective of the chosen parameters. Meanwhile, matrix $\mathbf{U}_i$ is intentionally designed to dynamically adapt based on the specific parameter values. Consequently, we obtain the following expression for a given parameter set:

$$\mathbf{A}_i = \mathbf{C} \cdot \mathbf{U}_i\left(P, \lambda_{\mathrm{MREA}}, \lambda^{h}_{\mathrm{ZFC}}, \lambda^{v}_{\mathrm{ZFC}}\right) \cdot \mathbf{R}^{T}. \qquad \text{(Equation 2.2)}$$

Finally, the matrix $\mathbf{U}_i$ can be derived for any set of parameter values by using the sparse proper generalised decomposition framework [37]. A visual representation illustrating the dynamic evolution of the temperature field within a specific thermal system is presented in Figure *2.6*. This system is defined by several key parameters. The thermal conductivity of the MREA is set at 0.89 $W \cdot m^{-1} \cdot K^{-1}$, while the horizontal and vertical thermal conductivity of the ZFC are 1.815 and 1.175 $W \cdot m^{-1} \cdot K^{-1}$, respectively. The heating elements operate at a power level of 675 watts per element, corresponding to a heat flux density of 128 watts per square meter (W/m^2). The problem at hand commences with an initial condition where a uniform temperature of 21°C is imposed across the entire system. It is crucial to emphasise that the external surfaces of this system are assumed to be adiabatic, signifying no heat exchange with the surroundings.

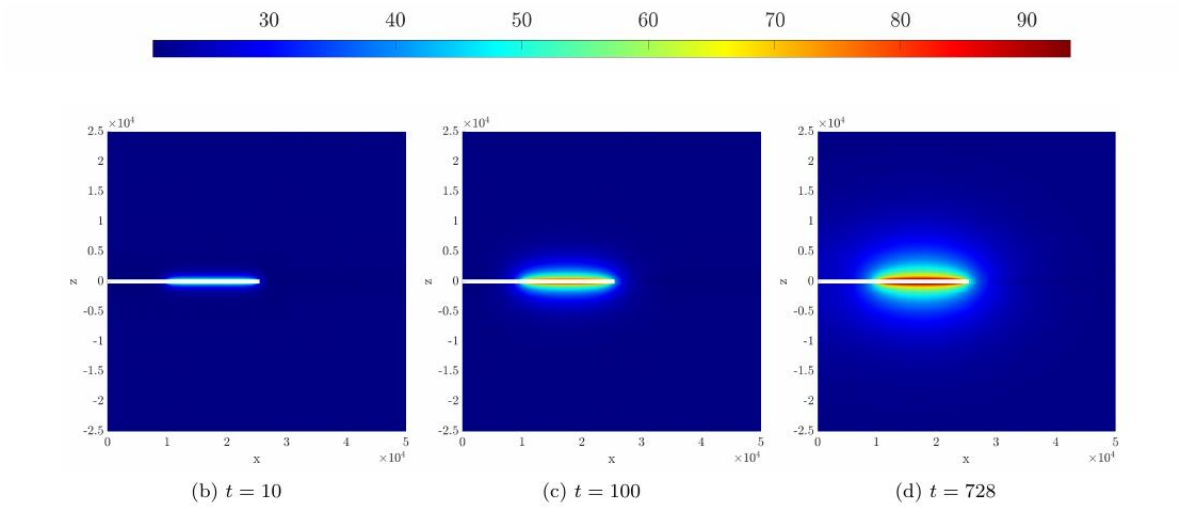


Figure 2.6 – Progression of the temperature distribution over time, as influenced by a specific set of parameters.

The essence of the hybrid twins approach lies in augmenting the physics-based solution with the available measurements at our disposal. The following section will explore the modelling of the disparities between the physics-based model described above and the actual behaviour observed through sensor data.

2.3.3 Data-Driven Model

The DDM plays a crucial role in representing a concept commonly referred to as "ignorance." As previously emphasised, the choice of employing the LSTM deep learning architecture stems from its good performance in handling sequential data. This decision is grounded in the aim to harness the capabilities of LSTM cells, intending to establish a robust methodology based on well-established temporal integration techniques—specifically, the forward Euler scheme [38, 39], as elucidated in Equation 3.3. This approach is commonly recognised as residual neural networks (NNs) [40, 41, 42]. LSTM cells seamlessly integrate short- and long-term memory units, incorporating an evanescent memory for the long-term path and a combination of long and short leads for the short-term memory response. Our primary objective is to capitalise on these characteristics to craft a highly effective model that excels in capturing the temporal dynamics inherent in the data sequences.

$$B_t = \frac{\partial B_t}{\partial t} \cdot \Delta t + B_{t-1}. \qquad \text{(Equation 2.3)}$$

In the context of our hybrid twin framework, the symbol B is used to represent the variable associated with the element of ignorance. The subscripts $t-1$ and t signify two consecutive time steps. The essence of our methodology aims to model the rate of change of ignorance over time. As a first step in this process, we seek to replace it with an unknown function, denoted as H, resulting in the following transformation:

$$B_t = H(B_{t-1}, \mu_t) \cdot \Delta t + B_{t-1}, \qquad \text{(Equation 2.4)}$$

where the function H not only depends on the current ignorance B_{t-1} but also on additional parameters denoted as μ_t. These additional parameters may include spatial coordinates, ambient temperature evolution, or material properties, all of which are integral to the model. In a similar vein, the function H can be substituted with a combination of two distinct functions, namely, f and g, expressed as follows:

$$H = f(B_{t-1}, \mu_t) \cdot B_{t-1} + g(\mu_t), \qquad \text{(Equation 2.5)}$$

The functions f and g serve as the components that the NN will model. Employing this approach is crucial to guarantee the stability of the temporal integration. Specifically, by constraining the values of the f function to be below 0, we aim to prevent the divergence of temporal integration [43]. This constraint is strategically implemented to ensure a robust and reliable performance of the NN in capturing and processing temporal dependencies. The NN architectures considered for describing functions f and g are both based on the use of LSTM layers combined with a deep dense neural network layer, as described in Table 2.3 and Table 2.4 for the ResNet. They were built by using Tensorflow Keras libraries.

Table 2.3 - The building blocks that model the function f.

Layer	Building Blocks	Activation
1	LSTM layer, Hidden size = 8	sigmoid + tanh
2	Flatten	no activation
3	Dense layer, #Neurons = 16	tanh
4	Dense layer, #Neurons=Sequence size	relu
5	Lambda layer returning -1 × inputs	no activation

Table 2.4 - The building blocks that model the function g.

Layer	Building Blocks	Activation
1	LSTM layer, Hidden size = 8	sigmoid + tanh
2	Flatten	no activation
3	Dense layer, #Neurons = 16	tanh
4	Dense layer, #Neurons=Sequence size	linear

Once the model has been defined, the next step is to train it effectively. However, before initiating the training process, it is imperative to preprocess the data to ensure its compatibility with the LSTM architecture. In its raw form, the data collected from the sensors typically resembles a tabular structure that encapsulates the readings from each sensor at various time intervals. To illustrate this concept visually, consider Figure 2.7, which presents a graphical representation of the sensor data. In this figure, the temperature readings recorded by each sensor across three distinct time steps are shown. This representation offers a clear visualisation of how the data is organised, with each sensor's temperature measurements evolving over time.

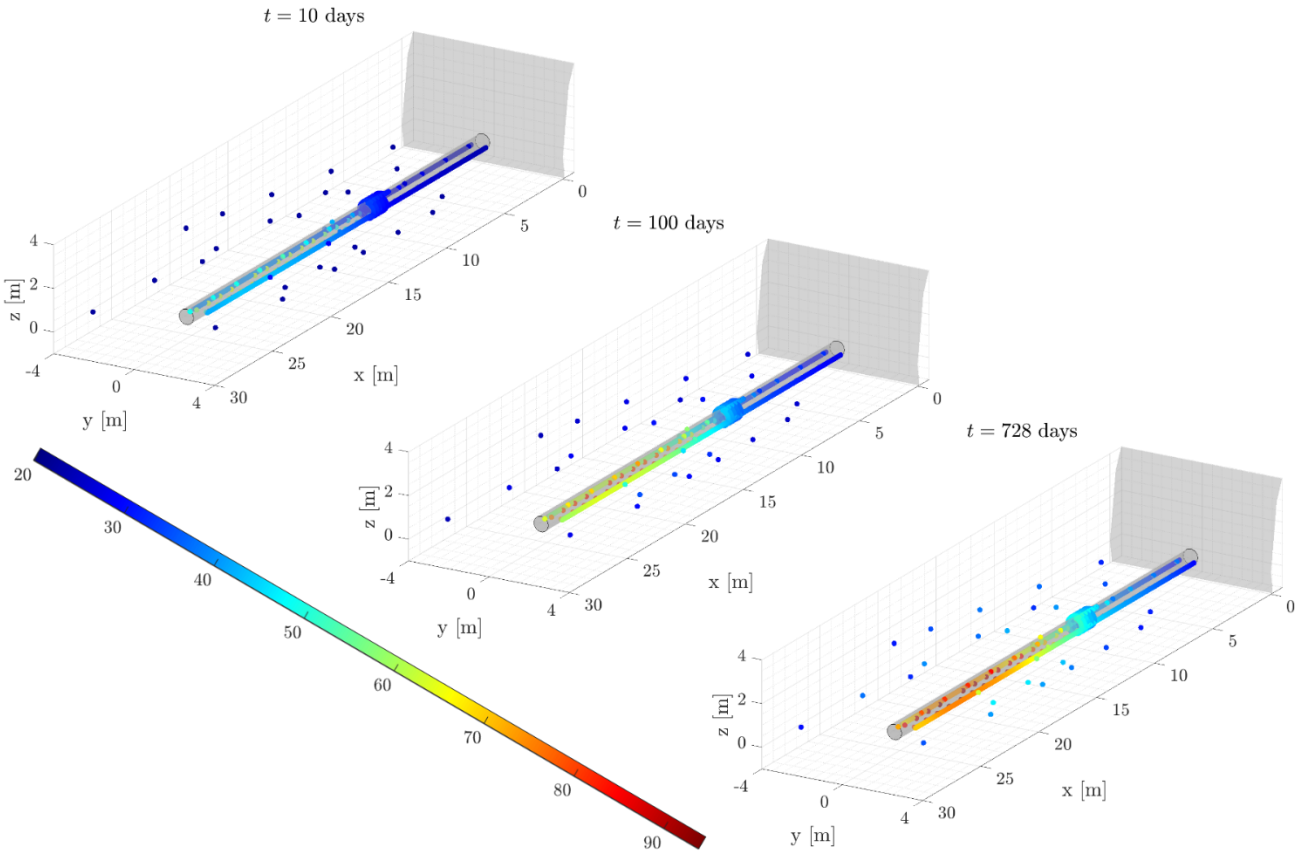
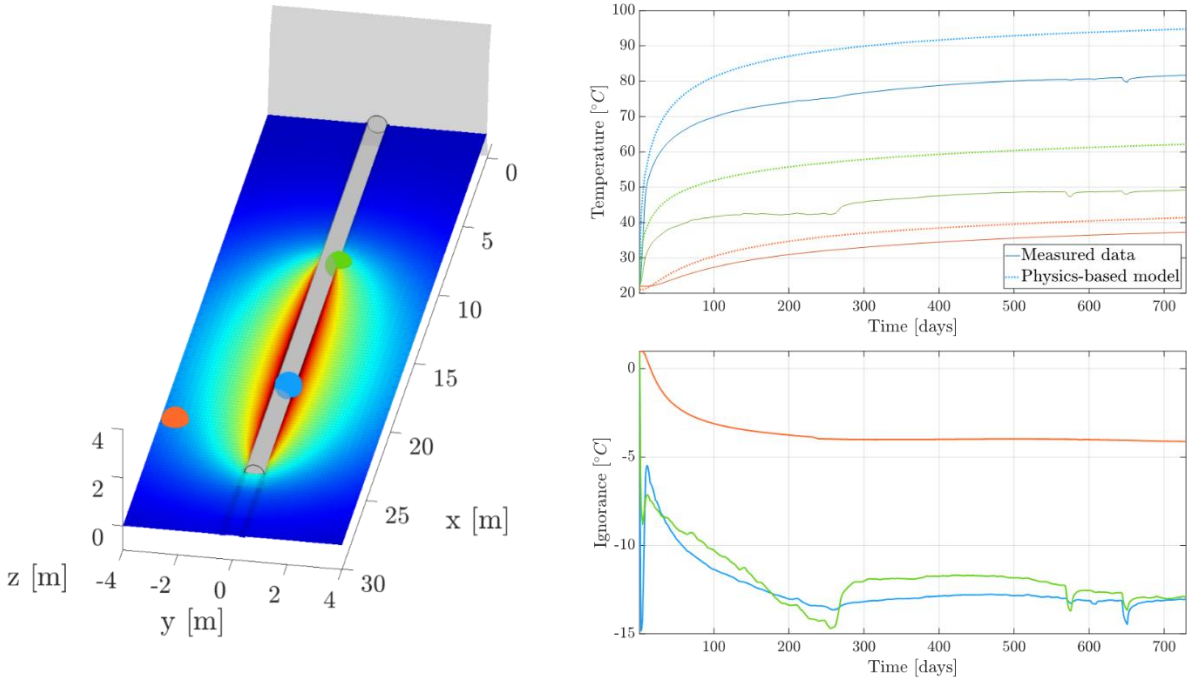


Figure 2.7 – Temperature measurements taken at three distinct time steps for each sensor. Each dot is a sensor that has been coloured to show temperature as indicated in the scale. The tunnel is shown in grey. The coloured line alongside the tunnel and in the tunnel mid-section are temperature data from monitoring using fibre optics.

To adapt this data for our model, it needs to be transformed into sequences of data. Figure 2.8a illustrates the position of all sensors while highlighting three randomly chosen sensors. Additionally, Figure 2.8b (top) depicts the temporal evolution of measured temperature for the set of three sensors. To enhance the visual representation, we have also included temperature data generated by the physics-based model at the same corresponding positions. By using the expression in Equation 3.1, we can determine the level of uncertainty denoted as "ignorance," represented by the variable B. This uncertainty is derived by subtracting the solution obtained through physics-based modeling from the temperatures measured experimentally. The change in this ignorance level over time is visualised in Figure 2.8b (bottom), for the same set of sensors.



(a) Position of three random sensors.

(b) Temperature evolution for a set of three sensors (top) and the corresponding ignorance (bottom).

Figure 2.8 – Position and evolution curves for three random sensors.

Once we have successfully defined the data, we are ready to proceed with training the model. The initial step in this process involves the proper partitioning of the data. To achieve this, we carefully select the first 250 timesteps for the purpose of training and validating our model, while the subsequent time steps are reserved for assessing the model's performance. Within the temporal data set, encompassing the initial 250 timesteps, we execute a partitioning operation to create distinct subsets for training and validation. This partitioning is accomplished with a specific split ratio of 80% for training and 20% for validation. This strategic division allows us to effectively evaluate the model's performance and refine its predictive capabilities.

When considering the myriad hyperparameters that influence the definition and training of a ML model, several key aspects deserve special attention. These encompass the ultimate size of the input sequence data, which can often be reduced to enhance flexibility in predictive capabilities. Additionally, the number of units residing within the LSTM cells plays a role in determining model complexity and performance. Further, the architecture's overall structure might entail auxiliary layers, necessitating careful calibration of the number of neurons in these layers to accommodate potential disparities in the sizes of the primary layers. The incorporation of an L_2 regularisation term is important for mitigating overfitting, ensuring the model's flexibility, and preventing excessive complexity. Lastly, one must fine-tune the batch size, a parameter that significantly influences the training dynamics, convergence speed, and memory requirements of the model. Therefore, meticulous selection and tuning of these hyperparameters are fundamental in the quest for effective and efficient ML models.

We train a model with a specific set of hyperparameters, enabling subsequent queries. This allows us to make forecasts over time by integrating Equation 3.4 into a closed-loop system, resulting in:

$$\begin{aligned} B_t &= H(B_{t-1}, \mu_t) \cdot \Delta t + B_{t-1} \\ B_{t+1} &= H(B_t, \mu_{t+1}) \cdot \Delta t + B_t \\ &\vdots \\ B_T &= H(B_{T-1}, \mu_T) \cdot \Delta t + B_{T-1} \end{aligned}$$

In the forecasting stage, we integrate insights from physics-based numerical simulations into our data-driven model. This fusion enhances the simulation's performance, enabling it to closely replicate real-world phenomena. The complete time integration scheme is represented in Figure 2.9.

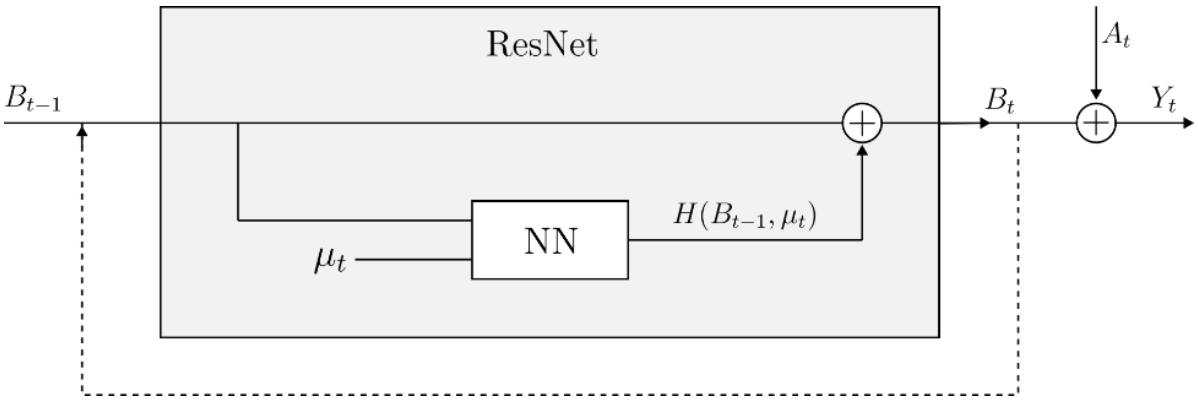


Figure 2.9 – Temporal integration scheme of the hybrid twin approach.

After forecasting the temperature, it is essential to assess the performance of the associated model. To accomplish this, we establish a set of error metrics. Initially, we define the error for a particular sensor at a specific time step as follows:

$$\varepsilon_{s,t} = \left(Y(X_s, t) - \hat{Y}(X_s, t)\right)^2, \qquad \text{(Equation 2.6)}$$

where Y denotes the measured temperature, and $\hat{Y}$ represents the predicted temperature. In this context, s serves as an indicator for a specific sensor, and t represents an individual time step. Consequently, we define the error for a particular sensor as the average of the sum of errors across all time steps, which can be expressed as follows:

$$\varepsilon_s = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{s,t}. \qquad \text{(Equation 2.7)}$$

Finally, the model's error is assessed by calculating the average summation of errors across all sensors, where each sensor's error is considered, as follows:

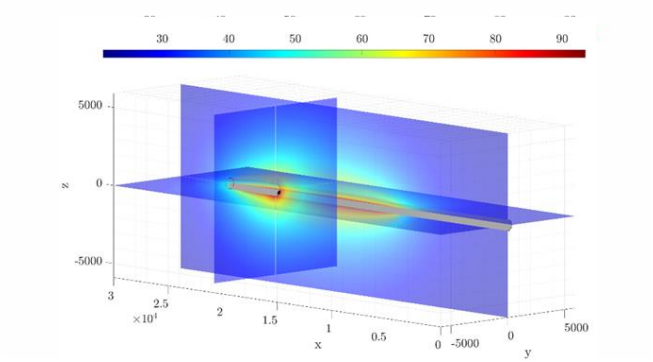
$$\varepsilon_m = \frac{1}{S}\sum_{s=1}^{S} \varepsilon_s. \qquad \text{(Equation 2.8)}$$

The choice of the best hyperparameter configuration is determined by the model with the lowest ε_m value.

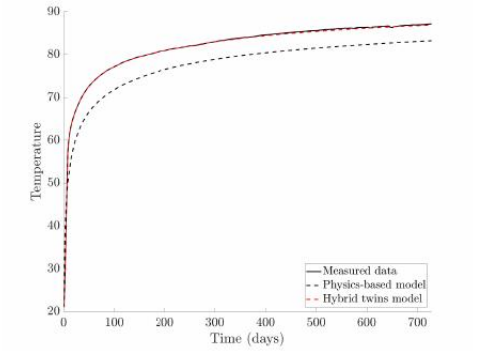
2.4 Results

This section illustrates the capabilities of the methodology, which leverages the hybrid twin approach, in the context of forecasting repository temperature. The insights and findings presented herein are derived from an exhaustive model evaluation process, where various hyperparameters were rigorously tested and refined. Through this extensive exploration, we identified the specific set of hyperparameters that yielded the lowest error, thus ensuring the utmost accuracy and reliability in our temperature forecasts.

The depicted visualisations in Figure 2.10a offer a comprehensive view of a random sample sensor within the test data partition. This sensor, depicted in black, serves as a representation of our data set's real-world observations. In this context, the figure presents the sensor's spatial position in the designated space. Furthermore, it overlays several planes where the physics-based solution is projected, specifically showcasing the temperature field at the final time step. On the other hand, Figure 2.10b illustrates the temporal evolution of temperature. Three distinct curves delineate this evolution: the solid black curve represents the actual measured temperature, the dashed black curve illustrates the temperature values projected by the PBM, and the dashed red curve signifies the temperature predictions generated by the hybrid twin strategy. Upon closer examination of these curves, it becomes evident that the integration of the DDM into the PBM significantly enhances its forecasting capabilities. This improvement underscores the value of adopting this hybrid approach over relying solely on physics-based methodologies. It's clear that this integration leverages the strengths of both data-driven and physics-based approaches, resulting in more accurate and robust predictions.



(a) Position of a random sensor along with the projection of the physics-based solution in different planes.



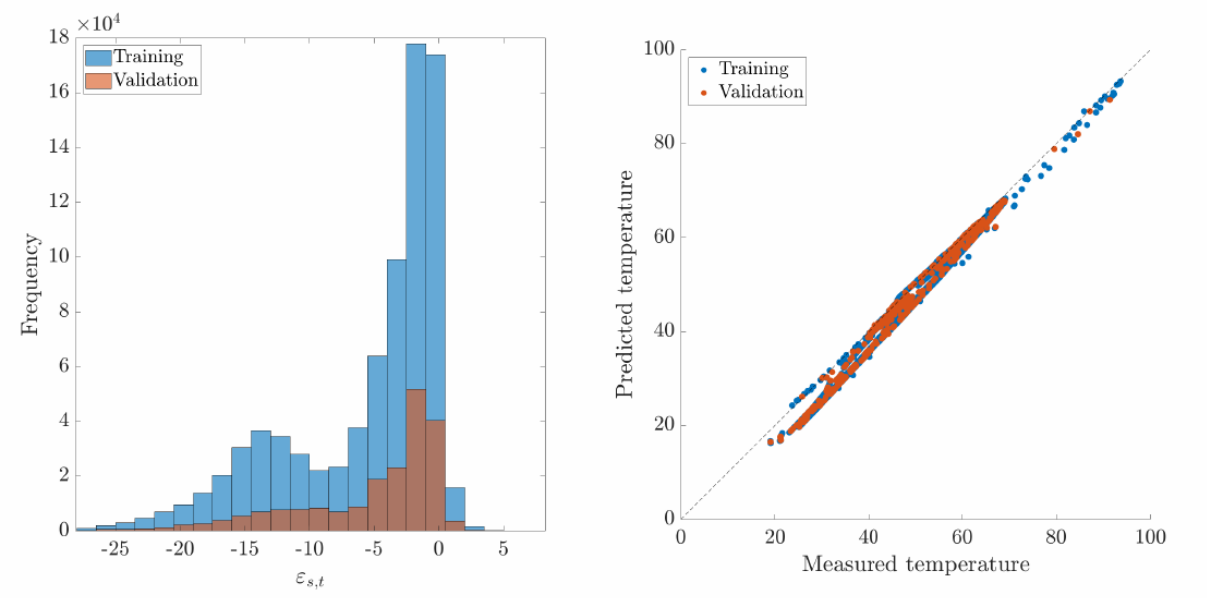
(b) Temperature evolution from the measured data (solid black), physics-based solution (dashed black) and hybrid twins approach (dashed red).

Figure 2.10 – Evaluation of different parts of the model for a random sample sensor within the test data partition.

To obtain a comprehensive understanding of the model's performance on a global scale, we introduce a novel error measure, which can be precisely described as follows:

$$\varepsilon_{s,t} = 100 \cdot \frac{\hat{Y}(X_s,t) - Y(X_s,t)}{Y(X_s,t)}. \qquad \text{(Equation 2.9)}$$

In Figure 2.11a, the frequency of errors is illustrated when predicting temperatures using the hybrid twin method. We looked at different periods and sensors to get a complete view of how well the method works over time. Conversely, Figure 2.11b illustrates the comparison between the measured temperature and the forecasted temperature using the hybrid twin approach, specifically for the final time step. This particular visualisation holds particular significance due to the accumulation of errors over the course of the temporal integration process. The last time step can offer invaluable insights into the model's behaviour, encapsulating the culmination of the forecasting performance, making it a critical focal point for evaluation and analysis.



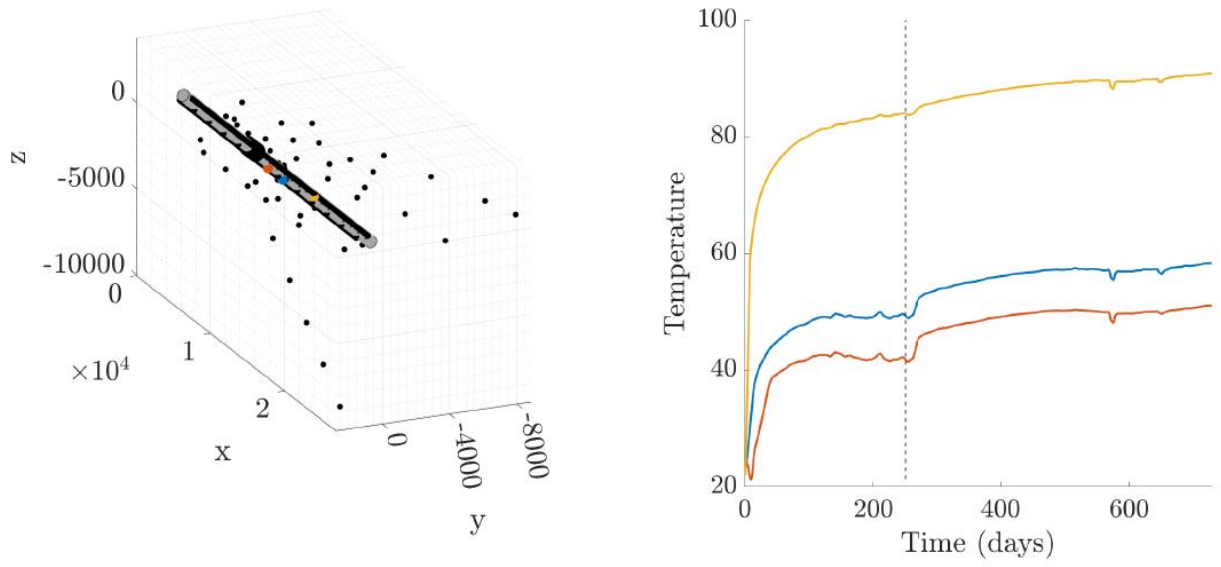
(a) Frequency of error measured as indicated in (9).

(b) Comparison between the measured temperature and the forecasted temperature for the last time step.

Figure 2.11 – Various methods for depicting the errors made by the hybrid twin model.

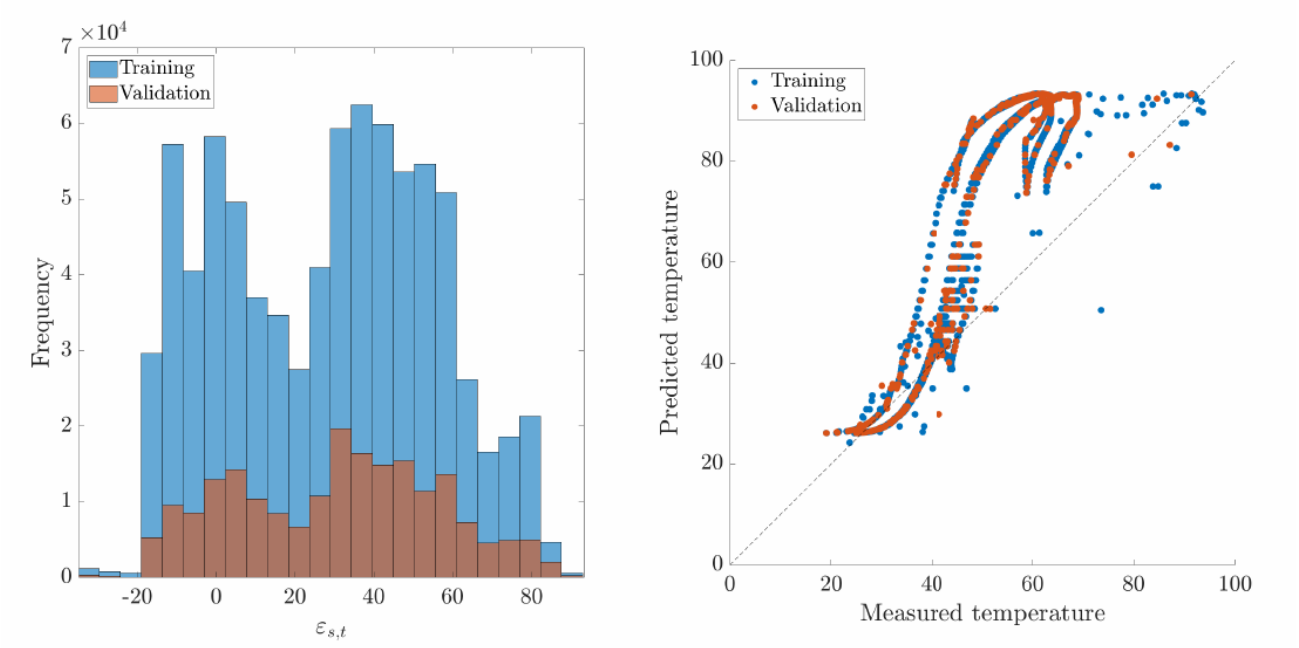
The apparent discrepancy in error rates among specific sensors, as depicted in Figure 2.11b, may initially seem higher than expected. However, there is a deliberate rationale behind this observed variation. Figure 2.12a displays the positions of three randomly selected sensors, while Figure 2.12b portrays the evolution of the measured temperatures on these sensors. Notably, a dashed vertical line in the latter figure underscores the time step corresponding to 250 days. After 250 days, the measurements appear to be influenced by an event, characterised by a significant temperature increase. We establish that this deviation from nominal behaviour is not reflective of the intended model. It is essential to emphasise that the choice of the initial 250 timesteps for training is not arbitrary but stems from a thoughtful decision. This decision is driven by a significant event within the ALC1605 experiment, marked by a pronounced impact on subsequent time step measurements. Despite this deviation, the intriguing aspect lies in the potential wealth of valuable insights that can be derived from this behaviour. The proposed methodology serves not only as a baseline model representing the expected performance of the tunnel and its surroundings but also possesses the capability to raise alarms when a sensor exhibits anomalous behaviour due to drifting measurements. This nominal model, therefore, not only functions as a benchmark but also provides a means to detect deviations from expected sensor performance.

Nevertheless, we have the opportunity to assess the outcomes generated by the hybrid twin model in comparison to those exclusively reliant on physics. This comparison aims to determine whether a notable enhancement has been achieved. To facilitate this evaluation, we present the error graph for the physics-based solution in Figure 2.13. The visual representation clearly demonstrates a significant improvement in overall performance when incorporating additional data-driven insights. Notably, this emphasises the concept that the precision of the physics-based solution does not necessarily need to be exceptionally high. Instead, it can be substantially elevated through the integration of data. This inherent capability has the potential to considerably streamline the modelling process, meshing activities, and calculations, ultimately reducing the time investment required.



(a) (b)

Figure 2.12 – (a) position of the three sensors and (b) the evolution of measured temperature on these sensors.



(a) Frequency of error measured as indicated in (9).

(b) Comparison between the measured temperature and the forecasted temperature for the last time step.

Figure 2.13 – Various methods for depicting the errors made by the physics-based model.

2.5 Application

The preceding section has demonstrated the potential of the proposed hybrid twin approach in forecasting temperature fields, showcasing its applicability across various industrial scenarios. Of particular significance are its applications in sensor diagnosis, as previously mentioned, and its capacity to extrapolate temperature data across all spatial coordinates within the domain, extending beyond sensor placement locations.

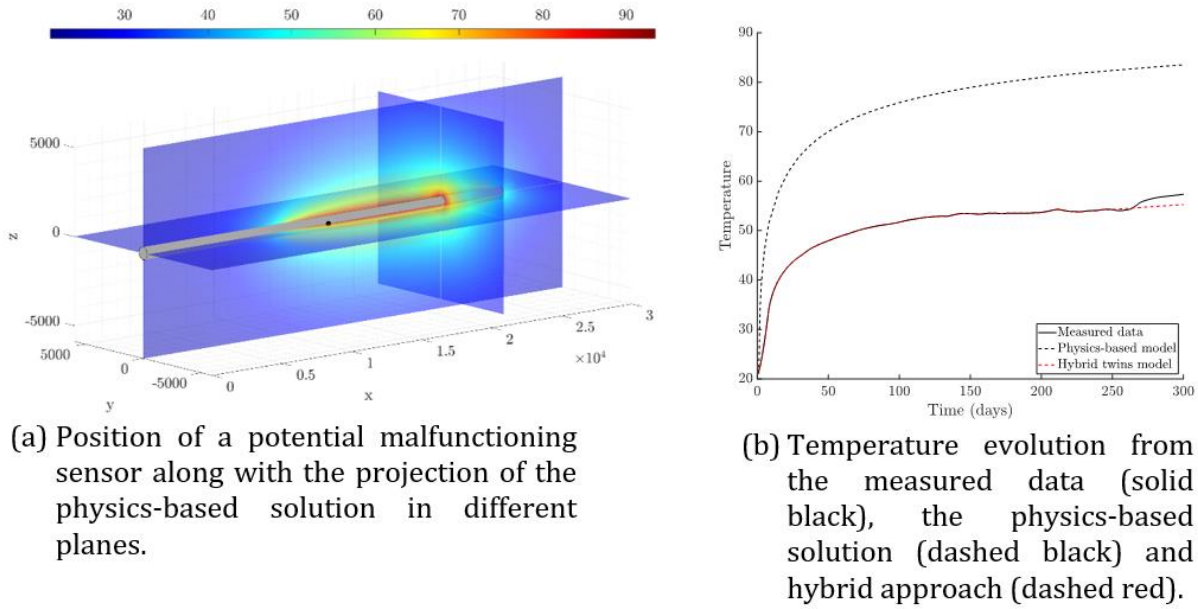
2.5.1 Sensor Diagnosis

A notable industrial application of our system lies in its ability to diagnose and monitor the performance of sensors. To showcase this feature, we'll revisit the previously mentioned database, where particular sensors experienced disruptions due to an incident within a laboratory setting. This incident involved the sudden closure of a security gate, causing a rapid rise in temperature within the controlled environment. It's crucial to note that the triggering event can vary widely, from gate closures as exemplified in our data set to sensor malfunctions or any unforeseen anomaly.

Our strategy for malfunctioning sensor diagnosis comprises two stages. First, within our approach that harnesses hybrid twins to model the anticipated behavior of the sensor network, any deviation from this expected norm signifies an early warning sign of potential malfunction. Figure 2.12a visually represents a sensor suspected of malfunction, displaying the sensor's spatial coordinates within the domain. Additionally, this figure incorporates three planes representing the projected physics-based solution. Figure 2.14b demonstrates three temperature evolution curves: one derived from measured data, one based on a physics-based model, and one based on a hybrid twin model. In the measured data curve, a sudden temperature increase suggests an incident that compromised the sensor's performance. By comparing this curve to the hybrid twin model and establishing a threshold, we can identify when the measured data exceeds it, potentially indicating an issue with the sensor.

Second, to effectively diagnose sensor malfunctions, we leverage the inherent redundancy in our data. This redundancy arises from the fact that numerous sensors are positioned closely together. When a sensor exhibits signs of dysfunction, this redundancy enables a comprehensive evaluation. By directly comparing the behavior of the suspect sensor to its neighboring counterparts, we can uncover any anomalies or deviations from expected patterns. This side-by-side analysis proves to be an invaluable tool in pinpointing potential issues with the problematic sensor. Figure 2.15 indicates the positions of the potentially malfunctioning sensor (in red) along with the closest five neighbors, along with the complete sensor database. Conversely, Figure 14 represents the curves of the measured data over time. The dissimilar behavior of the target sensor compared to its corresponding neighbors could indicate an actual sensor malfunction.

Once we can identify the malfunctioning sensor, it becomes the responsibility of the operator to thoroughly evaluate the situation, as the measured data or hybrid twin model cannot definitively determine the cause of the problem.



(a) Position of a potential malfunctioning sensor along with the projection of the physics-based solution in different planes.

(b) Temperature evolution from the measured data (solid black), the physics-based solution (dashed black) and hybrid approach (dashed red).

Figure 2.14 – Evaluation of different parts of the model for a random sample sensor that malfunctions within the test data partition.

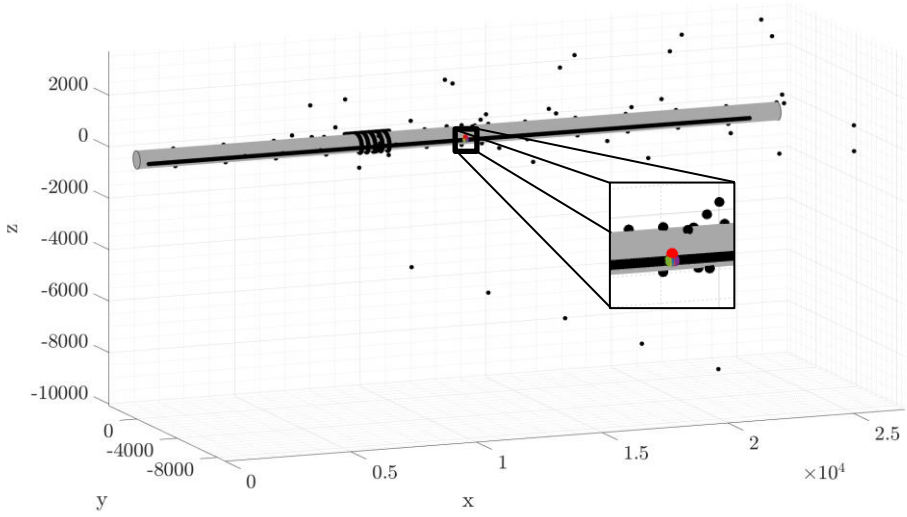


Figure 2.15 – Spatial coordinates for the sensor data set are presented, with the potentially malfunctioning sensor visually emphasized in red. Additionally, the surrounding context is provided by colouring the five closest sensors.

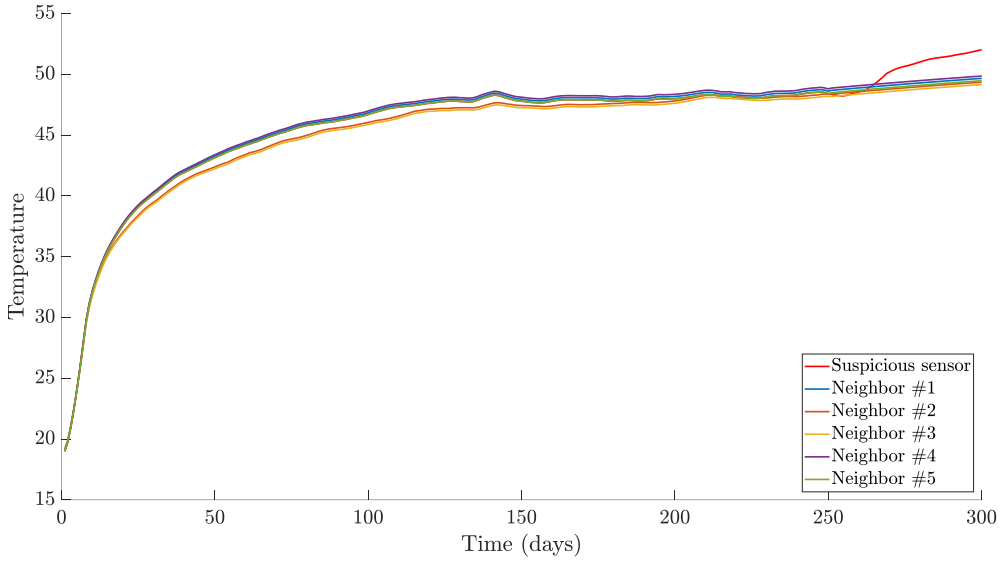


Figure 2.16 – The temperature progression, derived from the collected data, is illustrated for the sensor under scrutiny as well as its five closest neighbours.

2.5.2 Domain Extrapolation

In this section, we wish to emphasise another significant industrial application: the potential to expand the improved solution across the entire domain, rather than limiting it to sensors alone. This expansion will enable us to ascertain the accurate temperature at every point within the rock, not just at the sensor locations.

To fulfil the proposed task, our strategy relies on a conventional Proper Orthogonal Decomposition (POD) approach [44, 45]. Specifically, we work with the temperature matrix derived from the physics-based solution, where rows represent nodes, and columns represent time steps. We decompose this matrix as shown in the following equation:

$$\mathbf{A} = \mathbf{V}_A \cdot \mathbf{\Sigma}_A \cdot \mathbf{W}_A^T = \mathbf{V}_A \cdot \boldsymbol{\alpha}_A. \qquad \text{(Equation 2.10)}$$

Based on the preceding expression, we have the option to select a limited number of POD modes in order to reconstruct an approximate version of the original matrix, as follows:

$$\mathbf{A} \approx \mathbf{V}_A^r \cdot \boldsymbol{\alpha}_A^r. \qquad \text{(Equation 2.11)}$$

Furthermore, the inherent uncertainty arising from the disparity between numerical simulations and real measured data, referred to as "ignorance," can be initially represented as a matrix ($\widehat{\mathbf{B}}$). This matrix can then be decomposed as well. However, it is important to note that in this equation, the matrix exclusively contains the sensor data since this is the information that is available and known.

$$\widehat{\mathbf{B}} = \mathbf{V}_{\hat{B}} \cdot \mathbf{\Sigma}_{\hat{B}} \cdot \mathbf{W}_{\hat{B}}^T = \mathbf{V}_{\hat{B}} \cdot \boldsymbol{\alpha}_{\hat{B}}. \qquad \text{(Equation 2.12)}$$

Similarly, we can selectively retain specific POD modes in order to reconstruct an approximation of the original matrix, as follows:

$$\widehat{\mathbf{B}} \approx \mathbf{V}_{\hat{B}}^r \cdot \boldsymbol{\alpha}_{\hat{B}}^r. \qquad \text{(Equation 2.13)}$$

The vector set $\mathbf{V}_{\hat{B}}^r$ is defined within each sensor, while $\mathbf{V}_A^r$ is defined across the nodes of the FEM model. Since the sensors are confined within the model boundaries, we can derive $\mathbf{V}_{\hat{B}}^r$ from $\mathbf{V}_A^r$ by utilizing the characteristic shape functions of the mesh. This newly obtained set of vectors can be denoted as $\mathbf{V}_{\hat{A}}^r$ as it specifically relates to sensors. Subsequently, we can integrate this matrix into Equation 2.13, where the unknown variables are represented by the coefficients $\boldsymbol{\alpha}_{\hat{B}}^r$.

$$\widehat{\mathbf{B}} \approx \mathbf{V}_{\hat{A}}^{r} \cdot \boldsymbol{\alpha}_{\hat{B}}^{r} \rightarrow \boldsymbol{\alpha}_{\hat{B}}^{r} = \left(\mathbf{V}_{\hat{A}}^{r^T} \cdot \mathbf{V}_{\hat{A}}^{r}\right)^{-1} \cdot \left(\mathbf{V}_{\hat{A}}^{r^T} \cdot \widehat{\mathbf{B}}\right). \quad \text{(Equation 3.14)}$$

The final step involves using the recently acquired coefficients to extrapolate our understanding across the entire domain through the use of the matrix $\mathbf{U}_{A}^{r}$, as follows:

$$\mathbf{B} \approx \mathbf{V}_{A}^{r} \cdot \boldsymbol{\alpha}_{\hat{B}}^{r}. \quad \text{(Equation 3.15)}$$

To recover the improved model behaviour, we can use a matrix-based form of Equation 3.1. This equation involves the summation of matrices $\mathbf{A}$ and $\mathbf{B}$, ultimately yielding the actual behavior represented by $\mathbf{Y}$. This extrapolation method yields temperature fields, as illustrated in Figure 2.17.

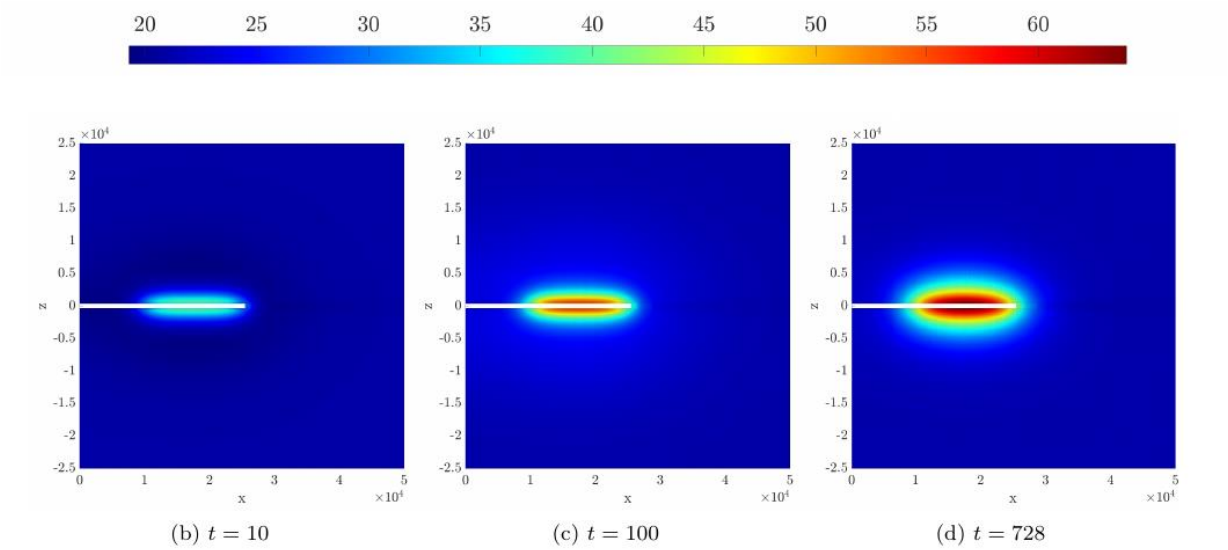


Figure 2.17 – Progression of the temperature distribution over time when the enhanced solution in the sensors is extrapolated to the whole domain.

2.6 Conclusions

In our research, we introduce an innovative method for forecasting the temperature evolution of a HLW disposal cell. Our approach involves creating a hybrid twin model that combines physics-based solutions with actual measurement data. This method offers several advantages, particularly in long-term forecasting, as it grounds predictions in the well-established physics of our application, enhancing the reliability of our results.

Our hybrid twin model is built upon data from the ALC1605 experiment, a demonstrator for the real repository. Through this methodology, we aim to gain insights that will enable its application to the actual repository. Upon analysing the results, it appears plausible that this approach will perform effectively in the final application.

Furthermore, our study of the demonstrator's hybrid twin model yields valuable information about the sensors within the actual repository. This includes determining the optimal number of sensors to be deployed and their optimal placement. We have also demonstrated that the physics-based model need not be overly precise, as the modelling of uncertainty can bridge the gap between numerical simulations and real-world behaviour.

In practical application, our model running in the real repository enables the identification of sensor drift, which can promptly alert operators to take necessary actions. Additionally, it allows for monitoring temperatures in locations far from the sensor installations.

3. Full-Scale Emplacement (FE) Experiment

This section provides a summary of the three test cases undertaken using information and data from the FE experiment. The experiment is described in Section 3.1. The test cases are then described in the three subsequent sections:

- Section 3.2 describes the test case undertaken by Nagra, which focused on data management and access, data quality control, and development of data cleansing tools.
- Section 3.3 describes the test case undertaken by PSI, which focused on development of methods for modelling of the thermal evolution of the experiment.
- Section 3.4 describes the test case undertaken by UFZ, which focused on the development of digital twins of the experiment and the wider URL for visualisation purposes.

3.1 Description of the Reference Experiment

Authors: Edgar Manukyan and Martin Schoenball (Nagra), Wilfried Pfingsten and Guang Hu (PSI)
Editor: Matt White (Galson Sciences)

The FE experiment simulates, as realistically as possible, the construction, waste emplacement, backfilling and early-stage THM post-closure evolution of a spent fuel and HLW disposal tunnel according to the current Swiss repository concept [46, 47, 48].

The objectives of the FE Experiment are:

- To investigate spent fuel and HLW repository-induced THM coupled effects on the host rock at full scale and to validate existing coupled THM models.
- To verify the technical feasibility of constructing an emplacement tunnel using standard industrial equipment.
- To optimise the bentonite buffer material design and production, in particular to produce bentonite blocks that are capable of resisting the ambient conditions during the emplacement and operation phases.
- To investigate (horizontal) canister and buffer emplacement procedures for underground conditions.

The FE Experiment was designed to replicate the emplacement tunnel of the reference repository design at 1:1 scale. The FE tunnel has an inside diameter of approximately 3 m and a length of 50 m, and is divided into four zones (Figure 3.1). In the main Test Section of the experiment, three heaters with dimensions similar to those of spent fuel and HLW disposal canisters were emplaced on top of bentonite block pedestals. The remaining space was backfilled with highly-compacted granulated bentonite material (GBM). In this zone, the rock is supported by mesh-reinforced low-pH shotcrete. At the far end of FE tunnel is the Interjacent Sealing Section, which comprises a concrete wall, a 2-m-long bentonite block wall consisting of manually installed bentonite blocks and a section filled with GBM. In this zone, the rock is supported by steel sets. A concrete plug section and an access section comprise the two sections closest to the FE cavern, which provides the entrance to the FE tunnel and hosts the monitoring data acquisition systems during experiment operation. The experiment began heating in December 2014 and has been in long-term operation since.

The experiment is densely instrumented with more than 1,000 sensors in both the EBS and surrounding rock mass to monitor the coupled processes during heating phases (Figure 3.2). Most point sensors are temperature and pressure sensors. Fibre-optic cables are deployed to monitor temperature using distributed temperature sensing (DTS). Other parameters monitored include:

- Saturation state, which is monitored by measuring relative humidity, suction, gravimetric or volumetric water content.
- Deformation, which is monitored by measuring convergence (prior to EBS emplacement), displacement or inclination (in the geosphere).
- The concentration of gas phases and pH.

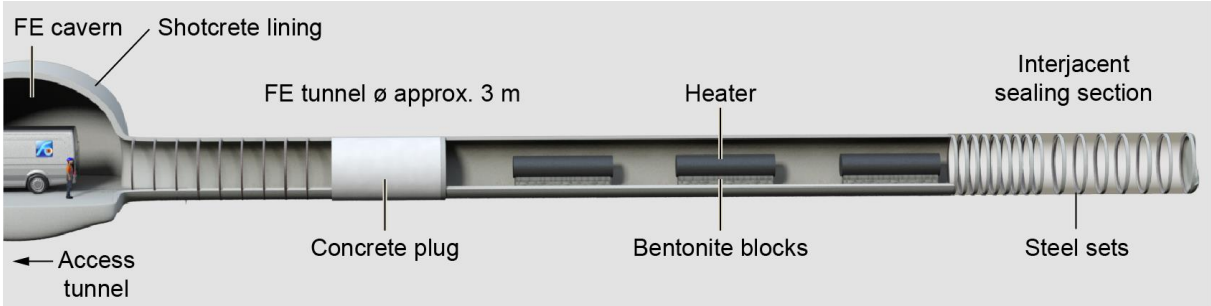


Figure 3.1 - Overview of the FE experimental setup Mont Terri URL in Switzerland [47]. From left to right is the FE cavern, where all data acquisition and interrogators are installed, the concrete plug that closes of the heated section, the three heaters as well as an interjacent sealing section with steel supports.

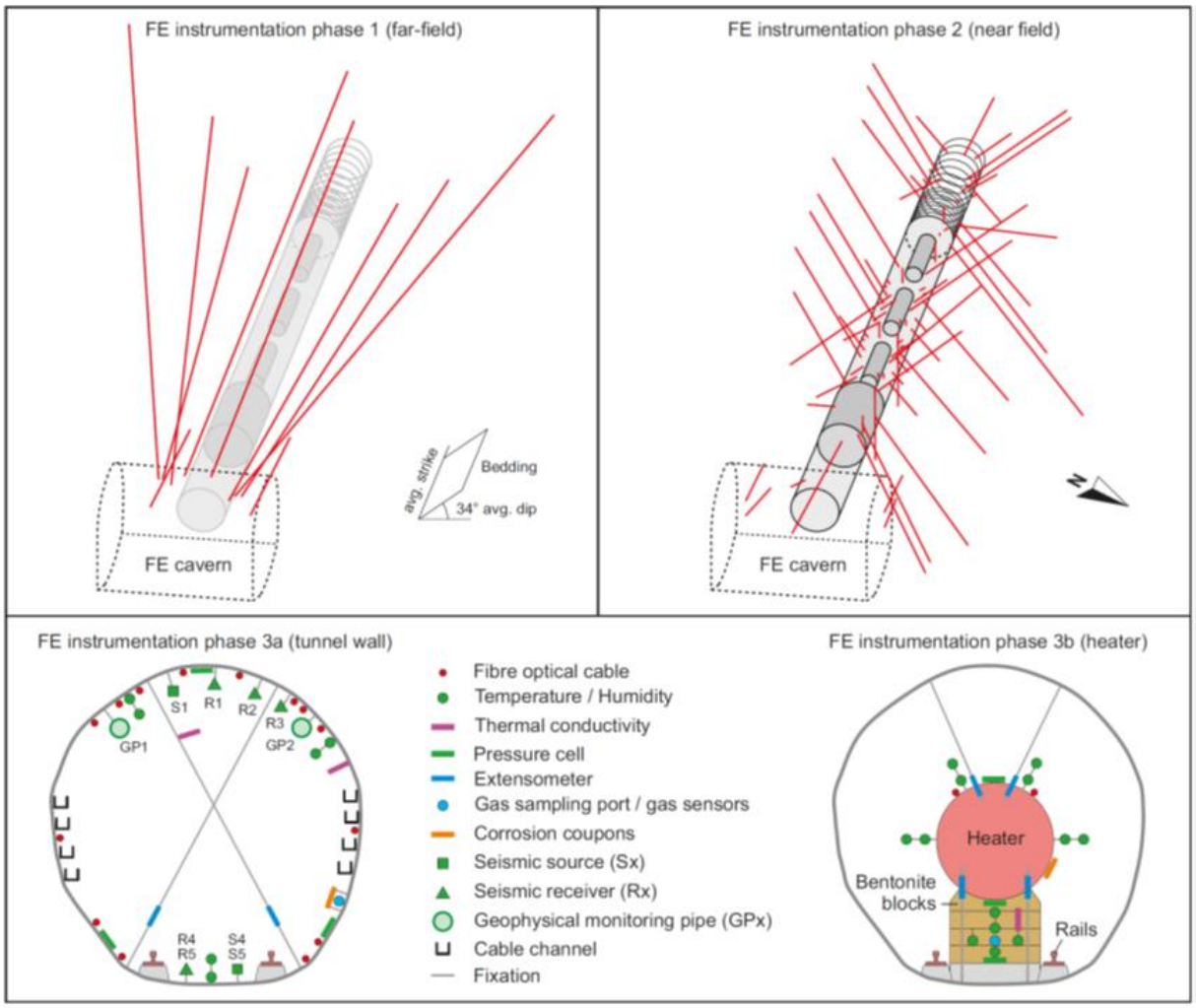


Figure 3.2 - The main instrumentation phases of the FE experiment [47]. The figures at the top illustrate the layout of boreholes around the 50-m-long FE tunnel. The figures at the bottom show simplified representations of the instrumentation cross-sections within the 2.5–2.7-m diameter FE tunnel.

3.2 Test Case Description and Results: Nagra

Authors: Edgar Manukyan, Martin Schoenball and Christoph Wetter (Nagra)
Editor: Matt White (Galson Sciences)

3.2.1 Introduction and Objectives

The FE experiment provides a comprehensive data set with lessons learnt from almost a decade of long-term monitoring. This section discusses some selected aspects of operating a complex monitoring system over such a time frame. The examples and approach developed and discussed below range from calibration of sensors, inferring drift in (fibre-optic) sensors owing, to system deterioration and identification of other types of sensor failure.

3.2.2 Data Management and Data Access

The monitoring data of the FE experiment is stored in an online database, the FE Information System (FEIS), that provides easy access for Nagra and its external partners. The FEIS tool is browser-based and provides relevant documents and data sets, maintenance operations as well as basic visualisations of the online data. Some preprocessing of the data is performed within the data acquisition system and the FEIS database. This includes conversion of the measured quantities into the target parameter through application of calibration relations. For the fibre-optic data this also includes a live calibration of each measurement against reference data that is acquired simultaneously. However, at the time of original implementation of the FEIS, no automated access to the actual data was foreseen, and the browser interface has limited capabilities for in-depth analysis of the monitoring data.

Therefore, to provide convenient access to the FE data to MODATS partners, it was decided to implement an application programming interface (API) to access the data. The system works as a wrapper that improves security by using a custom data port for MODATS users and a separate data port for the Nagra FEIS database connection. The data is accessed through the API and structured query language (SQL) queries, which provides access to only the desired data sets, and exposes only the functions and read-only table views required to retrieve the FE data. The FE database tables and data cannot be deleted or modified. Libraries to access the database exist for several development platforms, including Python, Matlab and C#. Separating the FEIS and the MODATS databases also helps to limit performance issues on the MODATS database from affecting the FEIS (for example, a MODATS user's query that runs for hours).

3.2.3 Calibration and Quality Control of Fibre-Optic Monitoring Data

The DTS monitoring of the FE experiment has been ongoing since December 2014. In this sub-section, some key lessons learnt from operating such a system under conditions relevant to repository operations in the long-term are identified.

Eight fibre optic (FO) cables with a total length of 2,160 m were installed on the tunnel wall as well as in several boreholes. Three interrogators were used during the first eight years of the experiment. In 2014, a Smartec DiTemp interrogator unit as well as a Neubrex NBX7020 unit were installed. In 2019, a Silixa Ultima S interrogator unit was added. The Smartec and Silixa units employ Raman scattering to measure temperature. The Neubrex unit uses Brillouin scattering, jointly measuring temperature and strain. Observations in the first eight years showed that Brillouin scattering is less practical for temperature monitoring under repository conditions than Raman scattering, since temperature measurements can be affected by strain built up inside the FO cables [49]. Here we focus on the data acquired by the Smartec and Silixa units.

It is known that DTS measurements are affected by external factors. Examples are the operating temperature and conditions of the interrogator unit, physical conditions, cleanliness of the FO cable connections, and the presence of strain or sharp bends at any location along the fibre [50]. These factors influence the Stokes (S) and Anti-Stokes (AS) raw values measured by the interrogator, as well as calibration parameters used for converting S and AS values to temperature [51].

Between December 2014 and April 2018, temperatures were calculated using static calibration parameters, determined at the start of the experiment. DTS accuracy, as compared with PT1000 point sensors, in that time span was around 1.5 °C (Figure 3.3, top) with all of the FO data higher than the PT1000 data.

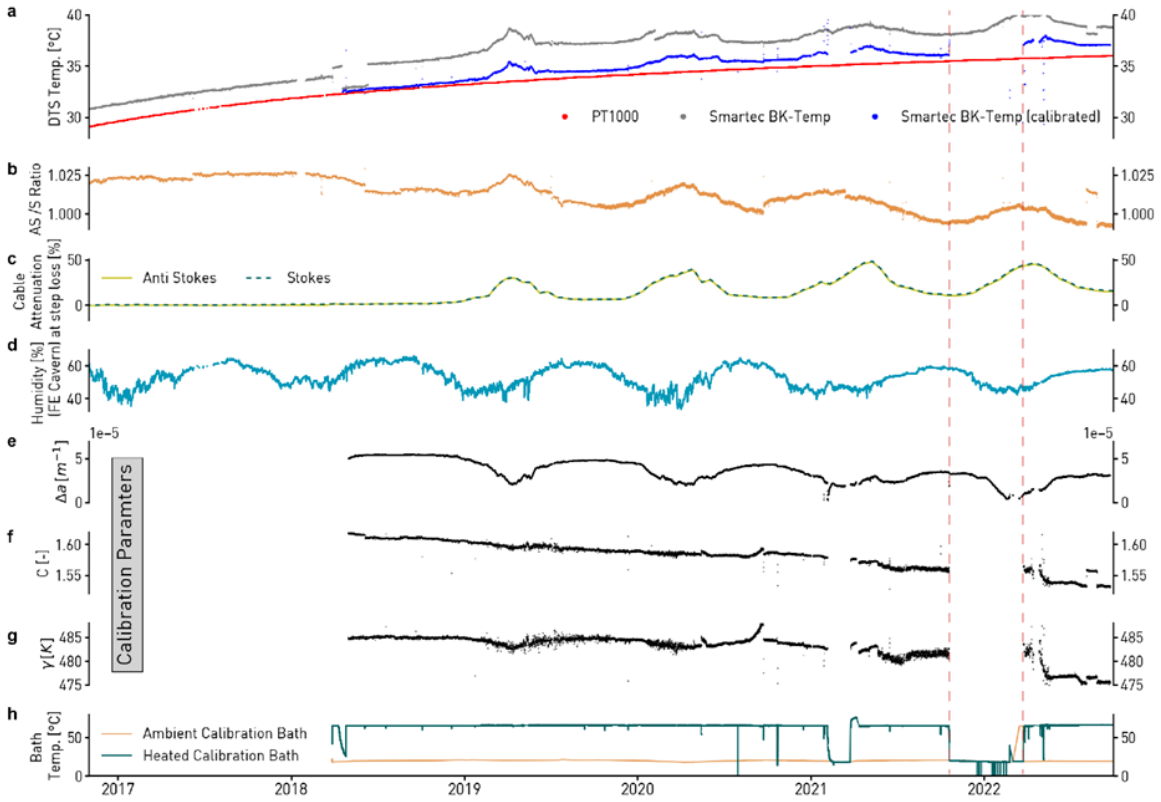


Figure 3.3 – Panel a: Temperature evolution from DTS as well as point sensors. Seasonal trends are noticeable. Panel b-d: Yearly trend and losses on DTS raw data, and a similar trend in humidity. Panels e-h: Evolution of calibration parameters $\Delta a, C$ *and* γ *(see below) as well as calibration bath temperature.Red dashed line indicates period when the calibration process failed owing to calibration bath heater failure, this further affects the calibration parameters. Also, the yearly trend is visible in the AS/S (anti-Stokes/Stokes) ratio (panel b) as well as in the* Δa *parameter (panel e).*

In April 2018, after 3 years of continuous operation, two 220-litre calibration water baths, providing known hot and cold reference temperatures, were installed inside the FE Cavern. Calibration was performed following the procedure in [51] (Figure 3.4). This process improved the DTS accuracy from 1.5 °C to 0.25 °C as compared with PT1000 point sensors (Figure 3.3, top). This calibration process has remained in use since April 2018. During this period, calibration was not working properly for 211 days or 12% of the run-time, owing to failures of the calibration bath heater. This is visible in the γ and C parameter in Figure 3.3, which scatter in periods when the heater was not working properly. During these periods, the determined calibration parameters were wrong and ultimately, erroneous temperature measurements were reported.

Heating tests using heatable fibres are performed yearly to ensure proper functioning of all systems. Optical time domain reflectometry (OTDR) measurements of cable attenuation are also carried out yearly, e.g., to detect changes in FO cable performance. Overall trends show increasing attenuation (i.e., step losses) at splice positions over time, which could be linked to aging effects. Additionally, yearly maintenance work is performed on the Smartec DiTemp interrogator unit including fan-filter replacement and dust removal. During the first eight years of operation, all three interrogators worked without failure.

The display of the calibration parameters along with the derived temperature measurements and comparison with reference measurements of point sensors (Figure 3.3) shows that calibration parameters as well as interrogator raw data (Stokes, Anti Stokes) can be used to gain information about the state of health of the DTS system. Calibration parameters are based on physical properties, for example, cable attenuation Δa or interrogator sensitivity C [51]. In theory, one should expect the three calibration parameters to be constant over time for stable conditions. The FE experiment showed that this is not the case, and small but significant temporal changes in the calibration parameters were observed in the first four years following installation of the calibration water baths. Calibration parameters are sensitive to DTS internal factors, for example, a change of interrogator sensitivity as well as external factors such as seasonal trends caused by atmospheric conditions. We propose that in case a calibration process, outlined by [51], is used, the obtained calibration parameters can also be used to monitor the state of the DTS system. Stable calibration parameters are a good performance indicator for DTS measurements. Changes in the automatically determined calibration parameters correlate well with changes made to the system (e.g., maintenance) but also apparent deterioration of cable splices.

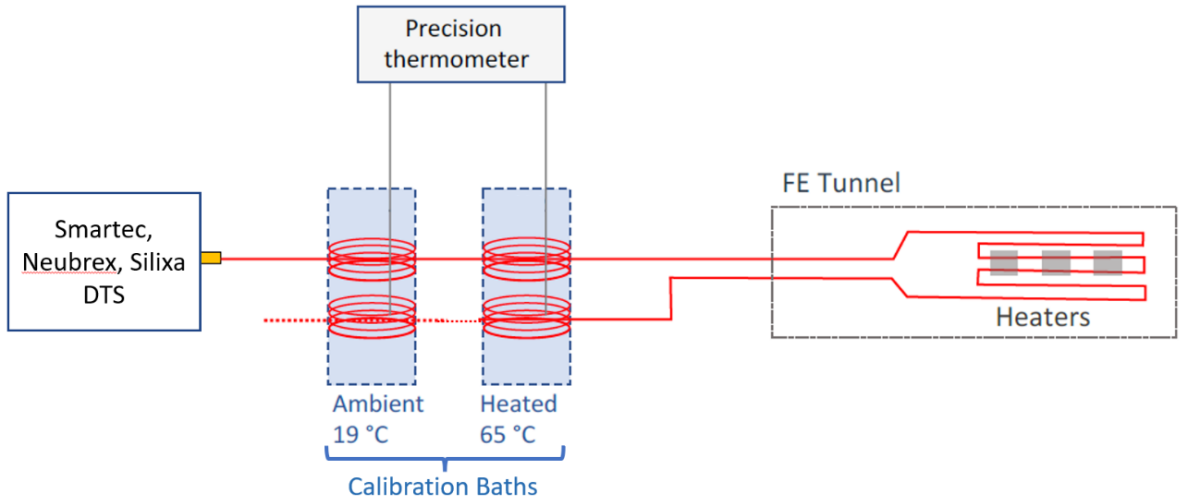


Figure 3.4 - DTS experimental design. Duplexed single-ended configuration, in which FO cable gets routed twice through two reference baths for which the temperature is known. Those reference temperatures can be used to calibrate the raw DTS temperatures. In the FE experiment, Reference Section 1 is designed as an ambient bath (~19 °C) and Reference Section 2 as a heat bath (65 °C) [50].

Seasonal fluctuations in the raw data, the calibrated temperatures, and one of the calibration parameters (Δa) were observed for individual fibre optic cables. These artefacts could be linked to the locations of splices in the FO cables, causing step losses. While the calibration method used can correct for continuous signal losses along the fibre (which may indicate cable aging), it cannot correct for step losses. The observed seasonal trends suggest that atmospheric conditions impact the fibre splices. DTS temperature anomalies (compared to point sensor references) reached amplitudes of up to 2 °C. Therefore, we suggest that those splice positions are thoroughly monitored using OTDR or comparable methods.

3.2.4 Approaches for Data Cleaning

3.2.4.1 Need for Data Cleaning

In the FE experiment, there are more than 1,000 sensors providing monitoring data. Despite efforts to acquire high-quality data, we observe malfunctioning of the monitoring system over time. Examples of malfunctioning behaviour include no data, data gaps, spikes, unexpected constant values (freezing of values), increased noise levels, unrealistic data trends, and unreasonable data values. These malfunctions are checked on-site. For some of them, the cause can be fixed by manual intervention

(e.g., cable connections to the data acquisition system have become loose). Some malfunctions recover over time on their own without providing a hint on the cause, and there are malfunctions that result in permanently false measurement values.

Erroneous data have only limited value and are a nuisance when it comes to modelling and visualisation. Hence the data need to be cleaned and bad data removed. Nagra's approach is to not manipulate the recorded data. This means that no data is deleted, modified, or added to the database. Instead, the data cleaning algorithm puts flags to each single data point identifying its status as either valid or discarded (because of a specific criterion). This approach makes sure that data is not modified, introducing subjectivity (e.g., interpolation or extrapolation), is fully auditable, and still provides the possibility to view the entire data set, including differently flagged data. It also means that data can be readily selected for different uses. In the discussion below, cleaning and flagging are used somewhat interchangeably.

3.2.4.2 Initial Data Cleaning

Considering the large amount of data produced in the FE experiment and the large amount of data expected to be acquired in future monitoring systems, automated and robust data flagging algorithms are needed. Before starting more advanced checks, clearly erroneous data are identified and flagged. These are spikes and implausible data values. In the plausibility check we determined hard limits on the measured parameters beyond which the recorded values cannot be true. For instance, in the FE experiment, it is not possible to have negative temperatures. Spikes are identified by checking if there is a sharp change in measured temperature (the difference in value is at least 1 °C and the rate of change is at least 1 °C/min) and the subsequent value is again a sharp change but with opposite sign.

Within MODATS, two different concepts for advanced data validity checking were implemented. These two concepts are described below.

3.2.4.3 Advanced Data Cleaning – Sensor Grouping

In the first approach, advantage is taken of the presence of many sensors and close-by sensors are analysed jointly (in this instance nearby refers to sensors measuring the same parameter in the same component of the multi-barrier system). The idea behind this is that, within a single group, sensors should show similar data trends. The algorithm checks all data points starting from the installation time of the sensors. It collects relative temperature changes for all sensors. If a relative data change from a sensor is larger than the sensor accuracy and is calculated to be an outlier (the value is more than three scaled median absolute deviations from the median, with the scaling factor of 1.4826, which connects median absolute deviation to standard deviation in case of normal distribution) compared to relative data changes of its group members, then the data point is flagged as an outlier.

Figure 3.5 shows the effect of data cleaning before and after the removal or invalid data points of 40 temperature sensors within the BFEA010 borehole. The data cleaning algorithm, performing plausibility check, spike removal and outlier identification, can significantly reduce the erroneous data points. The drawback of this algorithm is the need for several sensors that are either located close by or are expected to show similar behaviour (e.g., sensors in a borehole away from heaters). This requirement for sensor redundancy could be prohibitive for the actual repository. Furthermore, the number of functioning sensors will inevitably decrease over time owing to sensor failure. In the case of the FE experiment, we also noticed that we could not effectively group together all sensors within a specific volume. The two issues we identified are:

- Different sensor types having different accuracy and resolution.
- Different sensor groups having different recording time stamps and intervals.

It is important, therefore, to consider the compatibility of data from different sensors when designing monitoring systems.

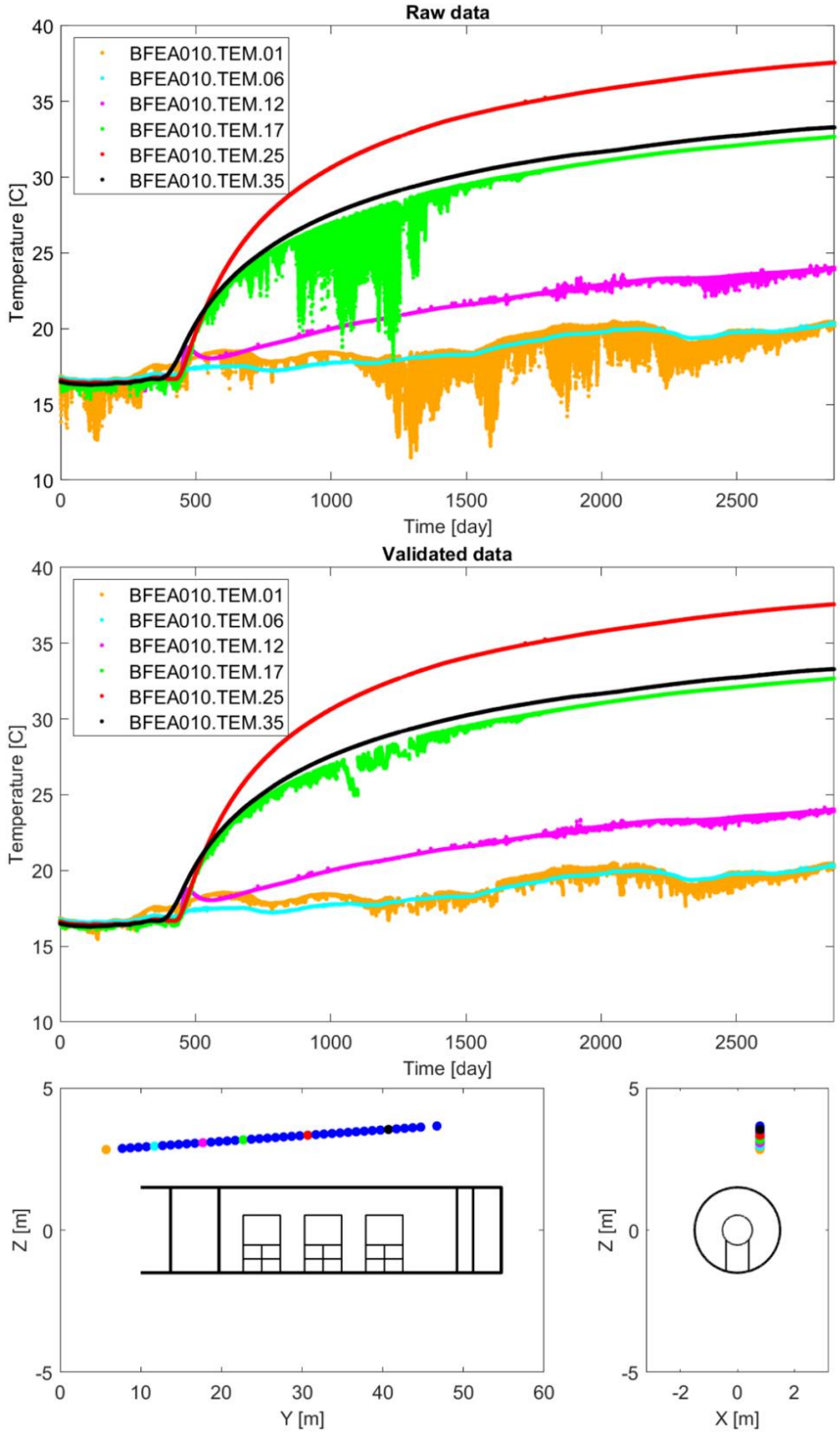


Figure 3.5 - Temperatures of selected sensors in BFEA010 borehole before (top panel) and after (middle panel) removal of outliers. Sensors were selected to illustrate a range of sensor behaviours including flawless (red & black), containing atmospheric influences (cyan) as well as various anomalous sensor behaviour (green, pink) or combinations thereof (orange). Each coloured dot in the top two panels corresponds to a data point measured by a single sensor. Locations of temperature sensors are indicated with respectively coloured dots in the bottom panels. The blue dots indicate the locations of sensors for which data are not plotted but were used in the analysis.

3.2.4.4 Advanced Data Cleaning – Numerical Modelling

In the second approach, data from each sensor is checked separately using the results of numerical modelling. Ideally numerical modelling would show the same data as each sensor and any mismatch would be due to sensor misbehaviour. However, there are several reasons why the modelled data is different to the true values. Possible reasons include:

- The model is only an approximation of a more complex reality.
- Modelled data is temporally and spatially interpolated to fit sensor time and location.

In addition to all kinds of erroneous data, the recorded signal also has peculiarities that are real but are often not considered in the numerical models. Such events include:

- Various QC tests in the beginning of the experiment.
- Electricity breakdown in the laboratory leads to heaters switching off and reduction of temperatures.

Hence, the flagging algorithm that uses results of numerical modelling needs to be robust to correctly flag the data. Like in the previous approach, the checking is done data-point by data-point starting from the first recording. Changes in the value between a new data point and the last valid data point are compared with the change expected from numerical modelling. As the modelling data is not expected to match reality exactly, lower and upper bounds for the next point measurement are extracted from modelling. Furthermore, room for inaccuracy is allowed owing to the accuracy of sensors. A data point is considered valid if the change in recorded data is within the bounds extracted from modelling. Further, a small gradient (0.1 °C/day) was implemented on both upper and lower bounds in order to recover from falsely approved data that drifted away from true temperatures. For the cases where data points were identified as possibly invalid, the algorithm first checks if the previous valid data point was frozen (showing exactly the same value for some period). A sketch of the concept is shown in Figure 3.6. For the cases where frozen data were identified, the data were marked as such, and validity was checked again using the last valid data point. Figure 3.7 shows an example of data flagging results for two sensors. The algorithm performs well in identifying and flagging various types of noise. Nevertheless, there are cases where good data points were identified as invalid and cases where poor data were identified to be good, both of which were identified via expert review of the data. Although such misjudgements cannot be entirely eliminated, we believe that the algorithm can be further improved.

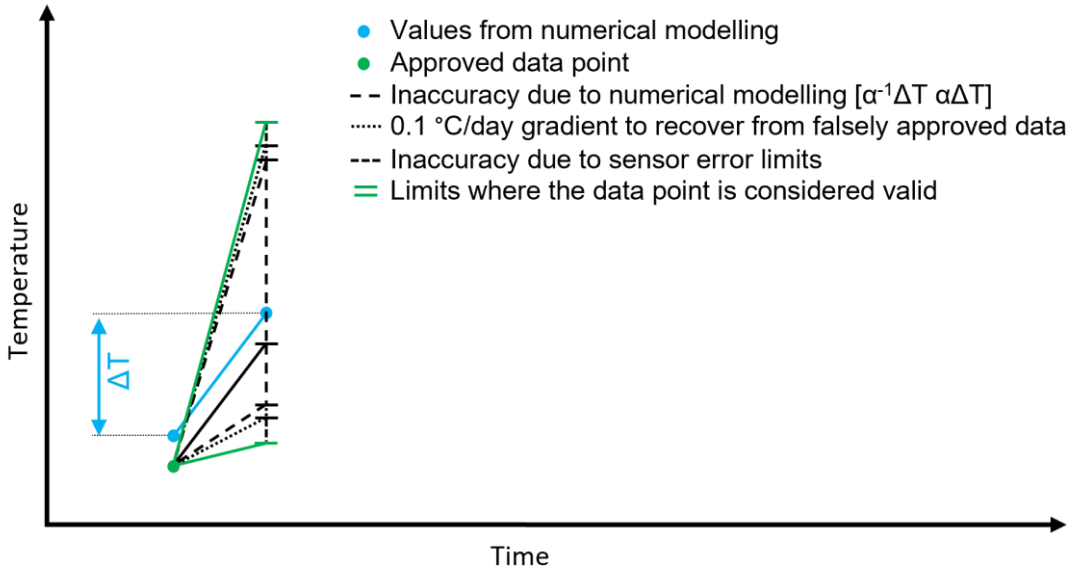


Figure 3.6 - A sketch of the concept showing how the limits are calculated for checking the validity of the next data point.

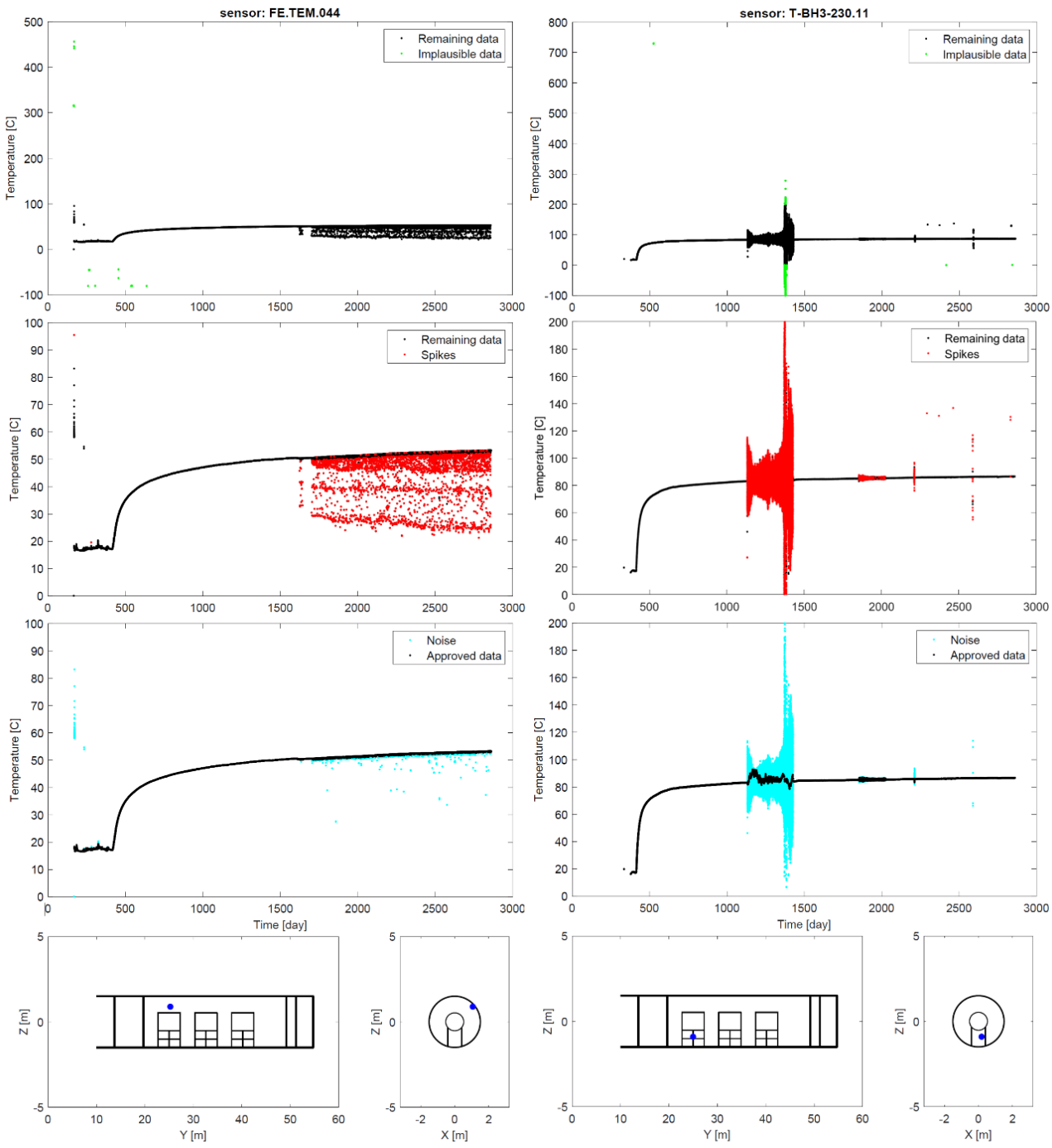


Figure 3.7 - Two columns show results of data flagging for two sensors. The results of plausibility check, spike check and noise identification are shown on the first, second and third rows, respectively. First row shows all recorded data, while the second and third rows show only recorded data that has remained after data cleaning step indicated in the plot above. The bottom row shows the location of respective sensors. Note changes of scale between some plots and that because of the plotting order some of the data points could be hidden behind data points with different colour.

3.2.5 Conclusions

We have made several improvements for validating and qualifying monitoring data of the FE experiment. On the practical side, we implemented access to FE database through Python and Matlab scripts, which maintains the necessary security over the full database. Such direct (easy) access is essential for post-processing data using algorithms developed externally of the FEIS.

On the monitoring side we had a close look at both DTS and point sensor temperature data. Our research showed that DTS calibration parameters together with interrogator raw data can be indicators for the state of health of a DTS system. In the FE experiment we identified that deteriorating cable splices can cause temperature anomalies resulting in seasonal trends. This finding suggests that in the monitoring system the number of splices should be minimal. Furthermore, despite the presence of calibration baths, nearby measurements of temperatures near DTS provide a means for quality control of DTS signals in the future.

For point sensors we have developed two algorithms for cleaning temperature data. Both approaches follow the principle of not adding or deleting any measured data point from the database, but rather flagging them. One approach makes use of nearby sensors to identify anomalous temperature values and the other approach uses data from numerical modelling. Both approaches performed well in flagging the untrustworthy data as invalid and trustworthy data as valid. Nevertheless, it was not possible to obtain 100 % flagging accuracy. Furthermore, the first approach requires presence of several sensors in close vicinity, which is possible for monitoring URL experiments but could be a limitation for a real repository where monitoring may be sparse. While the results of this work are encouraging, we believe that it is worthwhile to test the algorithms on other experiments and other types of data to further improve them before implementation in a real repository. For now, expert opinion remains an important factor for evaluating monitoring data.

3.3 Test Case Description and Results: PSI

Authors: Guang Hu and Wilfried Pfingsten (PSI)
Editor: Matt White (Galson Sciences)

3.3.1 Introduction and Objectives

Within MODATS, our focus is on the assessment of sensor data, especially temperature and humidity sensor data and the related modelling of the temperature evolution in the FE experiment tunnel near-field after switching on heater power using all kind of geometrical and material properties information available from the experimental setup and the emplacement description. PBM and DDMs have been developed and compared to FE temperature and humidity sensor data. Efficiency and robustness of different DDMs have tested and the best performing are compared with PBM and possible applications for a temperature digital twin are given, e.g., sensor failure identification or identification of dominating parameter uncertainties.

3.3.2 Data Treatment and Management

Temperature and relative humidity sensor data delivered by Nagra were used in this study. The data were extracted from the FEIS database and were delivered and analysed in two steps.

The initial sensor data covering the period from 15 December 2014 (start of heating) to 30 June 2017 (548 days) were used:

- For setting up PBMs and DDMs and data-driven models.
- To develop a general workflow for data analysis consisting of data assessment, data analysis, application of data within PBMs and DDMs, and data visualisation.
- To assess the specific FE sensor data for comparison and input to PBM and DDM respectively.

Direct access to the Nagra FEIS database was not possible owing to firewall restrictions at PSI. Therefore, sensor data were supplied and stored in Excel sheets. Upon receipt of the data, not-a-number (NaN) data were deleted. In addition, several procedures to "clean" the raw data were tested, e.g., deleting extreme spikes, but finally the raw data were used in the test case, except for deleting NaN values, which avoided issues associated with justifying which data should be used or ignored.

The second data set, which covered the period from 1 July 2017 to 1 March 2023 (both periods in total about 3,000 days) was handled and used in the same way. Both together yield a data file of 1.9 GB.

3.3.3 Physics-Based Modelling (PBM)

The heat transport equation (Equation 3.3.1) describes the temperature evolution of the thermal processes occurring in the FE experiment. As heat transport in a porous medium is dominated by conduction [52, 53], a reduced heat transport equation has to be solved for the PBM:

$$-\boldsymbol{\lambda}_{\mathbf{i}}\left(\frac{\partial^2 \boldsymbol{T}}{\partial \boldsymbol{x}^2}+\frac{\partial^2 \boldsymbol{T}}{\partial \boldsymbol{y}^2}+\frac{\partial^2 \boldsymbol{T}}{\partial \boldsymbol{z}^2}\right)+\boldsymbol{\rho}_{\mathbf{i}}\boldsymbol{C}_{\mathbf{P,i}}\frac{\partial \boldsymbol{T}}{\partial \boldsymbol{t}}+\boldsymbol{Q}=\mathbf{0} \qquad \text{(Equation 3.3.1)}$$

where λ_i is the thermal conductivity of the medium *i*; *T* is the temperature; *x*, *y*, *z* are the coordinates correspondingly; ρ_i is the dry density of the medium *i*; $C_{P,i}$ is the specific heat of the medium *i*; *Q* is the specific heat source.

For the 3D model setup, a high spatial resolution was chosen to reproduce the material properties for bentonite blocks, GBM, excavation damage zone (EDZ), clay host rock, heater wall and concrete in as much detail as possible, yielding a 3D material parameter distribution for the heat transport model. These parameter distributions are partly time-dependent, because important parameters such as the relative humidity of the GBM have a significant impact on heat transport and temperature evolution in the vicinity of the heaters (tunnel near-field). Near the heaters, initial heating causes humidity to decrease owing to vapour transport away from the heaters, whereas near the clay humidity away from the heaters is increasing. This means that there are different transients (temperature evolutions versus

time) for humidity (and related resaturation) evolution in the GBM depending on the location in the GBM influencing the thermal conductivity evolution, which depends also on the initial GBM dry density distribution.

We identified 40 pairs of sensors for temperature and humidity located at the same position to have correlated temperature and humidity information (Figure 3.8, also important for the DDM). This paired sensor information was used to generated a ML humidity surrogate model for thermal conductivity of the GBM. It includes a thermal conductivity relationship for the GBM as a function of humidity, the initial GBM dry density distribution (taken from slope scans for each of the GBM sections, Figure 3.9), a gravitational term, a cylinder symmetric tunnel term and a sinusoidal heater location term. The result is a ML input of thermal conductivity of the GBM into the PBM, i.e., a time and space dependent (surrogate) function of thermal conductivity of the GBM into the PBM (for details see [33]).

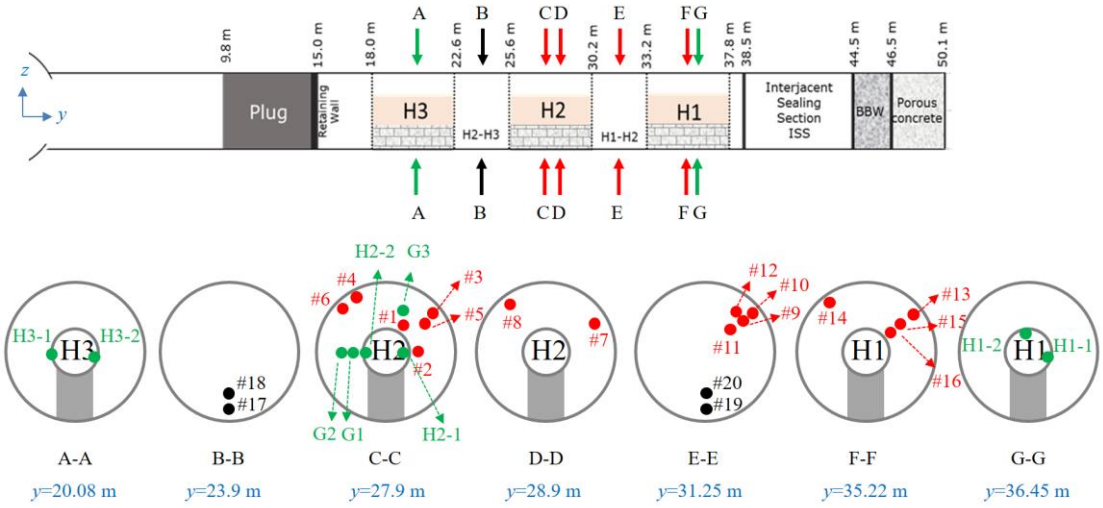


Figure 3.8 - Schematic diagram of positions of 20 pairs of temperature/saturation degree sensors and 10 temperature sensors in the upper and lower half of the GBM and heaters. (Red dot series of sensors represent pairs of temperature/saturation sensors #1~#16 within the upper GBM, black dot series of sensors indicate pairs of temperature/saturation sensors #17~#20 within the lower GBM, green dot series of sensors represent 10 temperature sensors (one temperature sensor in the clay, C1, is not shown) according to [33].

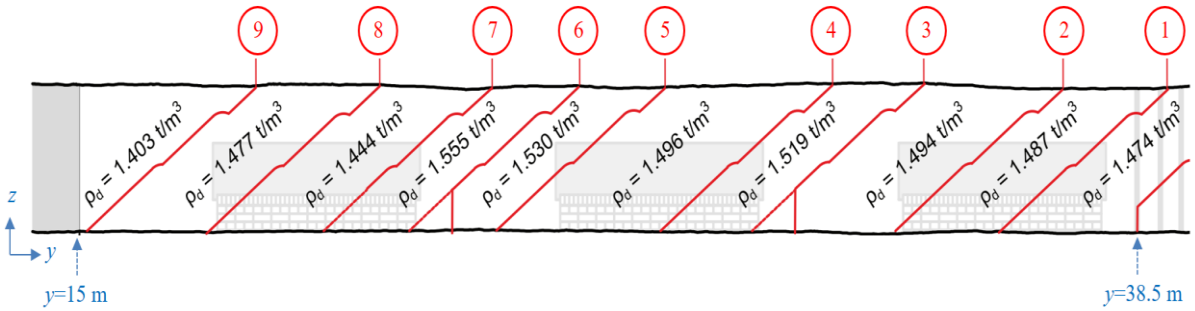


Figure 3.9 – Location of 3D slope scans (red lines) and calculated bulk dry densities of the GBM for each section between the slope scans in the backfilled FE tunnel [47].

A sensitivity analysis on mesh size and model domain size was undertaken to demonstrate a reasonable PBM set-up. Meshes with 1,253,142, 1,719,617 and 2,274,148 nodes were tested [33]) and the related influence of assumed temperature boundary conditions used for model validation (Figure 3.10). For the heat transport initial and boundary conditions, the temperature gradient in the clay host rock was chosen to be 0.06°C/m [54], whereas the power output from the heaters was taken from direct measurements (see Figure 3.11).

With this setup, we calibrated the PBM with 10 representative temperature sensors distributed in the near-field (near Heater 1, Heater 2 and Heater 3, in the GBM at several distances from the heaters, and in the host rock). Good agreement was achieved comparing temperature sensor data and modelled temperature data [33], i.e., during the initial phase, the maximum relative error was lower than 9% for all sensors, which might be influenced by the initial power switching of the different heaters. Later on, the maximum relative error was less than 3% for all sensors and less than 1% for individual sensors (Figure 3.12).

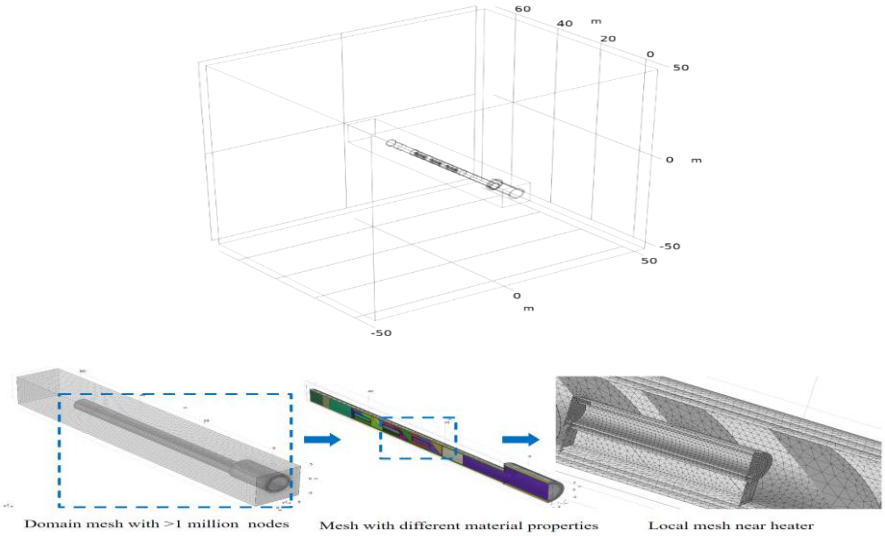


Figure 3.10 - Perspective view and enlarged view of 3D geometric model of the FE experiment. 3D perspective view (top) and enlarged view (bottom) with discretisation according to material properties (layers of density profile are shown in local mesh near the heater according to Figure 3.9) modified from [33].

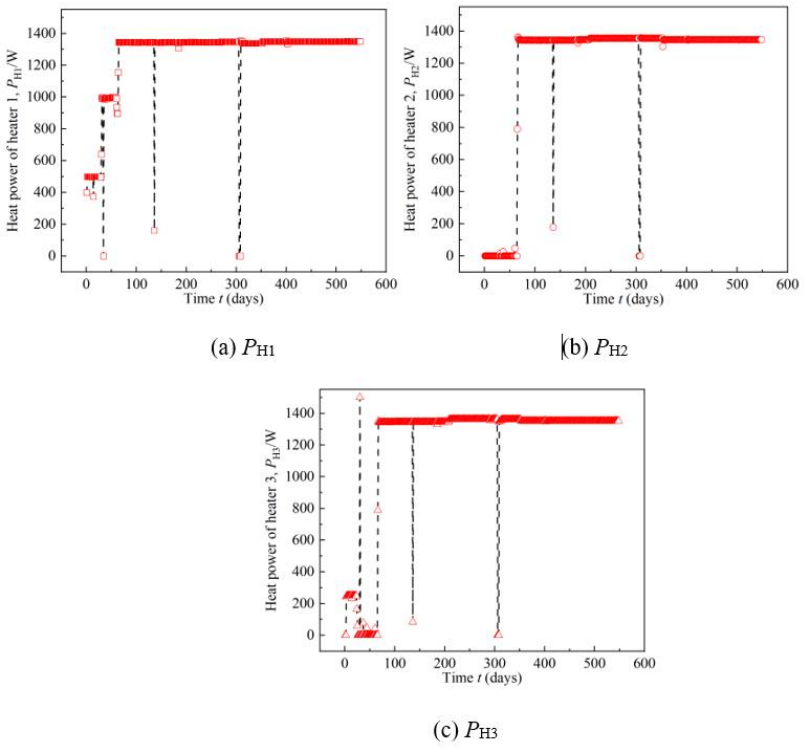


Figure 3.11 - Boundary conditions with heat power of three heaters (P_{H1}, P_{H2}, P_{H3}) from 15 December 2014 to 15 June 2016 (548 days) according to [33]. Dashed lines indicate interruptions to the heater's power supply.

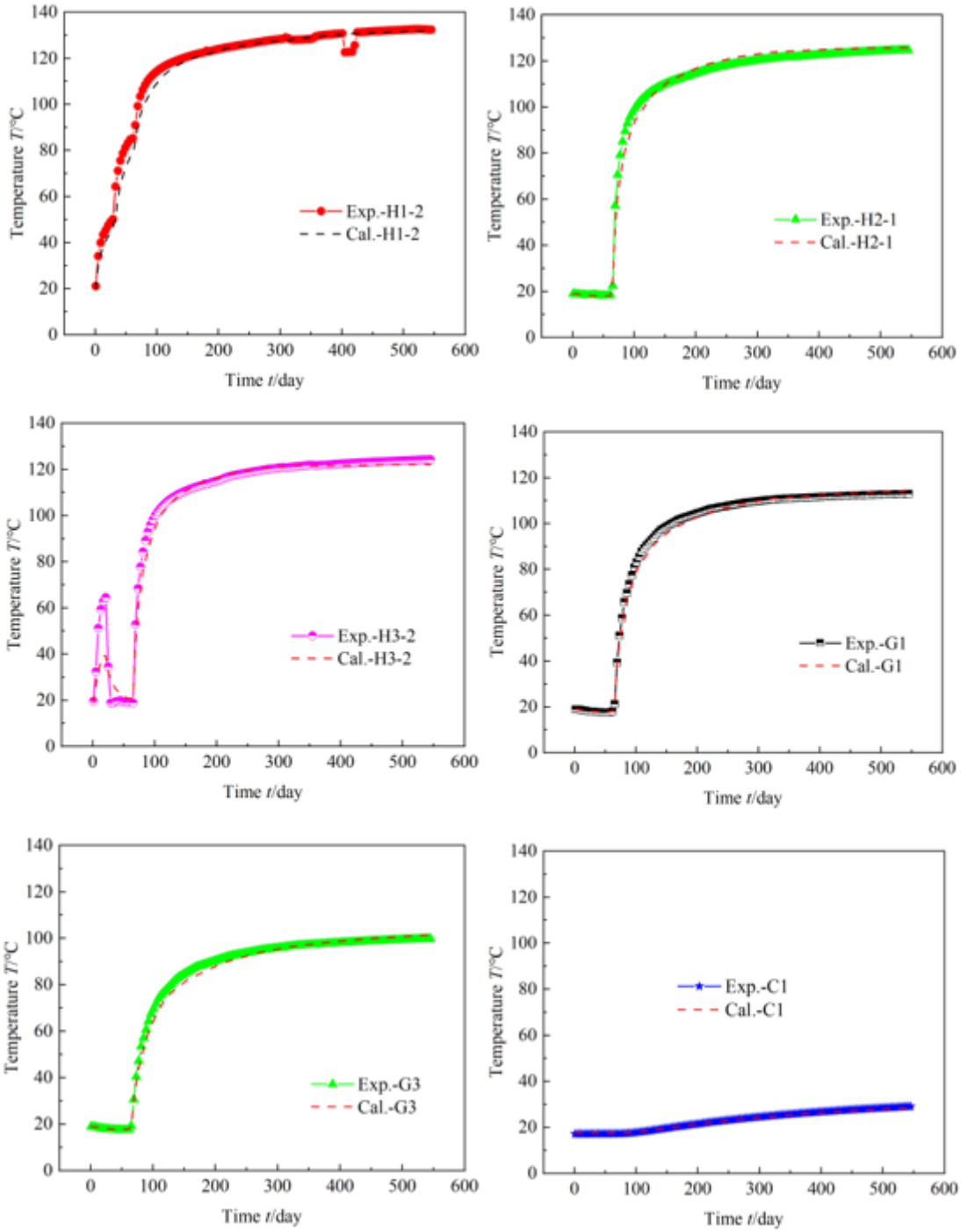


Figure 3.12 - Comparison of temperature evolution T and relative errors between calculation and experimental data at positions of 6 representative sensors over 1.5 years (heater positions: H1-2 (top left), H2-1 (top right), H3-2 (middle left), G1(middle right), G3 (bottom left), C1 (bottom right) according to Figure 3.9 [33].

After calibration of the PBM, known parameter uncertainty ranges (parameter ranges taken from literature, see [33]) and experimental setup details were used to predict the temperature at specified sensor locations, and hence to investigate the influence of individual heat transport parameter uncertainties. Multiple PBM calculations with defined uncertain parameter sampling using an orthogonal test method [55] have been undertaken for seven design parameters (heat power of the three heaters (PH1, PH2, PH3), thermal conductivities of the GBM (λ_{GBM}), the EDZ (λ_{EDZ}), the heater canister ($\lambda_{canister}$),

and the bentonite block (λ_{block})) yielding a matrix of 32 orthogonal test cases. Figure 3.13 shows that different parameters have different influences on the temperature evolution at different locations in the near-field. A small p-test value or a large range value indicate dominant parameter influence. Therefore, such "dominant design parameter analysis" can help to design a targeted temperature evolution or a targeted maximum temperature at a certain location in the near-field of a real repository by choosing or designing e.g., an appropriate heat power (amount of fuel in a canister) or an appropriate thermal conductivity of the GBM (initial GBM dry density distribution).

In addition, multiple PBM calculations with minimum and maximum uncertainty parameter values yield a bandwidth for the calculated temperature as a function of time at different locations. This predicted temperature bandwidth was used to identify sensor data outliers or abnormal data, i.e., measured temperatures that are out of the calculated temperature bandwidth, knowing that predicted temperatures should not differ significantly from measured values, when comparing them with a validated PBM and known range for individual parameters uncertainties (Figure 3.14).

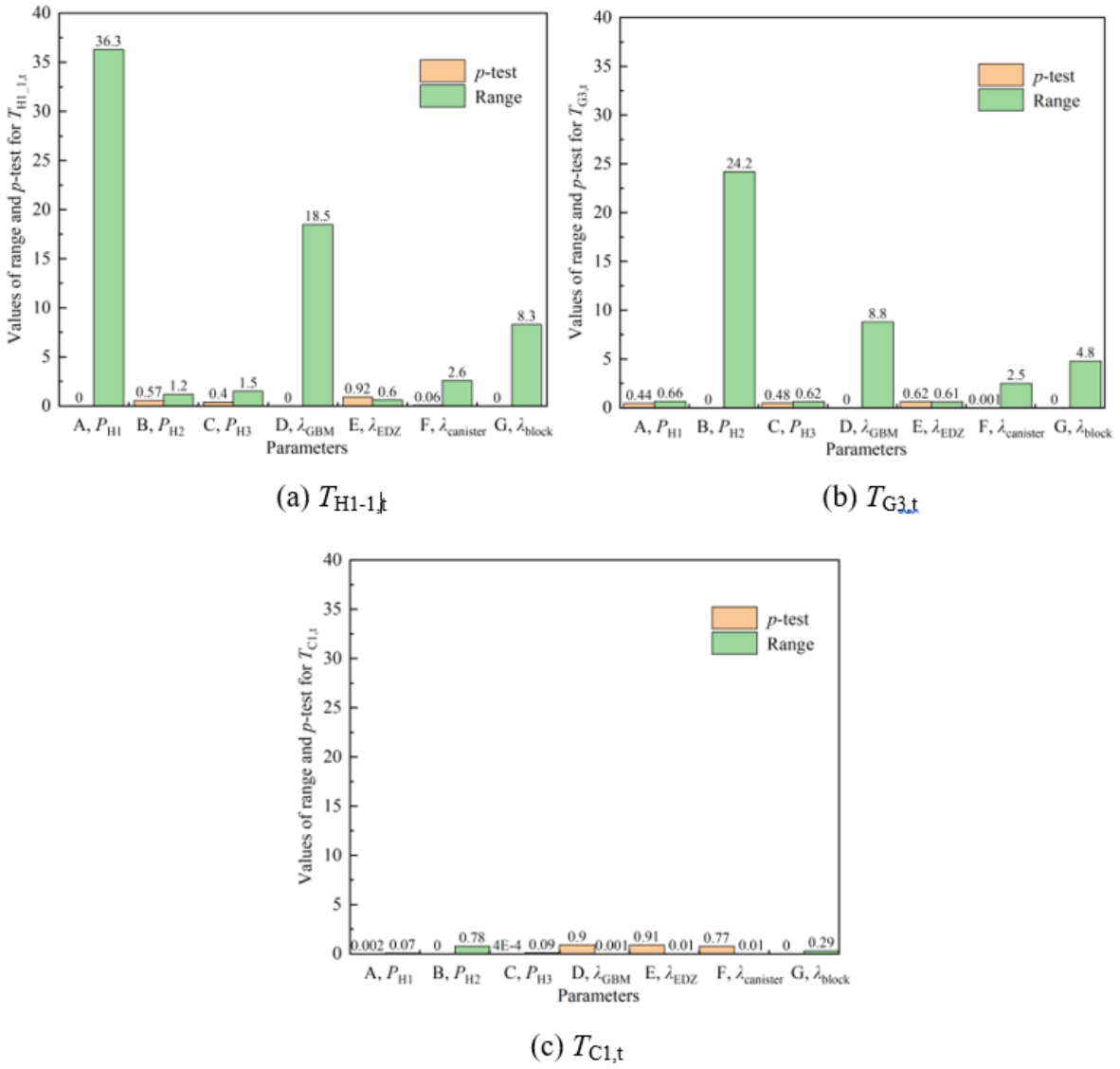


Figure 3.13 - Results of dominant parameters analysis for transient temperature evolution for sensors H1-1, G3 and C1 (compare Figure 3.8 for sensor locations).

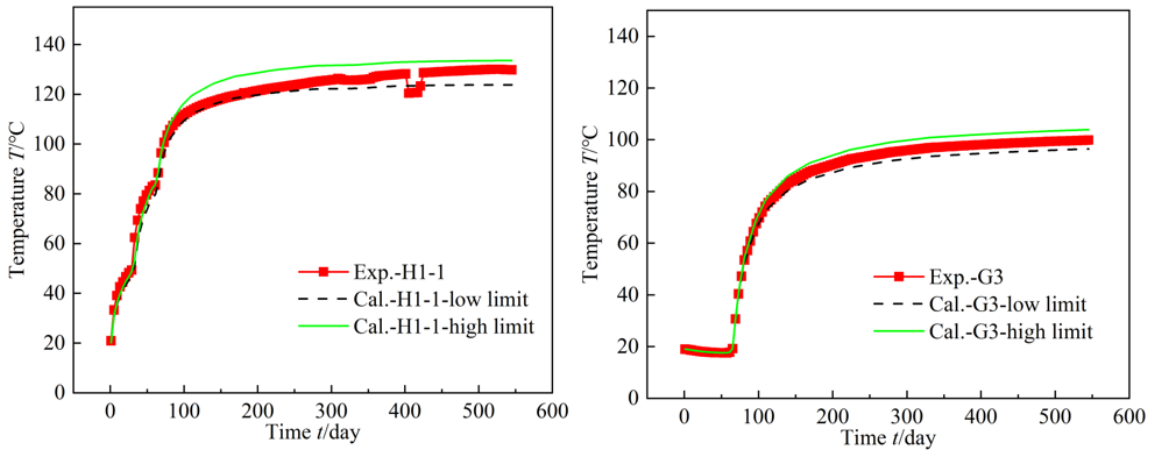


Figure 3.14 - PBM calculations with high and low parameter values yielding temperature bands, which should include the temperature sensor data exemplified for sensor at positions of sensors H1-1 and G3, positions according Figure 3.8 (modified from [33]).

3.3.4 Data-Driven Modelling (DDM)

In order to apply an optimised DDM to the FE experiment sensor data, we performed a review on existing ML methods applied in the area of nuclear energy [32] and identified several ML methods that could be applied to the FE experiment temperature data. We used different data sets with respect to the period covered and the parameters monitored (temperature (T) and relative humidity (RH)) and applied several ML methods. In total, eight ML models were evaluated in the prediction of temperature and relative humidity [33]. The models evaluated were the k-nearest neighbour (kNN), decision tree (DT), support vector machine, random forest (RF), LSTM, artificial neural network (ANN), polynomial regression (PR), and physics-informed ML (PIML, a combination of kNN with the heat power (P) or power density (Q)). The performance of the tools was evaluated using:

- Statistical methods, including mean absolute error (MAE), mean squared error (MSE), and Pearson's r coefficient (R).
- Computational efficiency tests including training time, memory and CPU usage.
- Model robustness analysis, parameter sensitivity analysis and parameter evolution, which reflect the independency of the learning data set (which data are used for learning and which for testing) or the ratio of training and test data (50% learning, 50% testing or 70% learning, 30 % testing) on the calculated results.

This yields an optimal ML tool for the FE temperature sensor data with respect to practical implications and applications, for example, preparation of learning data sets and assessing test data sets. All the models were trained to predict time-dependent temperature and relative humidity in 3D for 1.5 years following the start of heating. 40 pairs of temperature and relative humidity sensors represent the training and testing data set for the different ML methods. The training and testing are described in more detail in [56]. For the PIML methods, heater power as a function of time (for all 3 heaters) has been used as an additional physics-based boundary condition (Figure 3.15).

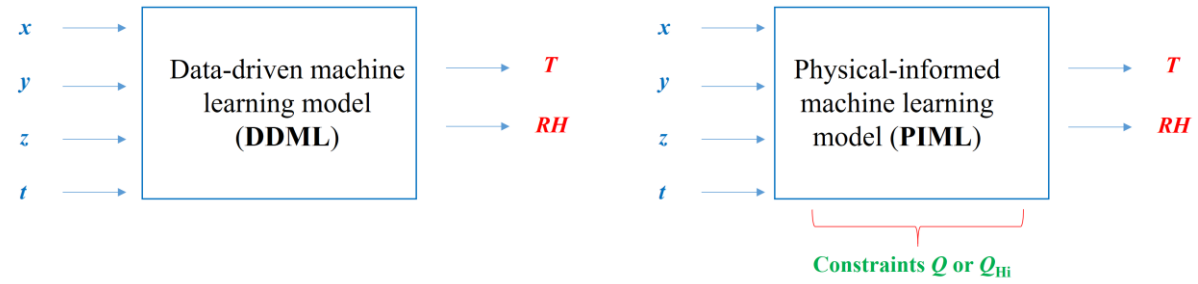


Figure 3.15 - Input and output parameters for ML methods and constraints in case of PIML [56].

According to the performance indicators MAE, MSE, and Pearson's r coefficient R, the PIML, kNN, and ANN models demonstrate good accuracy when compared to the other ML models. High computational efficiency is observed for the PIML and kNN models due to low training time and use of computing resources (Figure 3.16). PIML, kNN and ANN show robust stability in predicting temperature and relative humidity as well as sensitivities in the conducted sensitivity analysis (see [56] for further details).

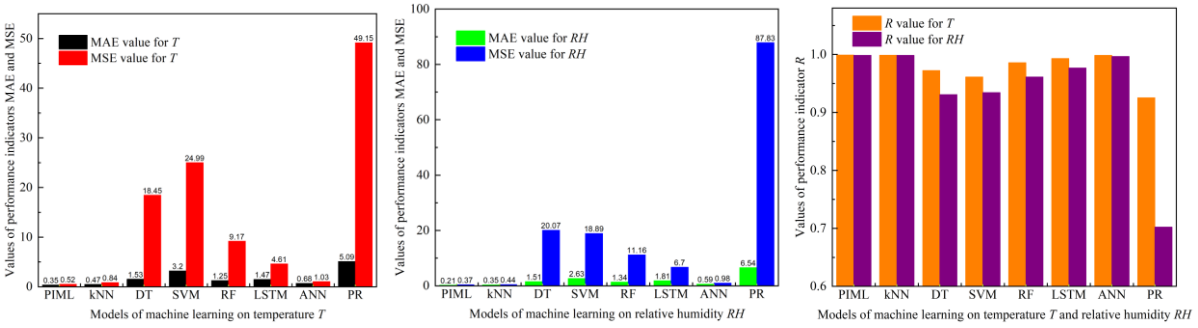


Figure 3.16 - Values of performance indicators MSE, MAE and R with 8 models of ML on temperature (T), relative humidity (RH) and for training-test ratio of 70%-30% [14].

The four best performing DDMs (PIML, kNN, LSTM and ANN) were then compared with representative sensor data for temperature (Figure 3.17) and relative humidity for the first 548 days of heating. There is a general good agreement between the models and the temperature and relative humidity sensor data. Early and late temperature data show some discrepancies from sensor data. However, most obvious is that only the PIML method reproduces the temperature data near Heater 3 well, all other methods fail for the initial heating phase, and the LSTM method appears to be the one with the largest differences. For this reason, we chose the PIML method (combined kNN with heat power P boundary condition) for further use in the DDM digital twin.

In [33] several DDM methods were used to access abnormal sensor data for many of the temperature and humidity sensors and identified a lot of sensors for which predictions in the near future were successful [14]. However, there were also sensors for which predictions were obviously wrong, which depends also on the individual DDM methods applied (Figure 3.17). So far, no clear dependency could be identified.

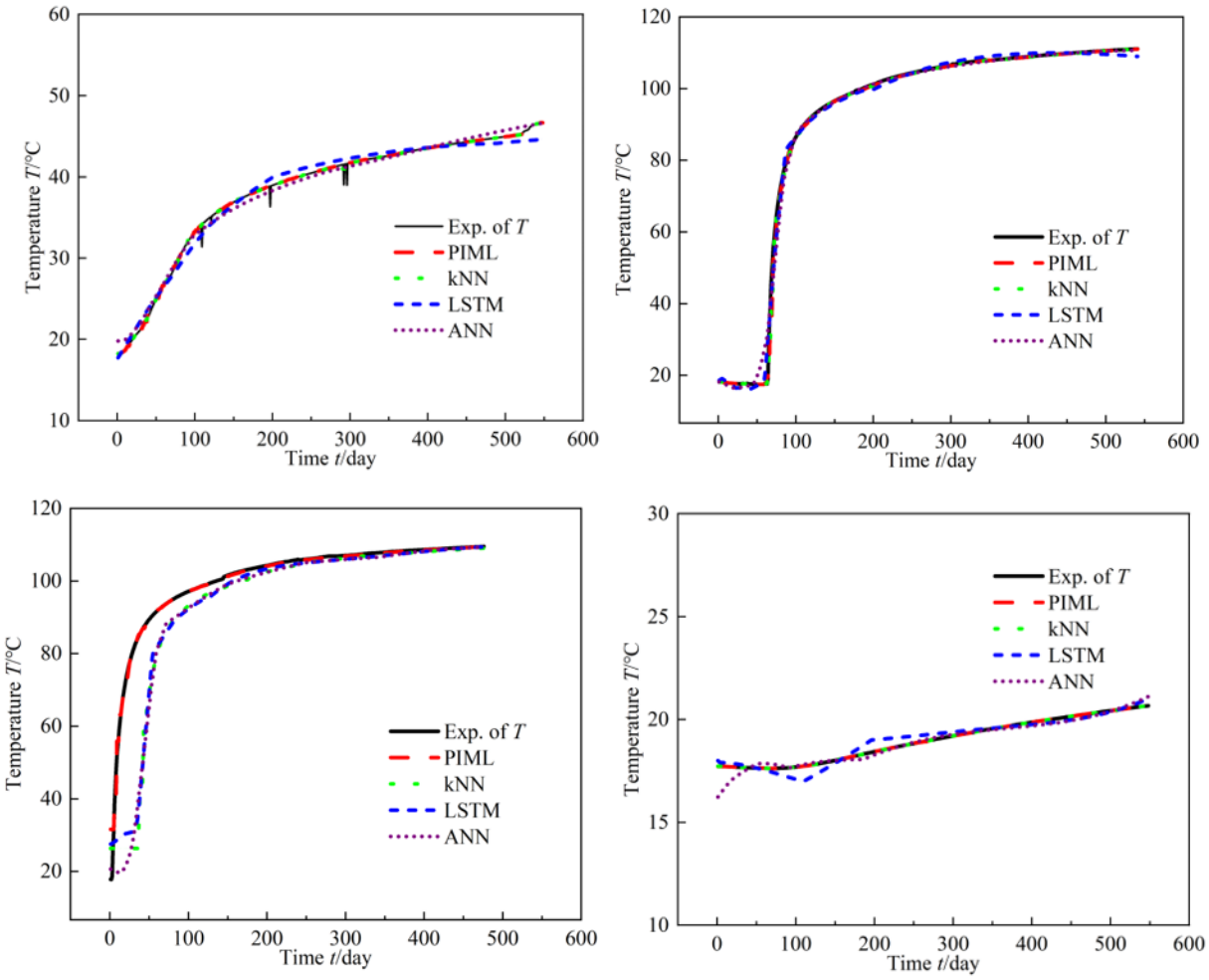


Figure 3.17 - Comparison between experimental data and predictions of temperature and relative humidity with PIML, kNN, LSTM and ANN models near Heater 1 (top left), near Heater 2 (top right), near Heater 3 – bottom left) and in the clay (bottom right).

3.3.5 Digital Twin

Two digital twins for temperature for the FE experiment were developed, a PBM and a DDM, respectively. Both models are described above, where for the PBM some components of a digital twin are already presented: the dominant parameter identification and the identification of sensor data outliers or abnormal data.

3.3.5.1 PBM Digital Twin

The ML-assisted PBM has been validated for sensor data up to 548 days, boundary conditions, and initial and evolving heat transport parameters and might be used for modelling longer periods than the first 548 days as a test for its prediction capability. However, the ML-assisted part for times larger than 548 would be ignored. The only additional input for modelling longer periods would be humidity sensor data to consistently describe the evolution of the thermal conductivity of the GBM appropriately, i.e., the surrogate function for the thermal conductivity has to be updated by upcoming humidity sensor data for times larger than 548 days.

As described above, the ML-assisted PBM of heat transport already allows the identification of dominant parameters using variance and range analysis. Parameters such as the heater power, thermal conductivities of the GBM (λ_{GBM}), the bentonite block (λ_{block}) and the canister ($\lambda_{canister}$) have been shown to dominate the temperature evolution at different locations in the near field. In addition, the parameter

uncertainty ranges of ±10% for λ_{GBM} and λ_{block} are analysed and bands of temperature uncertainties are compared with temperature sensor data, which allows sensor data assessment including identification of faulty sensor data (see Figure 3.14). The identification of faulty data is based on an assumption that that temperature shows a smooth evolution in space and time for such a near-field system; the PBM is validated for the time interval of interest; and the assumed parameter uncertainty is reasonable. The first step in the direction of a digital twin for the FE experiment is to use “incoming sensor data” for improved thermal conductivity calculations in PBM prediction of future temperature evolution in the tunnel near-field. The “incoming sensor data” are here mimicked by successively taking into account new data from the database up to about 3,000 days. Here we used data within a period of about 500 days, because direct access to data was not possible due to firewall reasons and for simplicity. However, the procedure of taking data from the database is applicable up to periods of seconds, generating each time an updated surrogate model for thermal conductivity as input in the PBM allowing modelling temperature evolution in the near-field. Herewith, continuously incoming new data will be used to feed the ML-assisted physical modelling framework allowing improved PBM temperature predictions and improved sensor failure identification.

An example for PBM of the temperature evolution for various locations is given in Figure 3.18 for a validated ML-assisted PBM up to 548 days and a continuously updated surrogate function for the thermal conductivity of the GBM up to 3,000 days using “incoming” humidity data in time steps of about 500 days. Differences might be expected to be minor because the surrogate model for humidity includes already a wide range of humidity values measured within the first 548 days for different locations. However, at sensors near the heater the humidity may develop quite differently from sensors further away from the heaters. For the surrogate function for the thermal conductivity of the GBM deduced from humidity sensor data for the first 548 days it is obvious that at some locations temperature decrease is calculated after reaching a maximum temperature. This would be the result of increased thermal conductivity with increasing humidity in the GBM yielding a more efficient cooling. Taking into account humidity sensor data within the GBM for times larger than 548 days up to about 3,000 days, a different behaviour is predicted (Figure 3.18). As shown in Figure 3.19 for representative humidity sensors, the humidity changes much more in the first 548 days than in the later 2,500 days. Therefore, the surrogate function for the thermal conductivity in the GBM needs to be updated every 500 days approximately to correctly consider the humidity evolution in the GBM (a daily update would have been also successful, but not performed owing to time restrictions). An extrapolation of humidity data from the first 548 days only would yield too large thermal conductivities in the GBM at later times. Such a procedure of stepwise use of incoming sensor data (500 days steps) replicates a PBM based digital twin improving stepwise temperature predictions into the future.

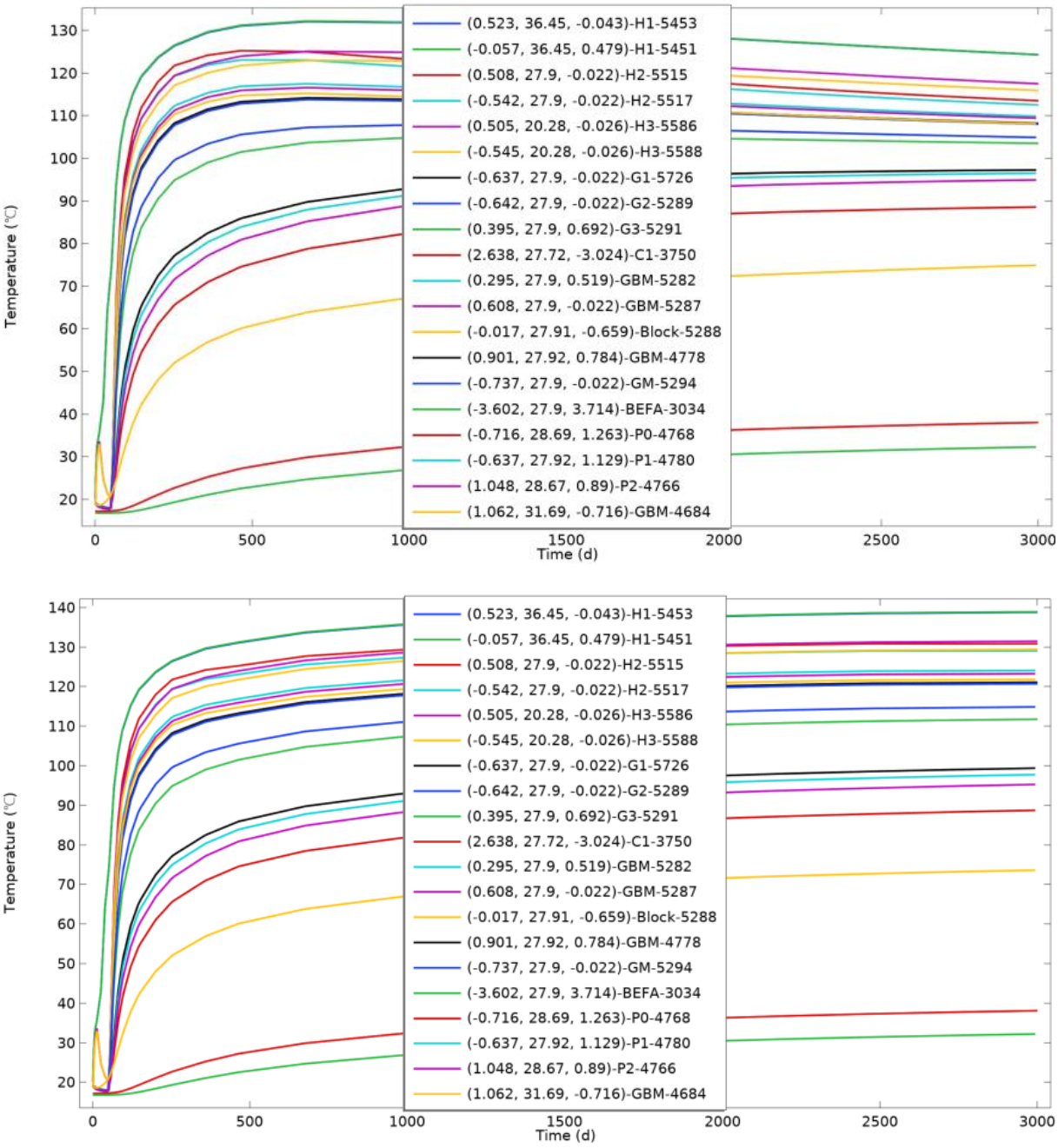


Figure 3.18 - PBM calculations for the temperature evolution at different sensor locations (x, y, z in brackets, and sensor name) in the FE near-field without (upper graph) and with (lower graph) an updated surrogate function for the thermal conductivity of the GBM.

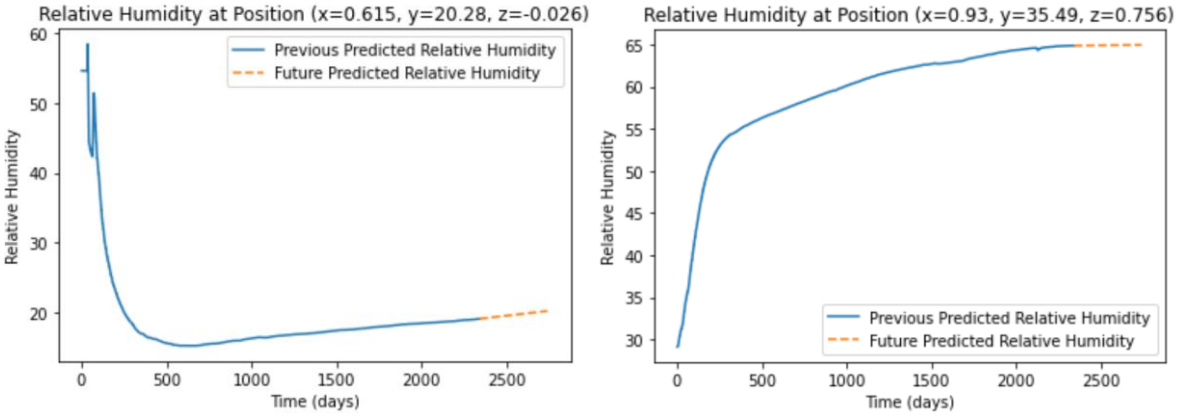


Figure 3.19 - Comparison of DDM of experimental humidity data and related DDM predictions into the near future at different humidity sensor positions. (compare also Figures 3.20-22).

3.3.5.2 DDM Digital Twin

For the DDM digital twin, the best performing DDM (PIML) was selected and applied stepwise to data from individual temperature and humidity sensors for periods up to 548 day, 1000 days, 1500 days, 2000 days and 3000 days. The results are shown in Figure 3.20 and Figure 3.21 exemplified by selected sensors for different DDM learning and testing times. The DDM reproduces well all temperature sensor data up to 548 days, even if there were sensor data failure identified using the PBM, and temperature spikes were reproduced by the DDM (Figure 3.21).

When it comes to predictions with the DDM digital twin, i.e., predicting in future times for which no data have been used for learning, then predictions seemed to be wrong in some cases (Figure 3.22). Predictions show no smooth temperature behaviour but edges from last learned data point to predicted future times. Also, sensor data with abnormal behaviour (spikes, longer temperature drops) seemed to be well reproduced by our DDM, which means that sensor data failure or abnormal data identification is not as straightforward for the DDM digital twin as for the PBM digital twin.

3.3.5.3 Future Work and Applications

The same digital twin framework has been used with a PBM and a DDM for temperature and relative humidity (DDM only) allowing just an exchange of the modelling tool, which would also allow a digital twin based on a combination of both, PBM and DDM (e.g., sequential or iterative). However, the predictive capability or abnormal data identification of a DDM based digital twin is limited due to the limited learning period, nevertheless, the DDM based digital twin might be much faster with respect to computational performance and an appropriate use of both PBM-based and DDM-based digital twin might be the most efficient one.

It should be noted that such a digital twin framework could be applied to a repository and help to design, optimise, and define the monitoring setup as well as contribute to the assessment of a repository providing detailed understandings of the performance behaviour of a DGR using parameter sensitivity and uncertainty analysis.

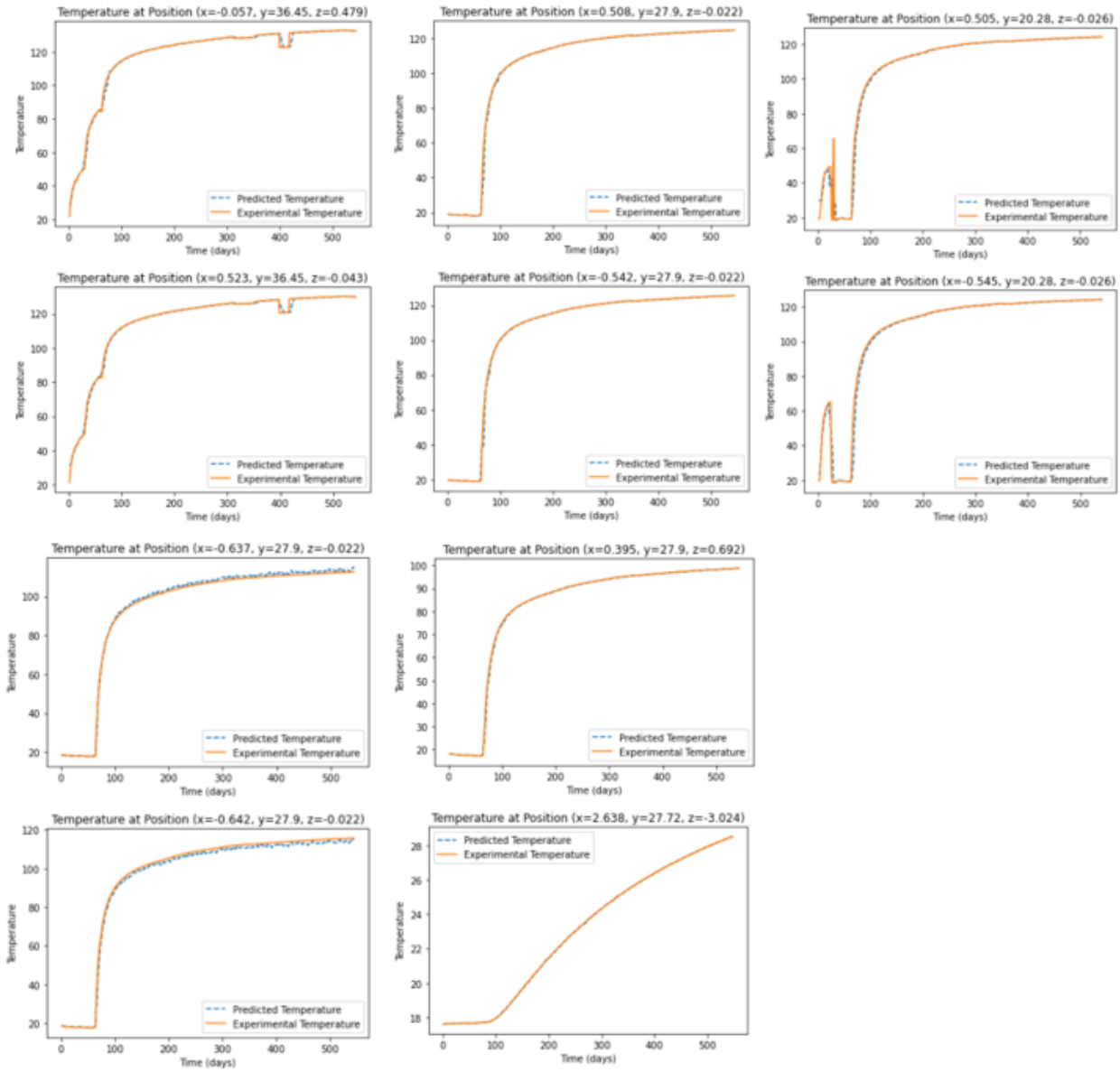


Figure 3.20 - DDM digital twin calculations for the temperature evolution at different sensor locations in the near-field for data up to 548 days learning and testing time.

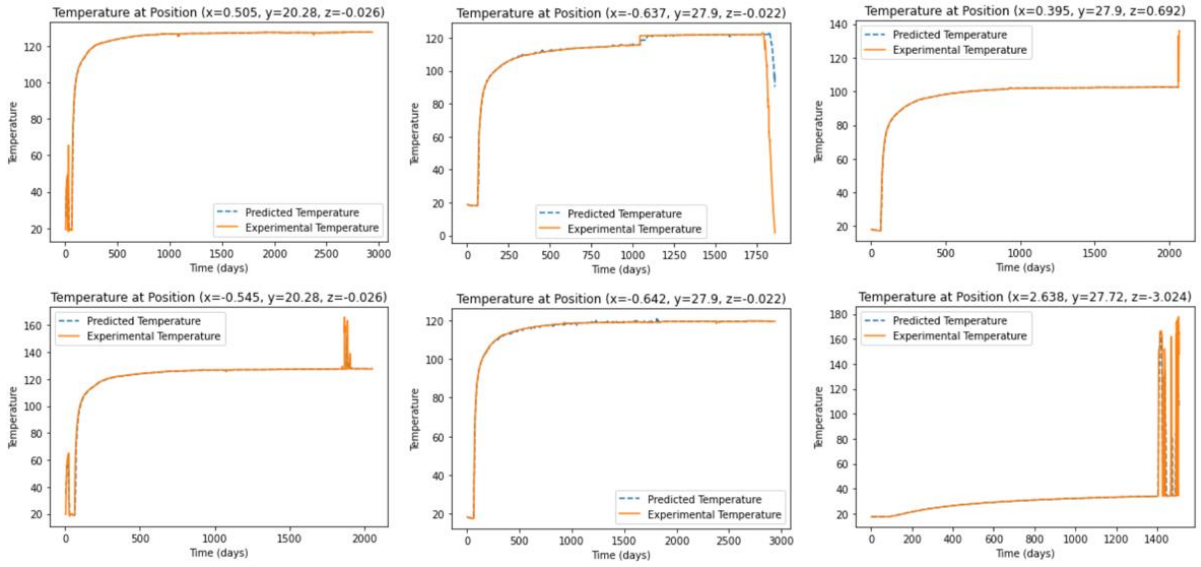


Figure 3.21 - DDM digital twin calculations for the temperature evolution at different sensor locations in the near-field for data up to 3000 days learning and testing time.

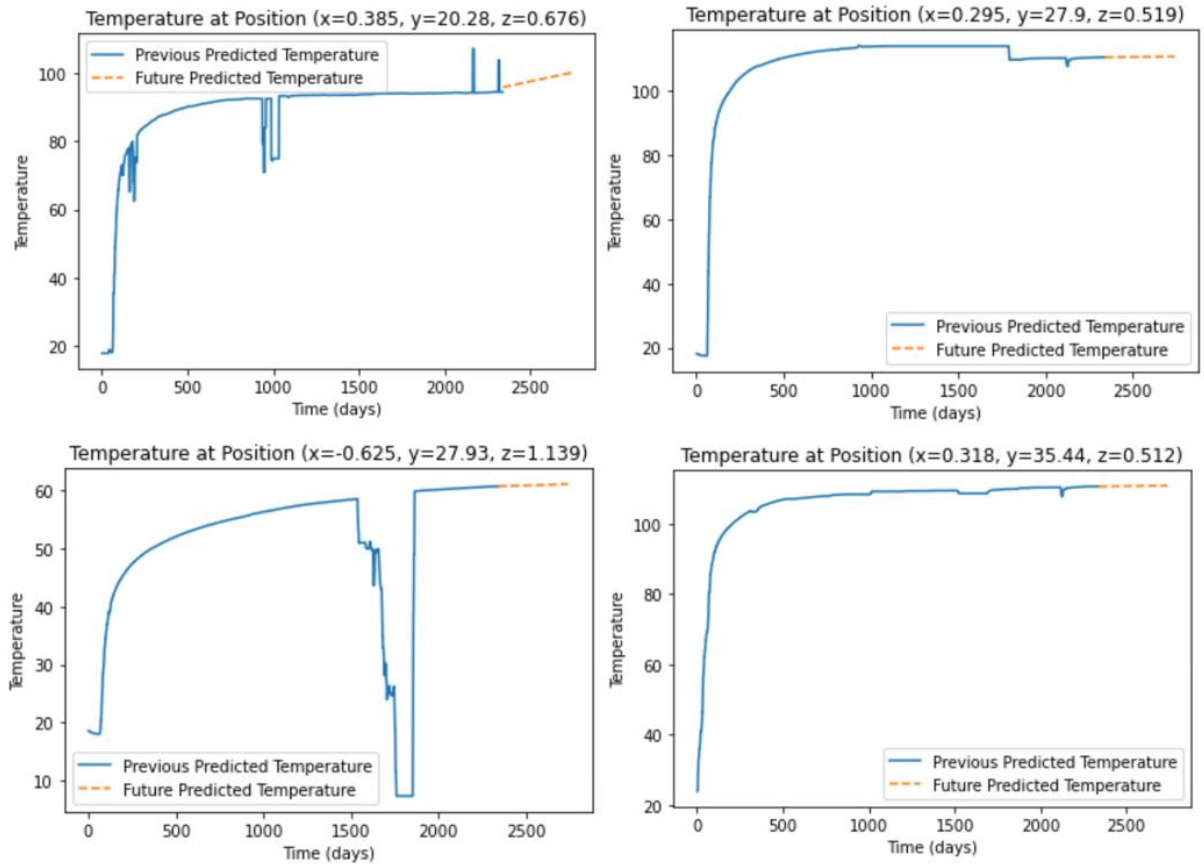


Figure 3.22 - DDM digital twin calculations for the temperature evolution at different sensor locations in the near-field with predictions using learning and testing time represented by the blue lines.

3.3.6 Summary and Conclusions

A concept of ML-assisted physics-based 3D heat transport modelling for the FE experiment has been developed and verified. ML is applied in this context to humidity in the GBM yielding a physically and NN-based surrogate model for the thermal conductivity of the GBM. The surrogate model is required to calculate the temperature evolution in the FE tunnel near field. The data could be described as sparse, in relation to the full 3D model data set available. Results of ML-assisted heat transport calculations are validated with the large experimental data set of the temperature time series data for 10 temperature sensors.

The PBM is used to investigate the dominant parameters influencing the temperature evolution in the vicinity of the FE experiment, the results of 32 orthogonal test cases have been analysed systematically. Estimated parameters uncertainty ranges of ±10% for thermal conductivities of the GBM (λ_{GBM}) and bentonite block (λ_{block}) are analysed and bands of temperature uncertainties are compared with temperature sensor data, which allows sensor data assessment including identification of the faulty sensor data. Such a ML-assisted physical modelling framework can be applied to future repositories.

A review of ML DDM methods has been performed and eight promising methods were evaluated in the prediction of temperature and relative humidity within the experiment. In terms of performance indicators MAE, MSE, and Pearson's r coefficient R, the PIML, kNN, and ANN ML models demonstrate good accuracy when compared to other ML models. The PIML and kNN models show high computational efficiency, the PIML, kNN, and ANN models show robust stabilities and sensitivities in the conducted sensitivity analysis in predicting both temperature T and relative humidity RH. The PIML model (combined kNN with heat power P boundary condition) performs best for this study. A lesson is that comparing different ML algorithms can be difficult owing to the many factors involved in a tool's performance. This test case has illustrated the criteria that can be used to qualify a ML method (including statistical methods, computational efficiency tests and the independency of the learning data set).

The general setup of two types of digital twins, one using a PBM, the other using DDM, has been realised. These twins allow temperature predictions for any location within the 3D model domain or the x, y, z data range of the DDM. Prediction into the future temperature evolution is possible with the PBM-based digital twin allowing also predictions including assumptions on uncertain model parameters (or their evolution) to identify abnormal sensor data or sensor failures. The PBM digital twin has been used for parameter uncertainty related to sensor data analysis. In addition, it has been used to identify the dominant parameters influencing the temperature in the tunnel near field. The PBM results for ~100 parameter sets containing variations of eight parameter values (including heater power and thermal conductivities) are used to define input and output parameter sets for ML to generate a surrogate model for the related input and output parameters. With such a surrogate model, millions of calculations for different input parameter and temperature output sets could be generated in a few seconds, allowing the quantification of dominant parameter uncertainty and identification of the most dominating parameters for temperature evolution in the tunnel near field. The DDM based digital twin could be used partly for predictions into the near future temperature development, maybe due to smooth temperature evolution for a system like the FE experiment, but results have to be handled carefully because they are not founded by related learning data of the corresponding DDM.

An extension of the temperature (humidity) digital twin for the FE experiment is planned in the direction of PBM of hydraulics, mechanics and gas migration looking at the available data from the FE experiment. However, one has to take into account that humidity sensor data as well as pressure, gas, stress, strain or displacement sensor data are not as smooth as temperature sensor data. Therefore, a combined PBM and DDM might be successful when developing a THMC digital twin for the FE experiment.

3.4 Test Case Description and Results: UFZ

Authors: Nico Graebling and Jörg Buchwald (UFZ)
Editor: Matt White (Galson Sciences)

3.4.1 Introduction

The work undertaken by UFZ within MODATS was multifaceted, and the different contributions are divided into three subsections.

First, in Section 3.4.2, interactive visualisation systems are described. These can be considered as digital twins with a focus on visualisation. The applications were developed primarily as desktop virtual reality applications and can be used for different purposes:

- As integrated information systems for domain experts.
- As a starting point for multi-party-dialogue.
- As teaching material for university students or interested members of civil society (referred to as intermediates).

This aspect of the work is described in more detail in [57 and 58].

Second, in Section 3.4.3, the results of the comparison of simulation and observation data for the uncertainty analysis in the scope of the FE experiment are summarised. This covers the preprocessing of temperature measurements, a statistical analysis and the visualisation of the results. The work is described in more detail in [59].

Third, in Section 3.4.4, the ongoing development of a generic tool for the visual comparison of simulation results, referred to as OpenGeoSys (OGS) - Visual Comparison of Simulation Results (VisCoSiR) is presented. OGS is open-source software for the simulation of THMC processes in porous and fractured media. OGS-VisCoSiR is a visual tool for the comparative exploration and analysis of simulation data for use alongside OGS.

3.4.2 Interactive Visualisation Systems

UFZ developed two virtual reality applications, which can be considered as digital twins with a focus on visualisation, as the applications contain virtual models of the URL and its geological context and measurements displayed in this virtual environment update automatically whenever sensors measure changes in the real URL. Virtual reality is a computer-generated simulation of a three-dimensional image or environment that can be interacted with in a seemingly real or physical way. The first application, the Virtual Experiment Information System (VEIS), is an integrated information system for domain experts, that includes the geometry of the tunnel system, the geological environment and information associated with experiments undertaken in the Mont Terri URL (see sub-section “Data Integration” in Section 3.4.2.1 for details). In a second step, the VEIS for domain experts was used as a basis for the development of a virtual field trip on methods and results in the context of radioactive waste management research at Mont Terri. This second application targets intermediates and can be used as a starting point for a multi-party-dialogue. Both applications share a similar architecture, that is displayed in Figure 3.23. The base is a 3D scene, created in the Unity Game Engine. In addition to data that does not change during runtime (“static data”, e.g. the tunnel system), information about boreholes and sensors from online databases is included in the application. Infrastructure for interactions and data integration comes from a module called UFZ Framework.

3.4.2.1 VEIS – A Digital Twin Prototype

We created a digital replica of the Mont Terri URL, that integrates heterogeneous data from several different sources. It allows data to be accessed and explored within its spatial context without prior technical knowledge. Both simulation results and observation data are displayed within the same system. In the following subsection, the development process of the application is explained. It is

followed by details about the data integration. Afterwards, visualisation of the experiments and the user interface is explained. The final subsection provides conclusion on the VEIS work.

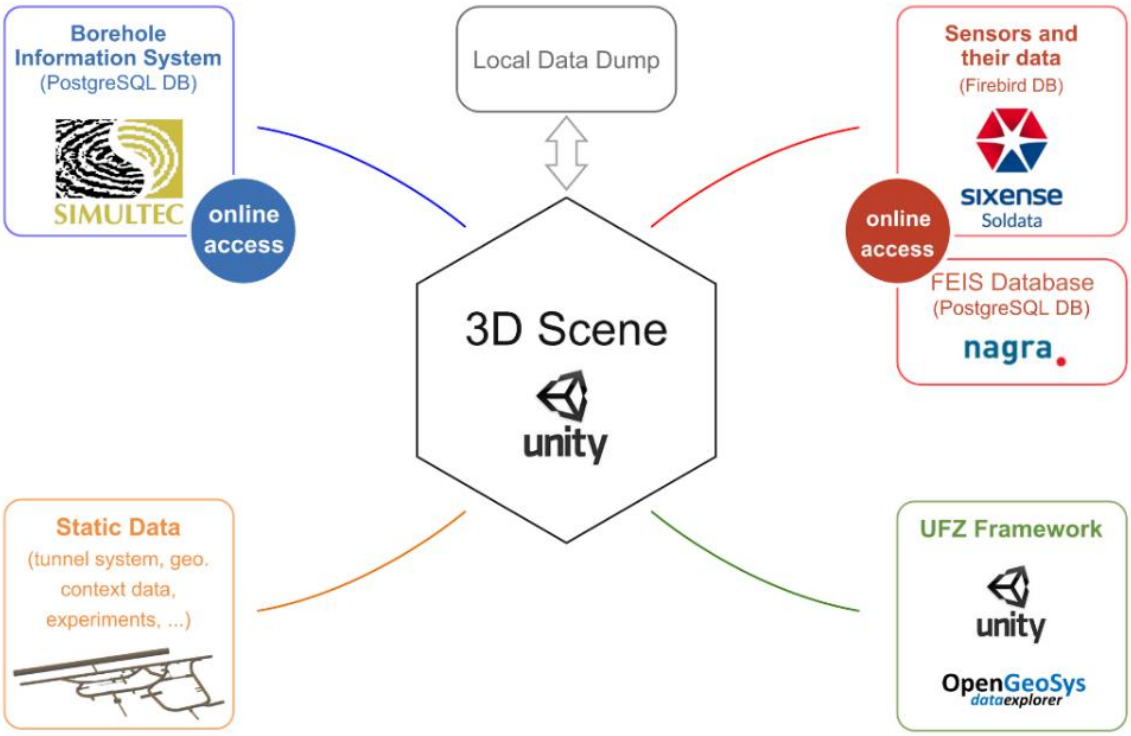


Figure 3.23 - Technical infrastructure and data sources that are included in the VEIS.

Development

We collaborated closely with domain experts to create a prototype of the VEIS for the Mont Terri URL. Highly heterogenous data sets from the URL were collected, converted, preprocessed and curated. The formats of the converted data include, for example, the Visualization Toolkit (VTK) format, which is well established in the scientific community, and the glTF file format, which provides high flexibility for use in 3D visualisation [60]. The application was created using the Unity Game Engine. Unity was chosen for its versatility, allowing deployment on desktop PCs, stereoscopic virtual reality devices like head-mounted displays, and in CAVE[5]-like setups (like the VisLab at the UFZ Leipzig, www.ufz.de/vislab). In this way, the application supports a wide range of input devices, from standard desktop controllers to specialised virtual reality controllers. Additionally, Unity offers the flexibility to extend functionality through additional modules and scripts, and enables the recording of interactive sessions, which can be useful for presentations.

Data Integration

Figure 3.23 presents an overview of the data sources integrated into the application. This data integration will be explained in detail in the following subsection. The model of the Mont Terri URL's tunnel system, originally created via laser scan, serves as a frame of reference for the other data sets that are included in the application. It is enhanced by a visualisation of the stratigraphic layers and the major tectonic faults (Figure 3.24) representing the URL's geological context.

The results of six numerical simulations have been integrated in VEIS. These simulations were performed in previous work, using the OGS simulation software [61]. The data sets vary in terms of structure (raster, mesh, and time series data), dimensionality (ranging from 1D to 4D), and scale (from

[5] A virtual reality CAVE is a virtual reality space; essentially an empty room in the shape of a cube in which each of the surfaces (the walls, floor and ceiling) may be used as projection screens to create a highly immersive virtual environment.

meters to kilometres). Despite the diversity of formats, a consistent processing approach was applied. They were first preprocessed and incorporated into a unified geographic context using the OGS Data Explorer, other OGS preprocessing tools [62] and ParaView [63], which are open-source tools providing interfaces for data integration and conversion, including geographical information system formats and geological modelling software platforms such as GOCAD and Petrel.

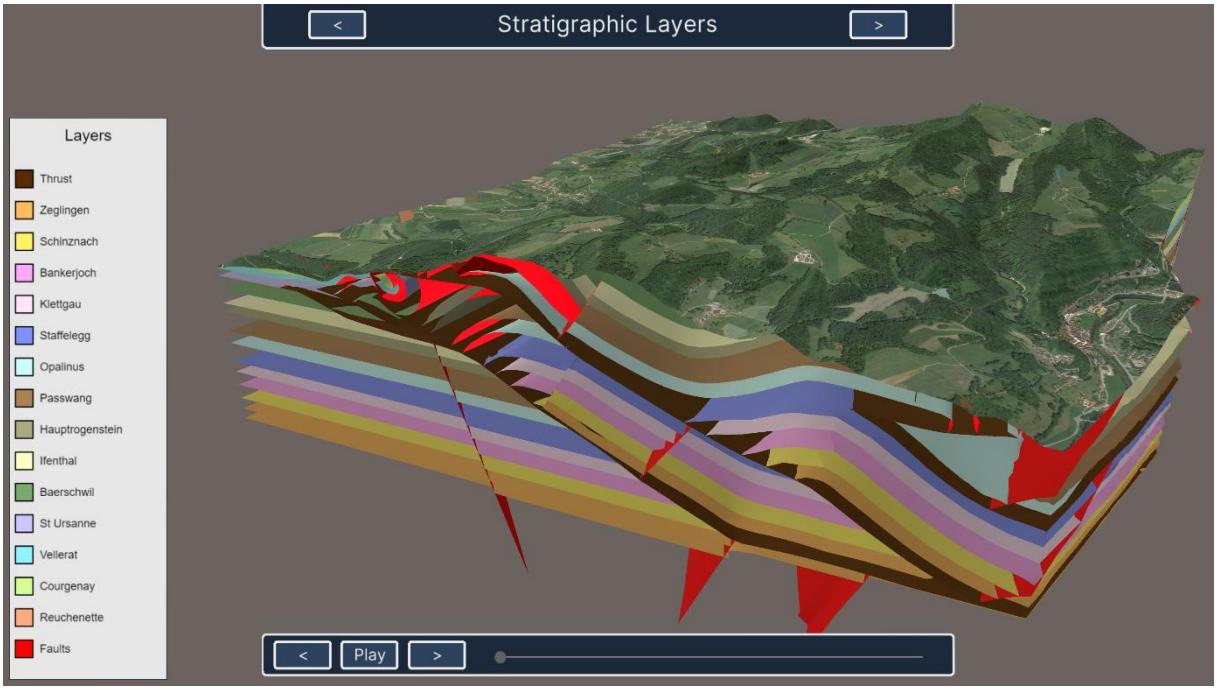


Figure 3.24 - Screenshot from the interactive visualisation systems for the Mont Terri URL illustrating the geological context in form of the stratigraphic layers and major faults.

Subsequently, a variety of post-processing and visualisation techniques were applied, including geometry clipping, surface extraction (e.g., contour surfaces), projections, and colouring (details of this methodology can be found in [64, 65]). The data were then exported as geometric objects in the glTF file format [60]. Importing these objects into the Unity scene involved the use of custom importer scripts (implemented in C#) and a dedicated Unity extension. Additionally, intermediate scripts were created to transfer supplementary data from ParaView to Unity. This supplementary data includes labels for the stratigraphic layers, colour mapping tables, and individually formatted time labels.

In addition to these static data, the system provides access to two productive databases, offering real-time updates on borehole and sensor data. First, data from the Mont Terri Borehole Information System (BIS) was integrated. The BIS is based on a productive PostgreSQL database, comprising 19 tables that represent various objects, including boreholes and their drilling sections. The borehole table holds details such as a unique borehole ID, links to associated documents, and geometric attributes like 3D starting and ending points, along with azimuth and dip angles. Furthermore, it provides information about drilling sections, such as the diameter and length.

To incorporate borehole and section data from the BIS into the prototype, PostgreSQL libraries were imported into the project and C# scripts were implemented to create a communication layer between the Unity application and the borehole database. This layer offers flexible access to live data and allows for switching between databases (remote or local) with ease. For the visual representation within the 3D scene, only valid boreholes with either both starting and ending points or a combination of the starting point, dip and azimuth, and the section length are considered. All boreholes are assumed to be straight. Boreholes lacking this geometric information are categorised as invalid. At the time of development, the database contained 1,365 boreholes, of which 16 were considered invalid.

In practical use, the prototype retrieves data as follows:

- Upon application launch, it loads the geometric data for all boreholes.
- This data set is then preprocessed to filter out invalid boreholes.
- Finally, meta data, such as document links and section information, is linked to the data objects associated with the borehole.

Internally, Unity uses single-precision floating-point vectors for positioning and orienting 3D objects (i.e., 32 bits are used for storing a single floating-point number). However, the BIS stores start and end points with double-precision (i.e., 64 bits per floating-point number). When displaying small boreholes in a large environment, this precision reduction can lead to significant artefacts. This issue has been mitigated by using an extension for double-precision floating-point vectors, at least for all calculations before displaying the visual representations of boreholes.

In addition to the borehole data, sensors and their measurements are included into the application. Therefore, access to a Firebird database for observation data of the URL in general and the PostgreSQL database specific for the FE experiment has been implemented. In this way, sensor measurements of various parameters, including pressure, temperature, displacement, and humidity are available in the application. Similar to the approach with borehole data, we implemented C# scripts to establish an additional communication layer between the Unity application and the sensor databases, ensuring maximum flexibility.

Several preprocessing methods for the sensor data were implemented, including aggregation (e.g., averaging measurements per day), and creating a version of the time-series without outliers. To identify outliers, the maximum realistic difference between consecutive measurements was determined in consultation with domain experts for each sensor type. Any aggregated measurements deviating more than this maximum delta from the previous aggregated measurement were considered outliers. For example, for temperature data, for which the database stores multiple values per hour, this threshold is set to 2°C per day. The outlier removal, including the criterion, is displayed in the chart view for transparency reasons (Figure 3.25).

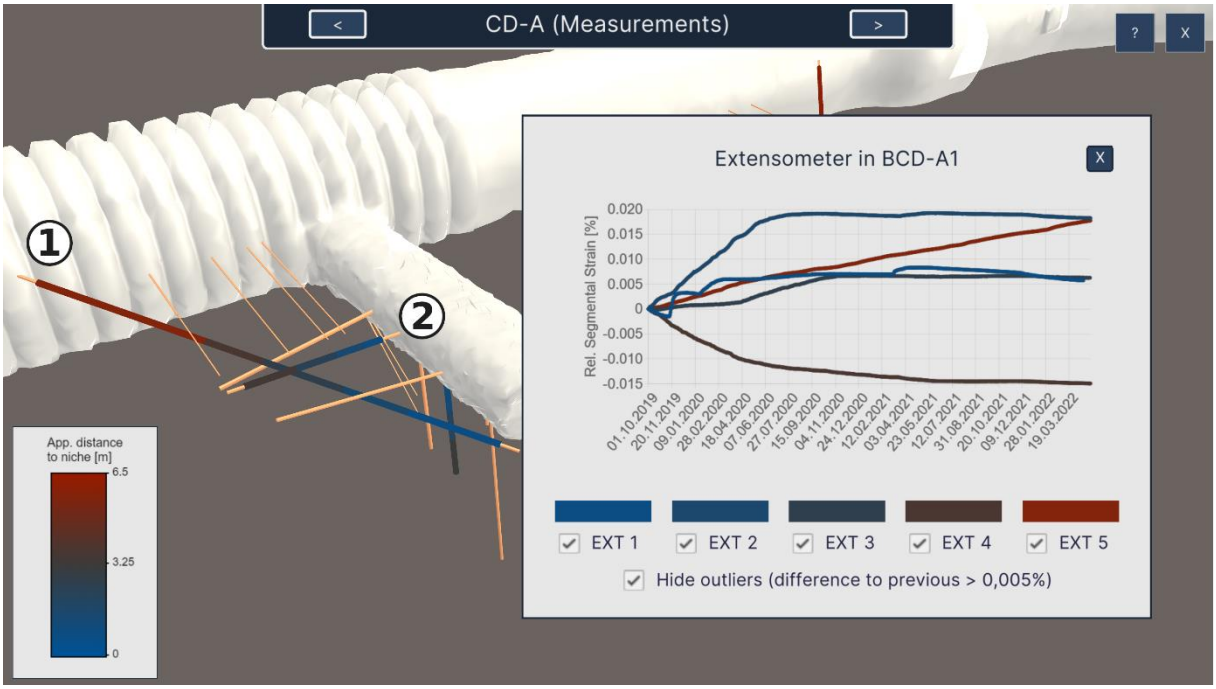


Figure 3.25 - Screenshot from the interactive visualisation systems for the Mont Terri URL illustrating sensor measurements from the online databases, depicted with outlier removal.

For the borehole data as well as for sensor measurements, the prototype also provides functionality to save a combined local backup copy of the data retrieved from the databases once the online databases have been accessed. This creates the possibility to use the application also in situations in which the

database is not accessible or the user's internet connection is not stable. The backup is an encapsulated combination of the following data sets: borehole positions and dimensions, links to borehole-related documents, sensor positions and sensor measurements. It can be considered a local file and must not be confused with a local database. From a technical perspective the data dump is a text file in the JavaScript Object Notation (JSON) file format containing all data objects in a structured way.

Visualisation-Based Digital Twins of Experiments

The VEIS contains not only data sets for the whole URL but also detailed information and visualisation for selected experiments. In the following, the incorporation of information and visualisations for the FE experiment is described. In addition to measurement data, the effect of heating has been modelled in a heat transport simulation with a time span of roughly 15 years using OGS. It contains 230 time steps and covers a volume of 100 x 80 x 100 m^3 around the tunnel. The modelled area consists of 164,472 hexahedra and 169,806 points. The VEIS displays these simulation results as a clipped volume (Figure *3.26*).

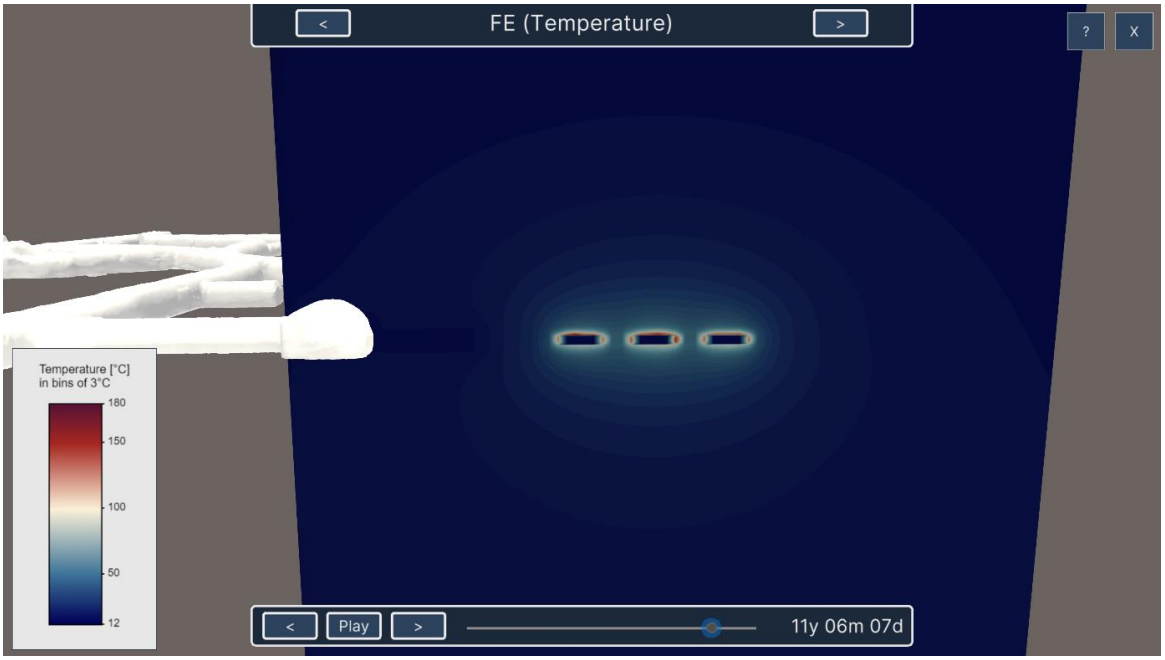


Figure 3.26 - Screenshot from the interactive visualisation systems for the Mont Terri URL illustrating temperature simulation results from the FE Experiment as a clipped volume.

The diverging colour map for temperature values allows to perceive high temperatures close to the heaters (white to red) as well as lower temperatures in the Opalinus Clay (blue to white). As structures in the blue-to-white area of the colour map were hard to perceive, the colour map has been sampled in discrete bins of 3°C. Although this approach reduces the displayed degree of detail, it allows to inspect the overall more important spatial structure of the heat emission and expansion.

The heat induced by the heaters affects not only the temperature but also other characteristics of the rock. Therefore, another model simulates mechanical parameters such as displacement, stress, and strain of the surrounding Opalinus Clay. This simulation and the underlying model have the same characteristics, such as domain extent and time span, as the temperature simulation mentioned above. In the application, the results of the vertical stress simulation are displayed as six contour surfaces of increasing values. A sequential colour map has been used for the surface colouring here because it allows to intuitively analyse spatial trends and interrelationships.

For both, the temperature as well as the vertical stress simulation, users can individually select a time step or play the simulation visualisations as an animation. Beyond these detailed data sets for the FE experiment, various data sets of the Cyclic Deformation Experiment and the Fault Slip Experiment have also been included in the VEIS. This covers geological context data sets, simulation results and

illustrations. In addition, we collected basic information for a selection of 63 experiments. The experiments' short names, short descriptions and positions in 3D space attached or close to the tunnel system have been extracted from [66] and the website of the Mont Terri Project. Based on these data, users can inspect the boreholes associated with an experiment by selecting an experiment's 3D marker in the scene (Figure 3.27).

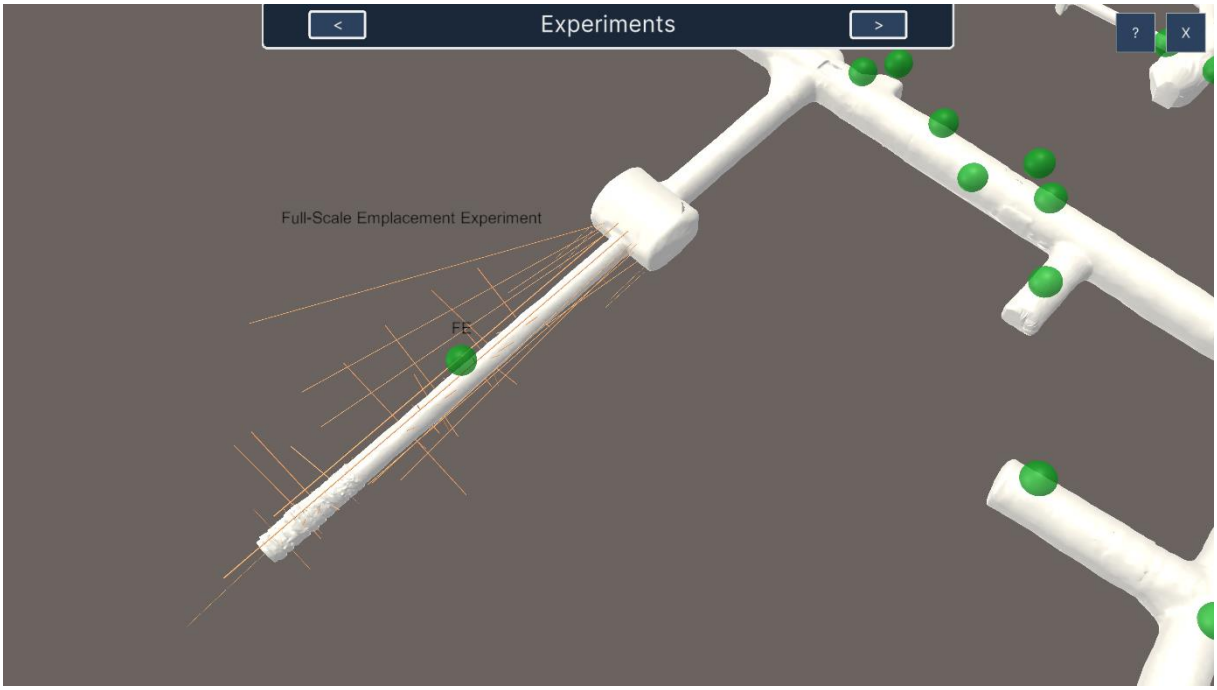


Figure 3.27 - Screenshot from the interactive visualisation systems for the Mont Terri URL illustrating experiment markers and the selection of the FE experiment's boreholes.

User Interface

All functionalities of the VEIS are reflected in the Main Menu. There, users can change the visibility of data sets, control animations of time-dependent data, select data sources and set up the visualisations in a custom way. In addition to the custom setup of the current view on the data via the Main Menu, users can visit viewpoints for which the object visibilities and interaction patterns are predefined. Furthermore, the application provides a Borehole Menu, which opens when users select the visual representation of a borehole in the scene. It provides basic borehole information, drilling sections details, a list of the installed sensors and access to the sensors' measurements in form of charts (Figure 3.24D).

Conclusion

By consolidating all data within a common spatial context, our application promotes effortless exploration of static and live updating data from the URL. The implemented system lowers the cognitive effort usually needed to gain insights from measurements, simulation results and contextual data. Moreover, the interactive visualisations significantly reduce the technical hurdles typically associated with such systems. As the system automatically updates when the underlying data sources (like boreholes and measurements) change, it can be considered as a digital twin prototype with a focus on data visualisation. The 4D visualisation approach is transferable to operating repositories.

3.4.2.2 Virtual Field Trip

We adapted the VEIS to be applicable as a virtual field trip for intermediates, defined as university students or interested members of civil society. To facilitate this, several additional features were added to the VEIS.

First, the illustrations of the experiments were extended. Their objectives and methods are less intuitive for intermediates than for experts. Therefore, illustrations were added with more contextual information

to introduce the users to the general idea and setup of the experiments in a more descriptive and vivid way. An example for such an illustration is the introduction to the FE experiment. The contextual information that was added covers the purpose of the heaters, their power output, and positioning within the experiment. This information helps intermediates to understand the simulation results and how they have been interpreted.

Second, tasks were added to six of the tour's viewpoints. A viewpoint consists of a predefined perspective (i.e., position and orientation of the virtual camera), an audio or text comment, the visibilities of data sets, and possible interactions. The tasks were designed as adjectives based on the learning objectives defined in Anderson and Krathwohl's taxonomy [67]. Users are motivated to "remember" facts about URLs, supported in "understanding" the significance of URLs' geological context, asked to "apply" context knowledge from the presentation, enabled to "analyse" results of numerical simulations, and invited to "evaluate" the effects of radioactive waste disposal on the host rocks. For example, one task is the exploration of the tunnel system's position and orientation relative to the surrounding stratigraphic layers. This motivates the users to visually combine information from two different spatial data sets to actively "understand" the geological structure and the types of rocks surrounding the tunnel system. Another task is to explore numerical simulation results of a determined *in situ* experiment that are available in the tour. In this case, users are asked to identify the time step when a simulated quantity reaches its maximum. In this way they are encouraged to not only passively view the processes but to "analyse" the data to obtain a better understanding. The solution to all tasks is included in the prototype and can be accessed on demand by clicking a button.

Third, the application's graphical user interface (GUI) was improved by adding controls for the playback of audio comments, the display of text comments, and access to the tasks and their solutions. This supports participants in attending the tour at their own pace as all content can be accessed repeatedly and audio comments can be paused if necessary.

Fourth, digital assessments were set up so that learning outcomes can be tested while the users attend the tour. These tests are designed as inline tests, i.e., they follow the corresponding block of viewpoints immediately. The test infrastructure is implemented in a dynamic way, which means that tests are not hard-coded parts of the application but can be provided as structured information in JSON file format, which are then read by the application in runtime. This allows teaching experts or science communicators to define their own tests based on the context of usage without changing the application itself.

Evaluation Procedure and Participants

In advance of the actual evaluation, a heuristic evaluation was conducted to identify and resolve major issues with use of the first prototype. An expert for user experience design analysed the prototype and categorised their findings based on Nielsen's heuristics [68]. In ten categories of usability matters, issues were assigned a severity rating from 1 (lowest severity) to 4 (highest severity). We resolved all issues with a severity of 2 or higher in a second version of the virtual tour prototype and discussed the adjustments with the user experience design expert again.

The virtual tour was then evaluated concerning its usability and learning outcomes at two different courses at master level at two universities: in "Numerical Methods in Geotechnical Engineering" at the TU Bergakademie Freiberg and in "Ground Water" at TU Dresden. Overall, 22 participants took part in the evaluation, 13 in Dresden and nine in Freiberg. Fourteen of the participants identified as male and 8 as female. At both universities, four female participants attended. The age of the participants was in the range from 21 to 32 years with a median of 24.5 years. The average age in Dresden was approximately 27 years while the participants in Freiberg were much younger with an average of 22 years. One participant stated that they have a red-green deficiency. None of the participants had detailed prior knowledge on the Mont Terri URL. 14 participants chose to attend the English version and 8 chose the German version. Attending the virtual field trip as well as the evaluation was voluntary. The participants were informed before the event that their data would be processed in anonymised form and

that they could request the deletion of their data at any time. The results of the knowledge tests were not considered for their course marks.

Evaluation of Usability

The perceived usability was measured using UMUX-LITE [69], which is a reduced version of the usability metric for user experience (UMUX). The questionnaire contains two items, that are answered on a seven-point-Likert scale. Its shortness is helpful to avoid overstressing the participants with too many questions. [70] showed that UMUX-LITE scores can be used to calculate an approximated score for the system usability scale (SUS) [71] as "the correlation between the SUS and UMUX-LITE [is] significant and substantial". Because SUS is an established questionnaire, it has been extensively analysed. Its scores range from 0 (worst) to 100 (best) and can be categorised into more intuitive categories from D (worst) to A+ (best) or using the attributes "okay", "good", "excellent", and "best imaginable" [72]. The evaluation proved the application's good perceived usability, as the median SUS score is 74.36. This value corresponds to the mark "B" and the attribute "good".

Evaluation of the Learning Outcomes

The learning outcomes were measured with multiple-choice tests, in which exactly one of the three possible answers is correct. There was no time limit for completing the tests and it was possible to not answer a question if a participant was not sure of the correct answer. For reasons of feasibility, the variety of question types was limited. This limitation could be addressed in future work. The learning outcomes were measured in four tests; one on the URL and three for the experiments shown in detail (e.g., FE experiment). These tests are designed as inline tests, i.e., they follow the corresponding block of viewpoints immediately. Before each test the participants are informed that there is no time limit for the test and that they have the possibility to go back to the viewpoints if they need more time for exploration. Because the contents of these inline tests are very specific, no prior knowledge was assumed.

Figure 3.28 shows a box plot of the virtual tour's perceived usability and the corresponding attributes and grades. One participant failed the test because they answered only 27.27% of the questions correctly. This person is considered an outlier as their result is significantly worse than the box plot's lower fence (approx. 45%). The median score of 86.36% correct answers shows that the group of participants reaches good results in the knowledge test and demonstrates the virtual tour's ability to successfully teach relevant knowledge about the Mont Terri URL in the context of radioactive waste management research.

Conclusion

By extending the VEIS of the Mont Terri URL, an advanced virtual field trip for educational contexts and as a base for multi-party dialogues was created. A strength of the work is the interactive visualisation of actual research data and its context, allowing intermediates to explore the data individually in an active learning process supported by the application's didactic design. The evaluation indicates a good perceived usability of the virtual tour prototype even for participants with low prior knowledge and little experience with 3D applications. Furthermore, the evaluation gives evidence of a significant knowledge transfer provided by the virtual field trip. A use for science communication to other stakeholders within participative processes for transparent monitoring concepts and site selection processes is conceivable and could be a potential base for contributions in EURAD 2.

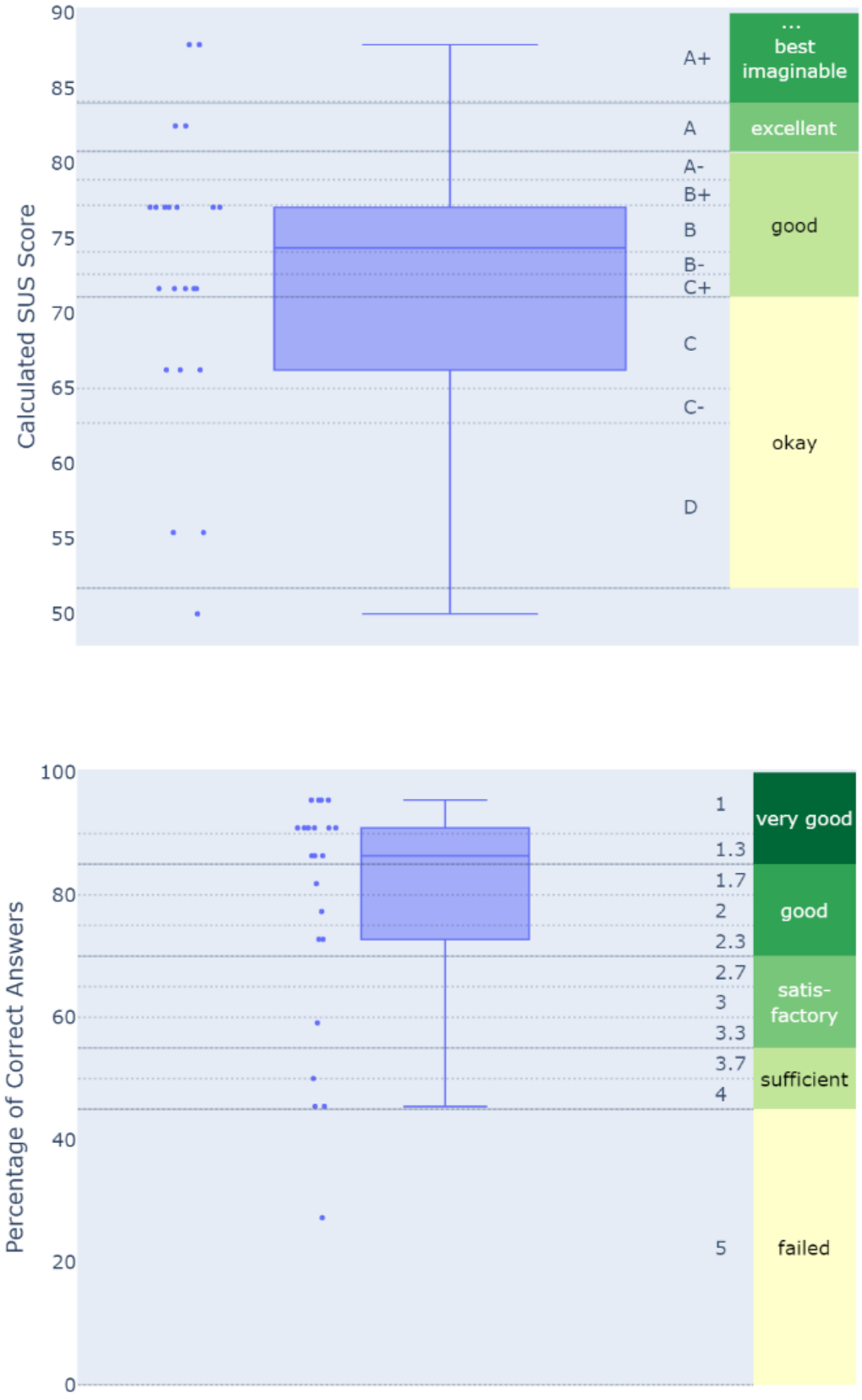


Figure 3.28 - Results of the virtual field trip evaluation concerning the perceived usability (left) and the knowledge transfer (right). The distributions in both charts are common concerning the absence of prior training with the system (usability) and also reflect usual exam results in university education.

3.4.3 Comparison of Simulation and Observation Data

3.4.3.1 Data Acquisition from the FEIS Database and Preprocessing

In order to make monitoring data available for visualisation, we implemented a Python script, that accesses the FEIS database and retrieves temperature measurements for a given list of sensors and dates. These data are then aggregated on a daily base and converted to the VTK file format. In this way, we provide preprocessing infrastructure for making observation data compatible with OGS tools (see Section 3.4.4) and other scientific visualisation software such as ParaView. Furthermore, the use of a common and widely applied data format, such as VTK, improves the robustness of current and future software developments compared to establishing a new and specialised file format used only in the domain. This is because established formats are founded on the technical support of a large number of users and are less likely to be replaced by new formats. The Python script is available under the license GNU GPLv3, and can be accessed online [73].

3.4.3.2 Uncertainty Analysis and Comparison

For the purpose of uncertainty quantification in the context of clay parameter uncertainties at the Mont Terri site in Switzerland, an experimental design-based history-matching workflow was applied to the data of the FE-experiment. The approach had previously been proposed to study uncertainties in coupled modelling of THM processes [74]. The approach discussed [74] is especially suitable for the given problem class, where experimental data can be compared directly to experiments, owing to several aspects discussed in the publication. To highlight here three of them:

- It is very general, i.e., it does not impose specific requirements on the model.
- It combines experimental and modelling responses, allowing also for the validation/study of model uncertainties beside parameter uncertainties.
- A proxy model is used to accelerate direct sampling/global sensitivity analysis, making it feasible to study complex three-dimensional problem.

A schematic sketch of the workflow is depicted in Figure 3.29.

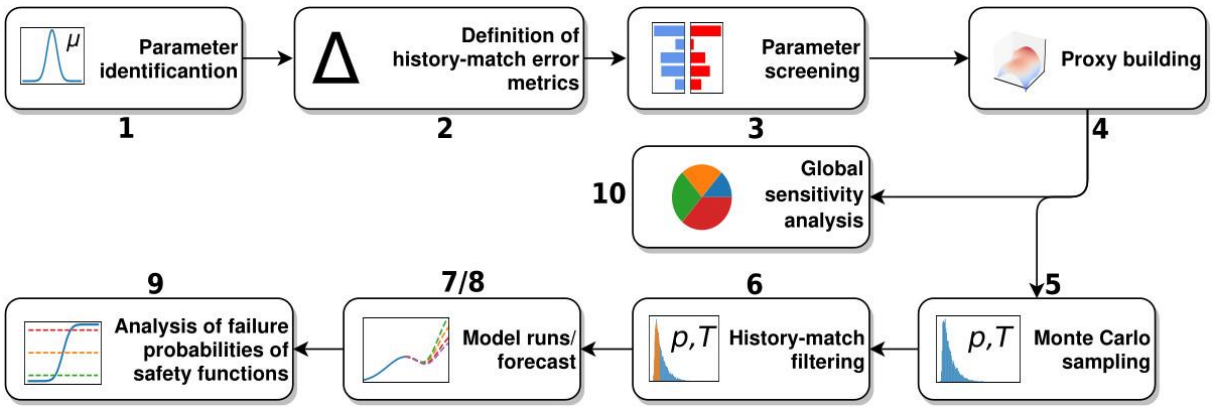


Figure 3.29 - Schematic sketch of the workflow.

The workflow starts with a parameter identification step, in which parameters with their corresponding suggested ranges are selected for the study. Detailed knowledge about parameter distributions is advantageous, as it increases the performance of the workflow. In a second step, the objective function is defined. In our case, it is a history-match error defined by the integrated squared difference between the model and the experiment responses. After that, an initial parameter screening is performed to limit the number of parameters considered in the following steps. All parameters with a significant contribution to the model output are used in a subsequent step to build the proxy model, which is in our case a Gaussian Process regression model. The proxy model is then used to perform a direct Monte Carlo sampling of parameter combinations and/or a global sensitivity analysis.

Based on estimation for model and proxy errors (and experimental errors), a history-match filtering is conducted accounting only for parameter combinations for which its responses are in agreement with the given thresholds. The matched runs can then be used for forward and predictions runs of the full forward model, for parameter estimation or an analysis of failure probabilities of defined safety functions. Observations are made at six distinct points within the Opalinus clay, each at varying distances from the tunnel wall. The history-match error metrics, based on the proxy error, experimental error, and model error, guide the selection of subsamples compatible with the experimental data. Model errors are estimated and thresholds are set for history matching. The model's history-matching capabilities are evaluated by comparing experimental and simulation results for temperature and pressure responses.

As deterministic forward model, we replaced the fully THM-coupled two-phase flow model with a non-isothermal Richard's approximation-based model. The model solves the equations for temperature and pressure as primary variables and contains scalar expressions accounting for the mechanical response in the mass balance (further details can be found in [75]) as well, which is why it is expected that temperature and pressure values to be close to the Richard's-based thermo-hydro-mechanical coupled model. This procedure reduced the number of six primary variables to two, i.e., temperature and pore pressure, with only limited loss in accuracy. While the temperature response exhibits good agreement in the matching procedure, the pressure response shows discrepancies, possibly due to oversimplified models or inaccuracies in material properties. This can be seen from the responses depicted in Figure *3.30*. The modelling curves were generated from parameter sets taken from a folded Placket-Burman screening plan[6]. The colours indicate the history-match error from smallest (blue) to largest (red). Parameter sets that minimise the history-match error of the temperature significantly over- or underestimate (depending on the observation point) the corresponding pressure response.

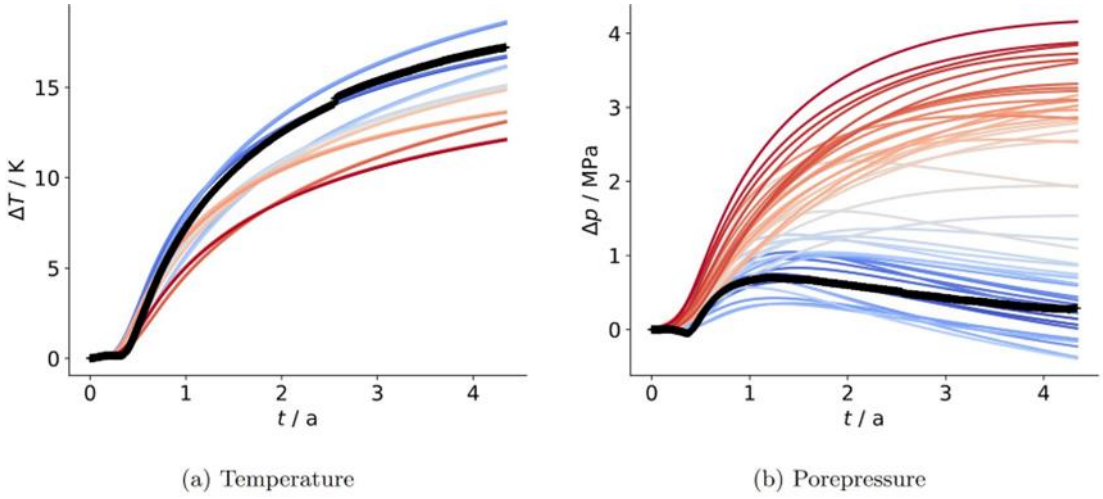


Figure 3.30 - Temperature and pressure responses for one observation point. Black: experimental response. Blue-to-red: Colour coded modelling responses of a folded Placket-Burman screening plan. Blue corresponds to the smallest history-match error, whereas red stands for the largest history-match error. The data is given for one observation point in clay.

Parameter estimation is carried out based on history-matching results and proxy error thresholds. The uncertainty in parameter distributions reveals unique behaviours for certain parameters, particularly regarding parameters that have the greatest impact on model results (e.g., hydraulic and thermal conductivity). Full forward model runs using these parameters highlight differences in pressure responses for different observation points, indicating model inadequacies. Point-based analyses show

[6] A folded Placket-Burman screening plan is an approach to investigating the dependence of a measured quantity on several independent variables (factors), to minimise the variance of the estimate of the measured quantity. Folding is a technique to double the size of plan for a given number of factors.

that parameter preferences vary across observation points, particularly for parameters like porosity and shear modulus. These findings suggest inconsistencies in the model description, potentially related to the representation of bentonite and shotcrete as discrepancies were most pronounced near the tunnel. It is important to note that only clay parameters were treated as uncertain in this study.

In summary, this study contributes to addressing the complex challenges of uncertainty assessment in THM processes for high-level radioactive waste repositories. While certain aspects, like geometrical and process-related uncertainties, are overlooked, this work serves as a foundation for further research and model improvement. We believe that most of the identified discrepancies can be resolved by taking into account uncertainties in the parametrisation of the bentonite and shotcrete material domains.

3.4.4 OGS VisCoSiR

As well as developing tools for processing monitoring data for visualisation purposes, the initial development of a visual tool for the comparative exploration and analysis of simulation data for use alongside OGS was undertaken. Before the development, we interviewed domain experts to gather information about the tasks needed to compare simulation results, their current workflow, and the visualisation methods used. Then, we planned the application based on the results of these interviews. The two use cases for the application are the comparison of results from parameter studies and the comparison of simulation results from different simulation software. The tool consists of two parts:

- A combiner, that allows to spatially and temporally combine (i.e., interpolate) the data sets and calculate metrics (like deviation).
- An analyser that presents the combination result and allows domain experts to explore the differences and similarities in an interactive, visual way.

The results of the combiner are stored in the VTK file format and then loaded into the analyser tool. The separation of the combination and the analysis into two different tools provides flexibility (other tools can be used e.g., for the analysis as well) and efficiency (the combination is only calculated once).

An advantage of a visual tool, compared to applications without a GUI (like Python scripts) or with a complex GUI (like ParaView) is the intuitive preparation and exploration of the data. A screenshot from the analyser tool is shown in Figure 3.31. To avoid overburdening the user, the analyser tool uses visualisation techniques familiar to domain experts (based on the interviews described above). The application provides functionality to visually explore the spatial and temporal data set created by the combiner tool. It supports the user with data summaries and filtering tools for details on demand. The analyser covers, for example, the creation of contour surfaces, the clipping of the data set and the sampling of the data sets at points. The software architecture is designed to be extendible and flexible to allow enhancements of the tool, e.g., by adding other data sources like observation data. In this way, we provide technical infrastructure for a visual in-depth data comparison to enable domain experts to reduce uncertainties by validating simulation results in general.

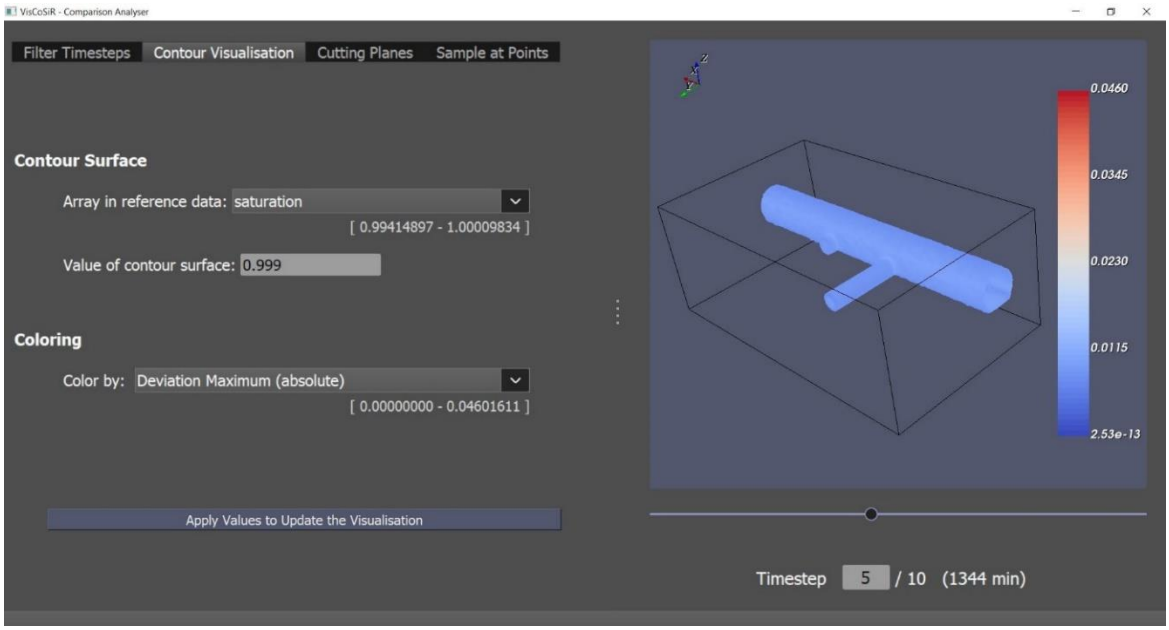


Figure 3.31 - Screenshot from the combiner tool: Users can specify the timesteps used for the combination (grey) based on the two input data sets (blue and pink).

3.4.5 Conclusion

In the context of exploring and validating simulation data, we provided three contributions: 1) preprocessing software infrastructure to convert measurement data into a common data format for visualization, 2) the actual comparison of simulation and observation data for the FE test case, and 3) a generic tool for the visual exploration of simulation results. With these contributions, we not only provided findings for the specific experiment but also generic infrastructure that supports researchers in exploring and validating simulation results.

4. POPLU and Prototype Repository (Posiva, SKB and VTT)

Authors: Arto Laikari, Pirkko Kuusela, Anu Purhonen and Tommi Aihkisalo (VTT)
Editor: Matt White (Galson Sciences)

4.1 Description of the Reference Experiments

4.1.1 POPLU Experiment (Posiva)

The POPLU experiment was implemented at the planned disposal depth of over 400 m in the ONKALO URL, which is located in Olkiluoto/Eurajoki, Finland, during the EC DOPAS project [76]. The objective of the experiment was to test the wedge-shaped design for a deposition tunnel plug. The plug was constructed using steel-reinforced low-pH concrete. A modified tunnel backfill with bentonite was located behind the plug. The dimensions of the plug are approximately 6 m (long) by 5.5 m in diameter. The layout of the experiment is shown in Figure 4.1. The experiment was constructed in 2015 and pressurised in 2016. The plug and part of the backfill are equipped with more than 100 sensors. These sensors include:

- Vibrating wire-based total pressure sensors with temperature measurement.
- Vibrating wire-based pore pressure sensors with temperature measurement.
- Strain gages.
- Thermocouples for temperature measurement.
- Two types of relative humidity sensors with temperature measurement.
- Displacement sensors.

The POPLU experiment also included a small-scale wireless sensor feasibility sub-experiment, where the main purpose was to test the feasibility of using radio waves to transmit sensor data through the rock. Data from this sub-experiment were not used in MODATS.

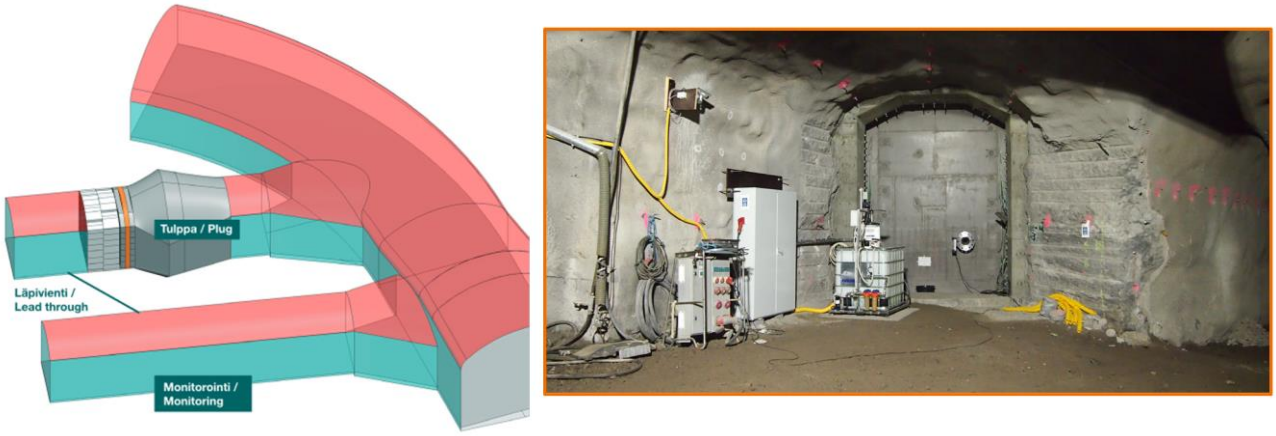


Figure 4.1 – The POPLU experiment layout and photograph.

Monitoring information was collected from the neighbouring tunnel and transferred to the ground level, where further management of the acquired data sets was performed. Monitoring of the plug commenced during the construction phase. After construction, the plug was pressurised by pumping water behind it from the neighbour tunnel.

In addition to the POPLU experiment data sets, Posiva also provided data from monitoring of the geosphere for use in MODATS. These data include:

- Extensometer data sets (temperature, displacement, relative humidity).
- Rock data sets (pressure).

All the data were provided in raw text format (e.g., comma-separated value (CSV) format) directly from the dataloggers. An outline of the data acquisition system is presented in Figure 4.2. In this data acquisition system, various datalogger units with multiple sensors are connected to the local area measurement network (LAN). Data is initially stored locally in network-attached storage (NAS) hardware. An industrial PC acts as a gateway to another network, which allows the data to be transferred to additional hardware placed at the ground level.

Erreur ! Source du renvoi introuvable.

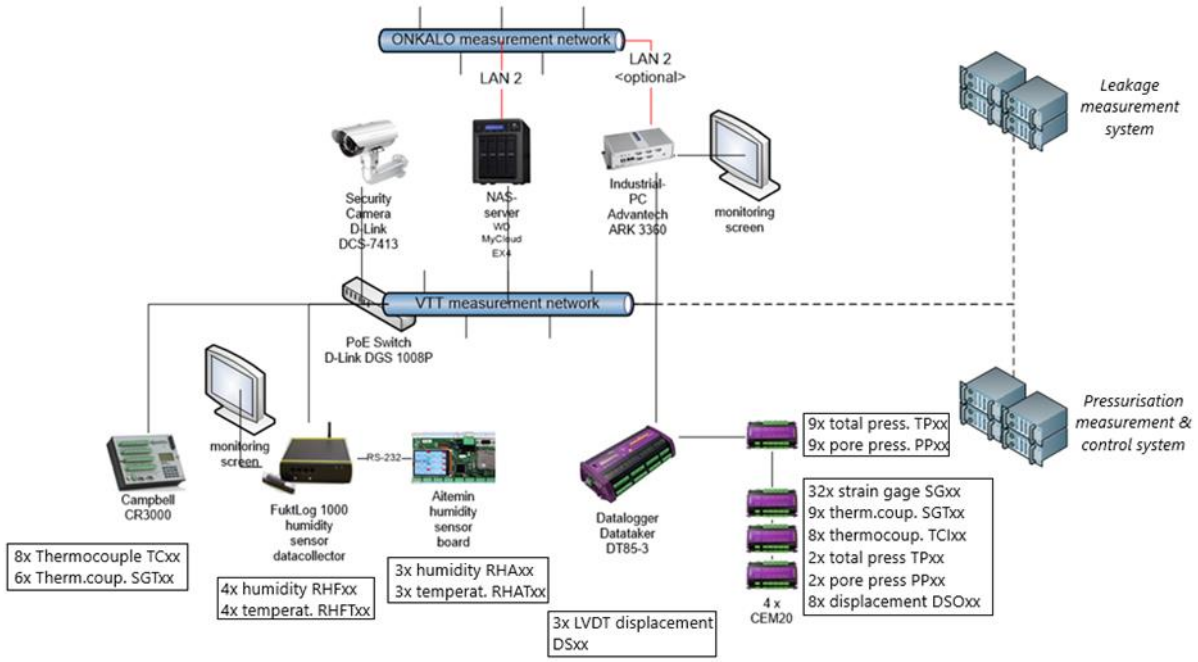


Figure 4.2 - POPLU data acquisition setup.

4.1.2 The Prototype Repository Experiment (SKB)

The Prototype Repository experiment was implemented in the Äspö URL, in Sweden [77]. The large-scale field test includes characterisation of the hydraulic regime in the rock, instrumentation of rock and engineered barriers, design and construction of engineered barriers, and modelling of the performance of these barriers, primarily buffer, backfill and plugs. In the experiment, six full-scale copper-shielded iron canisters with electrical heaters were placed in bentonite buffer. The tunnel was backfilled, and a plug separates the inner section - Section I - from the outer section - Section II. A second plug separates the Prototype Repository from the rest of the Äspö URL. All cables are routed via lead-throughs to the neighbouring tunnel. The Prototype Repository was constructed between 2001 and 2003. The outer part was dismantled in 2011 and inner part was dismantled in 2023. The Prototype Repository layout is presented in Figure 4.3 and Figure 4.4. The experiment was equipped with more than 350 sensors. These sensors include:

- Vibrating wire-based total pressure sensors with temperature measurement.
- Vibrating wire-based pore pressure sensors with temperature measurement.
- Strain gages.
- Thermocouples for temperature measurement.
- Two types of relative humidity sensors with temperature measurement.
- Displacement sensors.
- Water flow sensors.

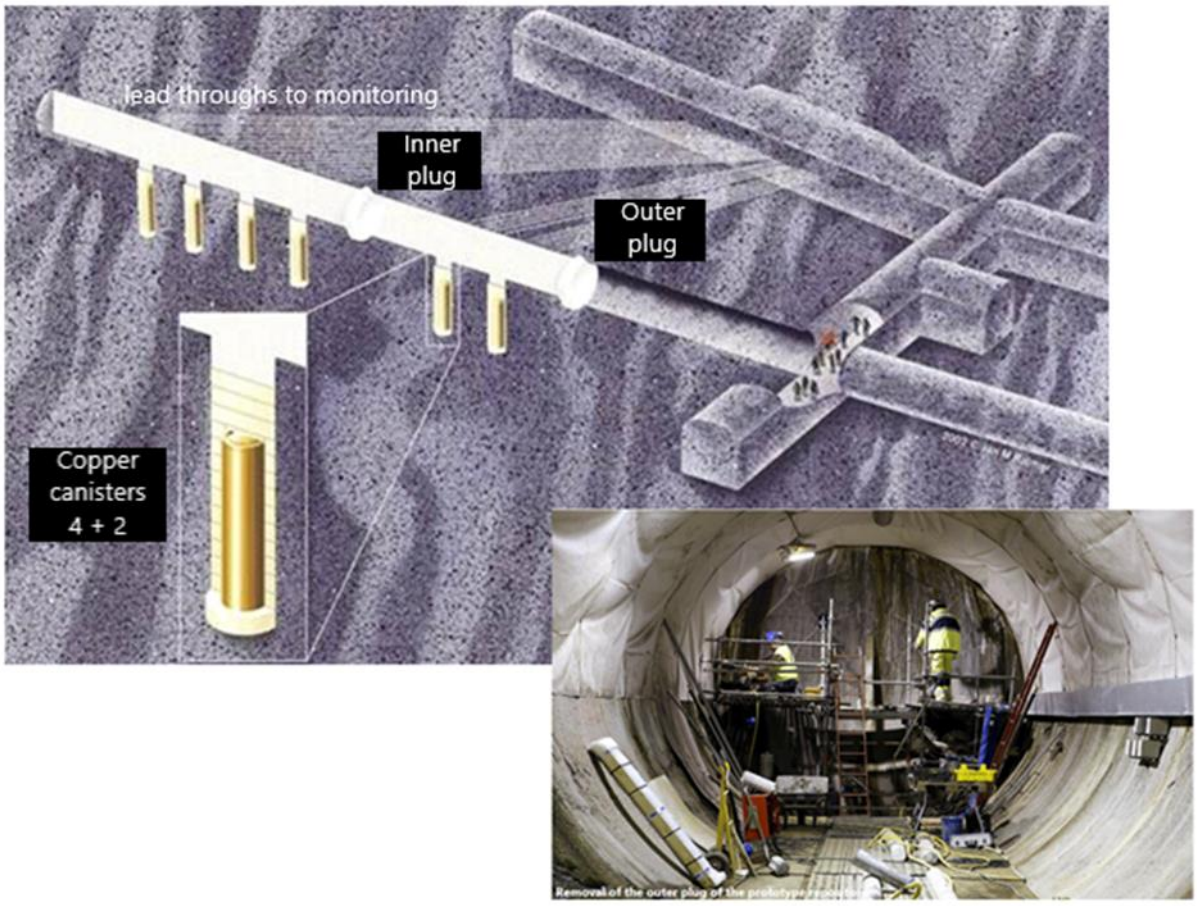


Figure 4.3 – The Prototype Repository experiment and removal of the outer plug (right side image).

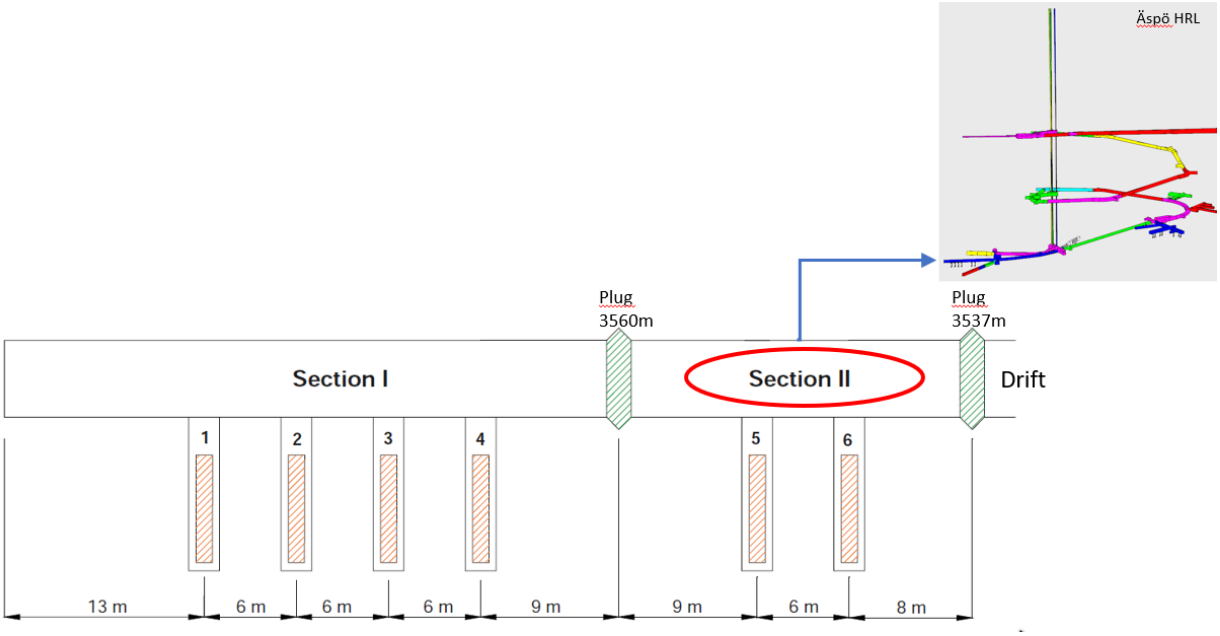


Figure 4.4 - The Prototype Repository experiment location and layout.

SKB provided the following data sets for use in MODATS:

- Experimental location/layout.
- Sensor coordinates.
- Sensor data from backfill and deposition holes.
- Sensor data from rock boreholes.
- Sensor data from canisters.
- Plug 1 and 2 leakage data.
- Key activities in the experiment.

These data sets were provided in manually preprocessed Excel-format or Word-/RTF-files. In the analysing phase of the received data sets, mind maps from the data were created.

4.2 Test Case Objectives and Approach

The objective of the POPLU and Prototype Repository test case was to create modular tools to support data use and management work especially in the data processing phase. These tools are expected to be useful to domain experts, for example, by accelerating their work and by offering new methods to analyse data.

The work was undertaken by VTT, a TSO without direct access to the WMO databases. As such, the work was undertaken with selected data sets chosen by the WMOs and taken from the real data environments.

4.2.1 Approach

As a starting point, representatives from Posiva and SKB supplied data sets from the POPLU and Prototype Repository experiments described in Section 4.1. These stationary data sets were reviewed and data management approaches considered. The data sets were from different stages of the data management lifecycle, so they required different handling procedures.

To get deeper understanding and guide the work, various experts from Posiva and SKB were interviewed to identify common data management procedures, and expectations concerning the test case from the perspective of the WMOs.

To support the work, literature reviews were performed to find applicable standards and data conventions from the radwaste and nuclear domain, especially concerning the metadata definitions of the data.

Various methods and libraries for processing and visualisation of the data sets were tested. This testing resulted in selection of Python libraries and Jupyter Notebooks for the work in MODATS (see discussion of these tools in Section 4.3.1, as they are open-source products and widely used in the data processing domain.

4.3 Data Treatment and Management

4.3.1 Preliminary Studies

A general dataflow framework from data acquisition to decision support is illustrated in Figure 4.5. This figure also shows the stage that the data sets provided by Posiva and SKB had reached in the data flow chain. Data from the POPLU experiment were raw data provided from various measurement devices; they were provided in multiple file formats. Data from the Prototype Repository experiment had been manually preprocessed; they were provided as daily averages of the sensor values, mainly in Excel-format.

Initial work on the case study focused on gathering background information on three fronts:

- Interviews with Posiva and SKB personnel were undertaken to learn about the current approach to data treatment and management within each organisation. The interviews considered data similar to the POPLU and Prototype Repository data sets.

- Analysis of the POPLU and Prototype Repository data sets.
- A review of approaches to data management in nuclear waste repositories to develop recommendations and identify good practice.

The aim of the interviews with Posiva and SKB staff was to identify current data management tools and current workflows, as well as challenges to, and opportunities presented by, use of the sensor data. The interviews were held with research coordinators/leaders, domain experts, data quality experts, information technology experts, and research data management experts. Tool demonstrations by the data user provided clear and detailed understanding of the current workflows used by the WMOs. The main finding from the interviews was that tools for processing this kind of data were rather limited (in both the scope of the tools and the number of tools available). Nor did the interviewed persons refer or compare to any more advanced data operations environment for some other data source.

The main tool used for data processing is plotting sensor measurements as a function of time. Plotting different types of sensors (e.g., temperature and pressure) on the same graph can lead to a loss of detail owing to the different ranges of sensor values.

In a typical workflow, monitoring data from various sensors is taken to Excel for processing and to produce suitable plots for reports. Excel was also the main tool for checking data quality and cleaning. This relates to the fact that both experiments have been established and started a long time ago and at that time modern data processing tools were not available.

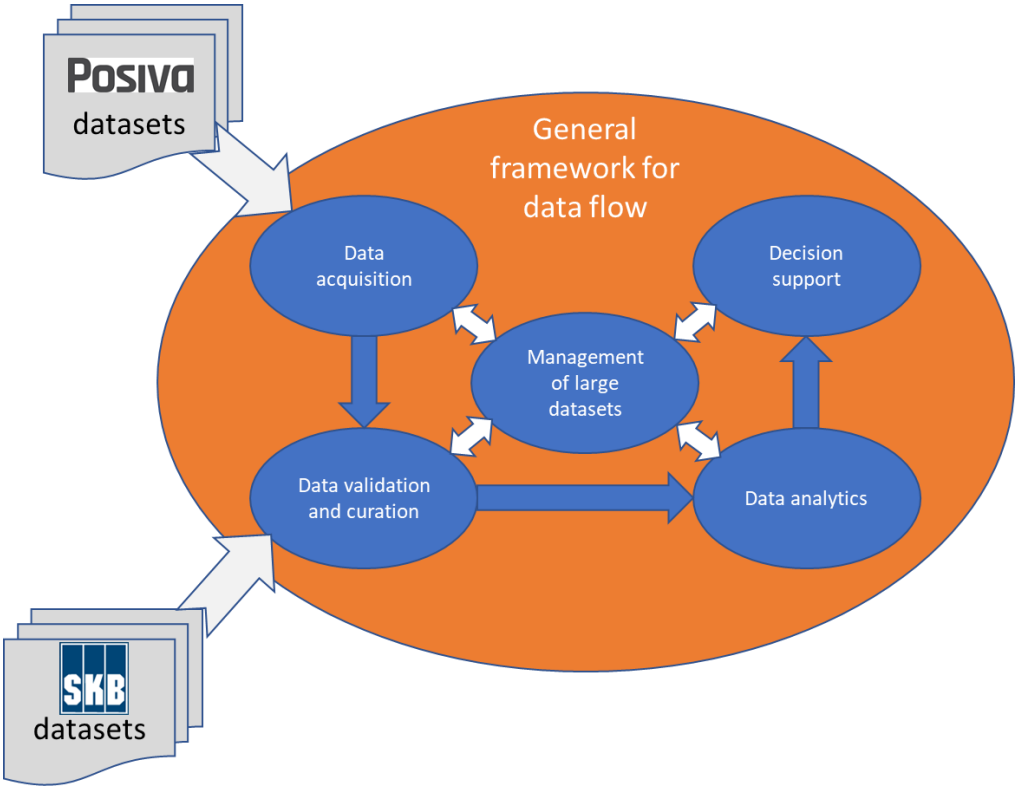


Figure 4.5 – General high-level data management framework, with the data flow stage of received Posiva and SKB data sets in MODATS indicated.

Posiva and SKB have implemented their own data management frameworks. Information on these systems is proprietary and company confidential, and was therefore not available for MODATS. Large number of experiments have been conducted over a long period and they have been implemented by various actors, resulting in company-specific and partly also experiment-specific metadata and data

treatment solutions. However, data sharing wishes and needs are growing, which has resulted in an increased need for data interoperability and harmonised information exchange.

Based on the interviews and the review of the data sets, it was concluded that practical tools for data treatment and management would provide better transparency and traceability than Excel. New practical tools could also facilitate documentation of the treatment and management activities, and should not require users to have skills in programming. Excel offers poor support for documenting data operations and finding an alternative tool supporting documentation and templates could improve workflows and knowledge transfer.

In this context, a tool means an algorithm that is implemented as in Python programming language. Python is a high-level, general-purpose programming language and it is a popular tool in data science owing to the wide availability of high-quality algorithms for data treatment and management. Many data environments support Python in data processing and these are usually invisible for the end user. In a research context, an interactive intuitive environment for using the tools is preferred. However, as data processing algorithms are controlled by user defined parameters, these algorithms require examples and guidelines on how to best use them.

It is proposed that WMOs begin to develop data processing tools in a Jupyter Notebook computing environment. Jupyter Notebooks is an open-source web application that allows a user to create and share documents that contain live code, equations, visualisations, and narrative text. Narrative text can be used to provide instructions for use of the live code, or to write conclusions of results obtained by applying code to data. A Jupyter notebook has two components: a front-end web page and a back-end kernel. The front-end web page allows data scientists to enter programming code or text in rectangular "cells." The browser then passes the code to the back-end kernel which runs the code and returns the results.

Preferred workflows can be promoted by developing notebook templates for users to adapt. Such notebooks may produce modified data sets, e.g., data cleaned for further purposes, together with a report on how the data have been manipulated. This improves transparency and traceability of data operations. Results, for example processed data or figures, can be stored in files, and the evaluated notebook can be converted to new formats (e.g., pdf and Word) in order to take results into reports, other documentation or storage. Jupyter, as such, is a simple notebook environment used in higher education and there are plenty of free beginner level courses on its use. The use of Jupyter is not limited to beginners; experienced data scientists also use notebooks owing to its ability to document data operations.

4.3.2 Metadata, Taxonomy, RepMet Initiative and Standards

As a part of the preliminary studies a literature review of the literature on data standards, recommendations, knowledge models (ontologies), and taxonomies was undertaken. The purpose was to identify and apply good practice in data processing during the POPLU and Prototype Repository test cases. In particular, the review looked for vocabularies for data anomalies, recommendations for labelling observed anomalies, and standards for such labelled data, etc.

The review identified the potential use of the outcomes from the Organisation for Economic Co-operation and Development (OECD) Nuclear Energy Agency (NEA) Radioactive Waste Repository Metadata Management (RepMet). RepMet included participation from Andra, Nagra, Posiva and SKB have been participating in this initiative together with other WMOs. The main aim of the RepMet initiative was [78]:

> *"To formulate a consistent set of guiding principles for capturing and generating metadata. This is to enable national programmes to create sets of metadata that can be used to manage their repository data, information and records in a way that is both harmonised internationally and suitable for long-term management and utilisation in safety cases and elsewhere"*

The RepMet initiative produced the following deliverables:

- RepMet/01 – Metadata in Radioactive Waste Management [78] provides an overview of metadata and its application within repository programmes, discusses issues around the implementation of metadata, and outlines the outputs of RepMet and how they may be used. It also provides specific recommendations concerning metadata for WMOs.
- RepMet/02 – Site Characterisation Library [79] deals with data and related metadata that should be considered during the site investigation, leading up to site selection.
- RepMet/03 – Waste Package Library [80] deals with data and related metadata that should be considered for packaged waste and spent nuclear fuel that, after proper treatment and conditioning processes, are ready for final disposal at the repository.
- RepMet/04 – Repository Library [11] deals with data and related metadata relating to the engineered structures and waste acceptance requirements of radioactive waste repositories.
- RepMet/05 – RepMet Tools and Guidelines [81] supports the libraries, providing several tools, methods, guidelines and approaches that were either used in developing the libraries or will be useful for WMOs when adopting and implementing the libraries.

However, the guidelines and recommendations for data available from RepMet are high-level, and not focused on practical processing of a given data. Because specification of taxonomies, metadata and data formats is the responsibility of the data owner, which in this case is the WMO, the work in the POPLU and Prototype Repository test case was more focused towards developing tools for detecting anomalies, trends, spikes, and other events in the data. The RepMet initiative is highlighted here to encourage WMOs to consider applying the RepMet guidelines in their own data environments, which would increase data interoperability and common practices.

4.3.3 Raw data cleaning process

The data cleaning framework is illustrated in Figure 4.6. First, there needs to be a collection of device specific tools to convert non-homogeneous raw data sources into a format unified data. This is the starting point in using the data. Using a transparent and flexible data cleaning pipeline similar to the workflow shown in Figure 4.7, it is possible to generate datasets that are suitable for solving specific problems. It is important to recognise that different problems require different approaches in the data cleaning. For example, for some purposes data need to be cleaned of anomalies (i.e., the anomalous data removed), whilst in other cases anomalies need to be found and highlighted from the datasets.

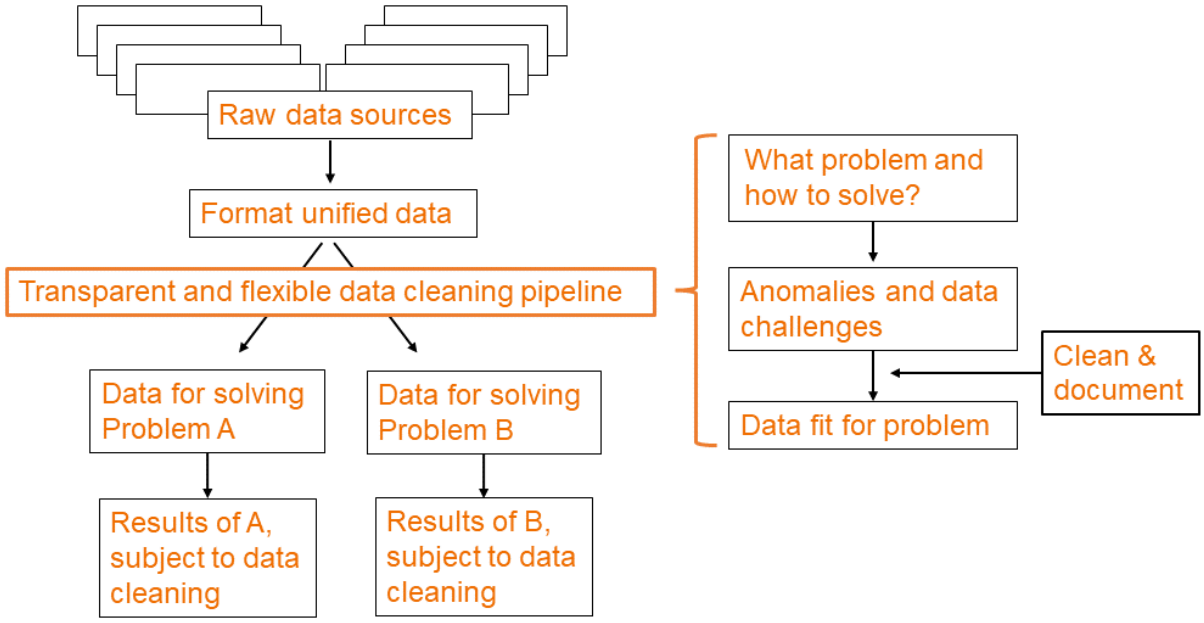


Figure 4.6 The framework of using sensor data.

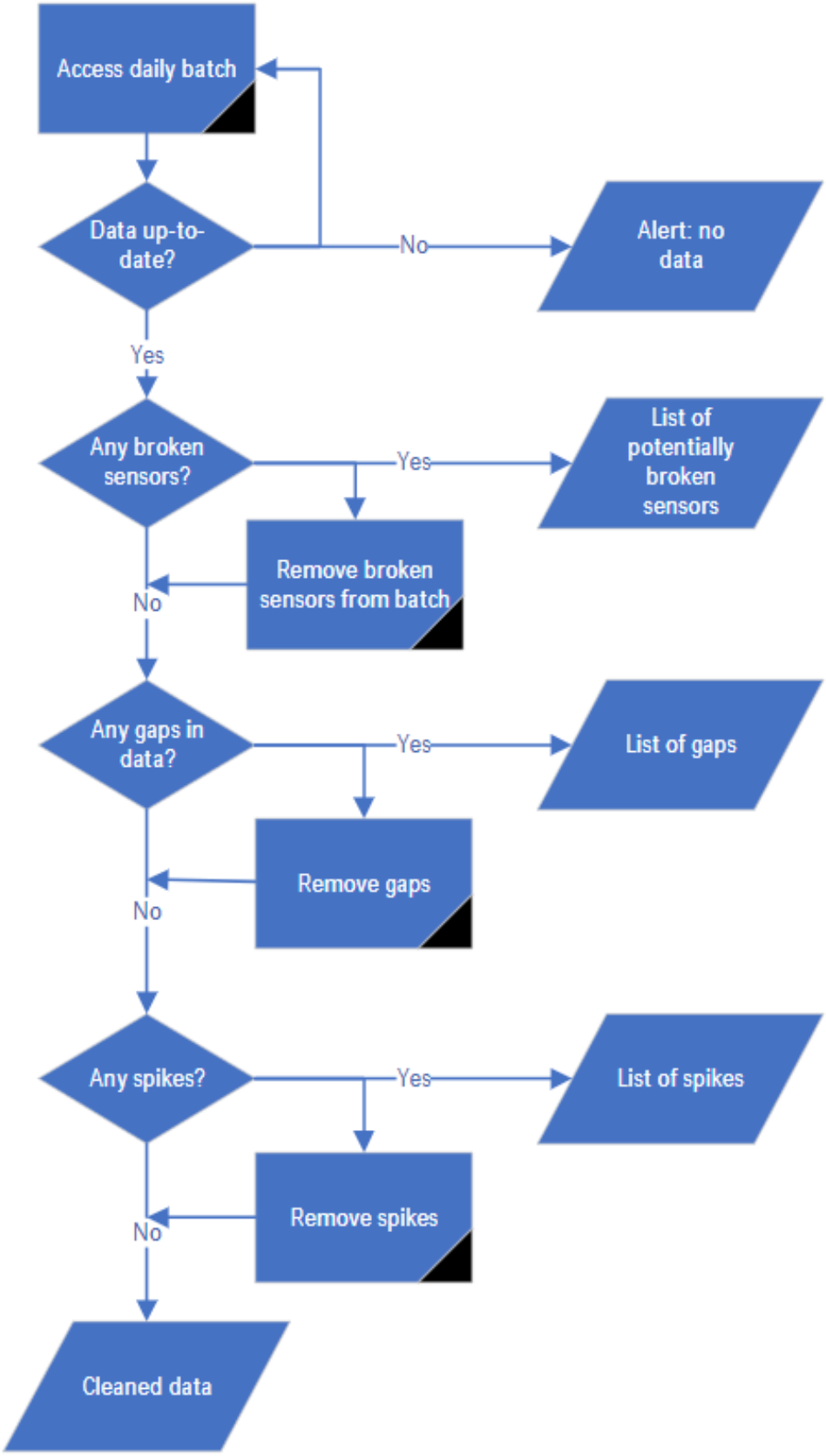


Figure 4.7 Conceptual data cleaning pipeline for daily processing.

4.3.3.1 Preprocessing

From the technical point of view, raw sensor data requires processing before it can be used. The first step in this processing is to reformat the data from different sensors into a common format and to unify the data. The goal is to maintain all of the original information and prepare the data for further processing.

As noted above, the data provided from the POPLU experiment was in a raw data format. It had the following characteristics:

- It was constructed from hundreds of files, the format of which was dependent on the data source.
- The data included duplicate values for the same datapoint.
- Data files could include values from multiple types of sensors.
- Different units were used for parameter values from the same type of sensor.
- The data included gaps and anomalies such as spikes or other deviations from expected values. Anomalies can be caused by known incidents such as electrical disturbances or they can be phenomena that need further study.

To be able to use the Python tools for processing the data were formatted and combined into one file per data source. A CSV file format was used for the combined file. Timestamps were converted to Epoch time[7] and duplicate rows were removed. For traceability purposes, the original source file of each row was identified in an additional data column.

As noted above, the data provided from the Prototype Repository experiment were already preprocessed and were provided in Excel-format. Such data can be processed using *pandas* (a Python Data Analysis Library) if the data are stored in CSV format. Because the preprocessing of the Prototype Repository was already done manually by SKB, data files preprocessing mainly involved modification of the timestamp column, i.e., the name and format of it, and after that the data can be fed into other tools for further analysis and processing.

4.3.3.2 Cleaning

Whereas the goal of data preprocessing is unification of the data, the goal of cleaning is twofold. First, the goal is to prepare the data for the specific analysis that is going to be performed. Second, the cleaning should give an overview of the quality of the data set. Consequently, there is not just one cleaning pipeline, but the cleaning tasks should be configured based on the needs of the user. Figure 4.7 presents an outline of the modular data cleaning pipeline developed for cleaning a daily batch of data. Similarly, a larger set of history data can be cleaned for example, to be used as training data for specific analysis models.

A cleaning pipeline may contain basic or more advanced operations. For example, typically the algorithms provided by the open-source Python libraries do not process data with gaps successfully, so the basic cleaning contains interpolation of short gaps in addition to selection of the measurements that are to be analysed from the data files. If the source data that are used have varying sampling rates, then resampling of data is needed. However, however the data is cleaned, the process used should be transparent i.e., the data points that have been modified have to be traceable to be distinguished from the non-modified data.

A modular data cleaning pipeline is proposed to provide a flexible approach, depending on the end user needs, for example:

- Cleaning pipeline:
  - Data report and cleaned data.
- Example use cases:

[7] Epoch time traditionally corresponds to 0 hours, 0 minutes, and 0 seconds (00:00:00) Coordinated Universal Time (UTC) on a specific date, which varies on different computer systems. The date and time in a computer are determined according to the number of seconds or clock ticks that have elapsed since the defined epoch for that computer or platform.

- Visualisations (remove spikes).
- Checking/labelling data (all anomalies important).
- Study sensor problems (spikes important).

Some of the developed tools may be both the goal of the analysis and intervening step for cleaning data for further analysis (e.g., identify spikes for visualisation or identify spikes before removing them).

In addition to basic cleaning functionalities, the main tools that are related to identification of anomalies and quality of data set include:

- Trend detection and removal (especially required before outlier detection).
- Outlier detection of single timeseries, outlier detection of groups of timeseries.
- Detecting spikes in timeseries (time range, height); removing spikes from timeseries.

Anomaly detection in general requires expert knowledge to what is considered as an anomaly. Spike detection, especially, requires identification of what kind of spikes are looked for. Depending on scale, something that seems like a spike in a yearly plot may be actually a change of values that lasts for several days, On the other hand, quickly oscillating data may be detected as set of spikes depending on what parameters are given for the spike detection. In POPLU data, spikes very often go up in just one data point and then slowly go back to previous values. Depending on the type of the sensor and reason of the spike returning to original value may take from minutes to days, these kinds of spikes are typically difficult to remove as the existing removing algorithms usually expect that the spikes behave symmetrically.

Python code and the Jupyter notebooks developed in MODATS are summarised in Section 4.6 and listed in Table 6.2.

4.4 Data Analysis

Several data analysis tools have been developed, experimented and applied to the data sets from the POPLU and Prototype Repository experiments. Data cleaning tools developed and presented in the previous section can be used for various data analysis use cases as standalone tools for quick testing, but for normal use this is often not sufficient. The requirement is that all operations on data are documented and approved by authorised persons when the data is used and archived. Figure 4.8 outlines the data analysis phase. To make this process easy for the user, Jupyter notebook templates are developed to be used to automate the analysis and documentation process. Using these notebooks, responsible persons approving the curated datasets can see transparently, what modifications has been done for the data and the approval can be also recorded in the notebook.

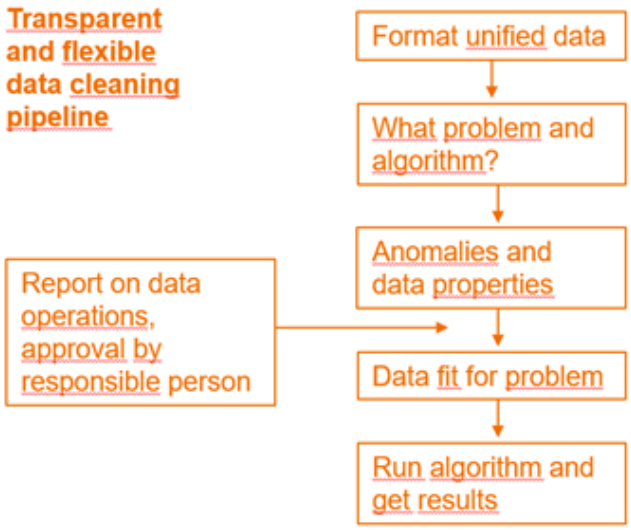


Figure 4.8 – Data analysis phase outline.

4.5 Case Study Examples

4.5.1 Analysis case study: Did the POPLU experiment have an impact on nearby rock?

The case study question of POPLU experiments having an impact on the nearby rock is a causality problem which cannot be solved with available observational data. Ideally, we should have two similar rocks to compare; one without any potential disturbances and one potentially disturbed by POPLU experiments. However, we formulated a general data-based problem:" Are rock sensor measurements different before and after POPLU experiments?" In order to study the rock data before and after POPLU experiments, we present here as an example results of Principal Component Analysis (PCA). PCA is practical basic tool in data analysis and easy to perform by domain experts with a template notebook for code and explanations.

We selected 5 pressure sensors from Posiva pressure hole PH22 with good sensor quality and no clear change in behaviour after POPLU experiments when sensor values were plotted as a function of time. PCA looks at the shape of the cloud produced by these sensors in a 5-dimensional space. We examined how this cloud of points evolves in time. This corresponds to studying the mutual relationships of these pressure sensors. PCA provides new coordinates where the first principal component (PC) is the direction of the largest variance in the cloud. Each subsequent PC is orthogonal to the previous ones and describe a decreasing amount of variance in the data. Typically, first two to three components are of interest and the last ones describe only noise.

The main PCA result of PH22 is illustrated in Figure 4.9. PCA is also a dimension reduction method as plotting the data in the coordinates of the first two PCs covers in this case already 95% of the variation in the 5-dimensional data. Data in blue-green shades and red-yellow corresponds to time before and after POPLU experiments, respectively. Time evolution, illustrated by shade change, shows that qualitative behaviour in After data is different from the hovering in the Before data. Inspection of how different sensors contribute to the PCs also provides information on the nature of PCs and how sensors produce information. PCA was performed in such a way that the long-term trend of pressure sensors did not impact the result.

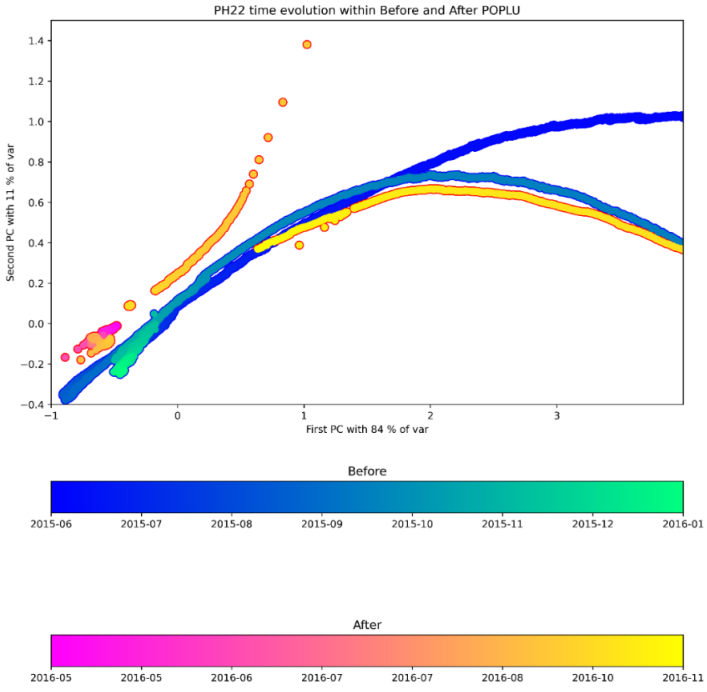


Figure 4.9 – Illustration of a change in pressure relationships before and after POPLU experiments detected by principal component analysis.

In terms of repository monitoring the outcome of the PCA poses needs for further studies and input from domain experts. Before this data analysis the conclusion has been that POPLU experiments are not visible in the nearby rock data. At this point we have a data analysis indication that the joint behaviour of 5 pressure sensors is qualitatively different after POPLU experiments, at least for some duration after the experiments. Unfortunately, the sensor data time range and quality does not allow to study if the "hovering" behaviour returns after a transient behaviour. The further questions include:

- What is the normal behaviour of a similar rock under no possible disturbances? Compare PCA results of the two rocks. Is there normal annual variation in pressure?
- Could there be an explanation that is unrelated to POPLU experiments? Or, is this an impact of POPLU construction work and the rock is in a transient phase before new equilibrium?

At this stage, we can conclude that even simple, easily understandable and transparent data analysis methods can provide new insight into the data and their regular use may improve understanding of the repository phenomena (natural, intended actions or unintended consequences).

4.5.2 Analysis case study: Cleaning pipeline example utilising Jupyter notebook

The Jupyter notebook *Cleaning_pipeline.ipynb* presents an example on how different functionalities in the cleaning.py can be used to clean a dataset for potential later use. In this case, it is assumed that data is cleaned for modelling purposes. The phases in the workflow include:

1. Overview of dataset with visualisation of relevant sensors. Before visualisation any additional information is removed, for example, raw data files contain information where the data was initially originated.
2. Detailed gap analysis (*get_length_gaps*). Detailed gap analysis reveals the location and length of the gaps in timeseries. Gaps mean no information, so for modelling purposes it should be found a time frame where the gaps are minimised.
3. Selection of time frame. The most representative time frame is selected. The gaps that are left are removed with interpolation.
4. Detection and removal of spikes (*identify_spikes*, *remove_baseline*, *remove_spikes*). The specified type of spike is detected and removed from selected timeseries. Remove_spikes use timeseries where the baseline has been removed as input.
5. Visualisation of results.

In all the phases, the intermediate results can be stored for traceability purposes.

4.5.3 Analysis case study: Detecting anomalies

In this example (*Anomaly_detection.ipynb*), *detect_outliers* and *detect_outliers_group* methods from *cleaning.py* are used for identifying anomalies in a dataset. Anomalies can be identified either for each variable separately or for variables as a group. The purpose of the example is to show potential solution for detecting them, so the actual sources of the anomalies in the dataset are not known.

Both outlier detection approaches are based on the *LOESS* and *LOF* algorithms. Seasonal-Trend decomposition using *LOESS (STL)* is from *Statsmodels library* [82]. First, *LOESS* is used to remove trend from the original timeseries. After that the *Local Outlier Factor* (LOF) is used to extract outliers from the residual. *LOF* algorithm from *Scikit-learn library* [83] is an unsupervised anomaly detection method. The number of detected outliers depends on the given parameters.

In Figure 4.10 the sensors are treated as a group and the outlier detection notifies when there are exceptional events in those sensors as a group. The timeslots where the outlier is found is marked as - one marker in the outliers graph bottom line.

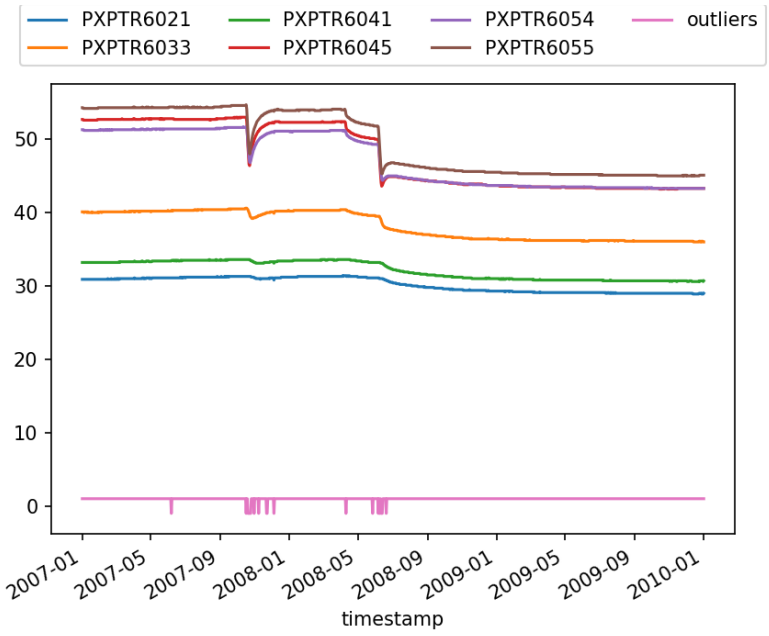


Figure 4.10 Detecting outliers from sensors as a group.

In Figure 4.11 only one sensor is studied and the visualisation shows the original timeseries, the extracted trend and residuals as well as the detected outliers in subplots.

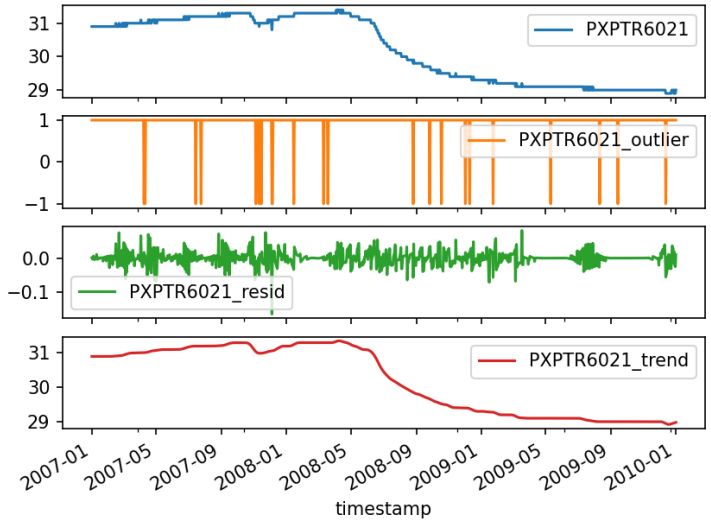


Figure 4.11 Detecting trend, residual, and outliers from a sensor.

4.5.4 Analysis case study: Detecting spikes in timeseries

Spike detection and removal (*Spikes_detection_removal.ipynb*) shows in more detail how spikes can be detected and removed. It uses *identify_spikes*, *remove_baseline*, and *remove_spikes* methods from cleaning.py. First, *identify_spikes* returns the locations of the highest peaks, they are shown as dots in Figure 4.12. Some of the dots may hide other dots under them when the sensors are behaving similarly. Consequently, this figure may reveal events that have affected several sensors. Method *remove_spikes* looks for the start and end of the spikes and returns the cleaned timeseries. Instead of the original timeseries it uses a timeseries where baseline has been removed as input. Result of one of the sensors

after spike removal is visualised in Figure 4.13. The original timeseries is shown in blue, red dots reveal the locations of peaks, and the cleaned timeseries is in orange. In Figure 4.14 one cleaned spike is shown in more detail. As can be seen in removal of the spikes there is still room for improvement in finding the start and end of the spike.

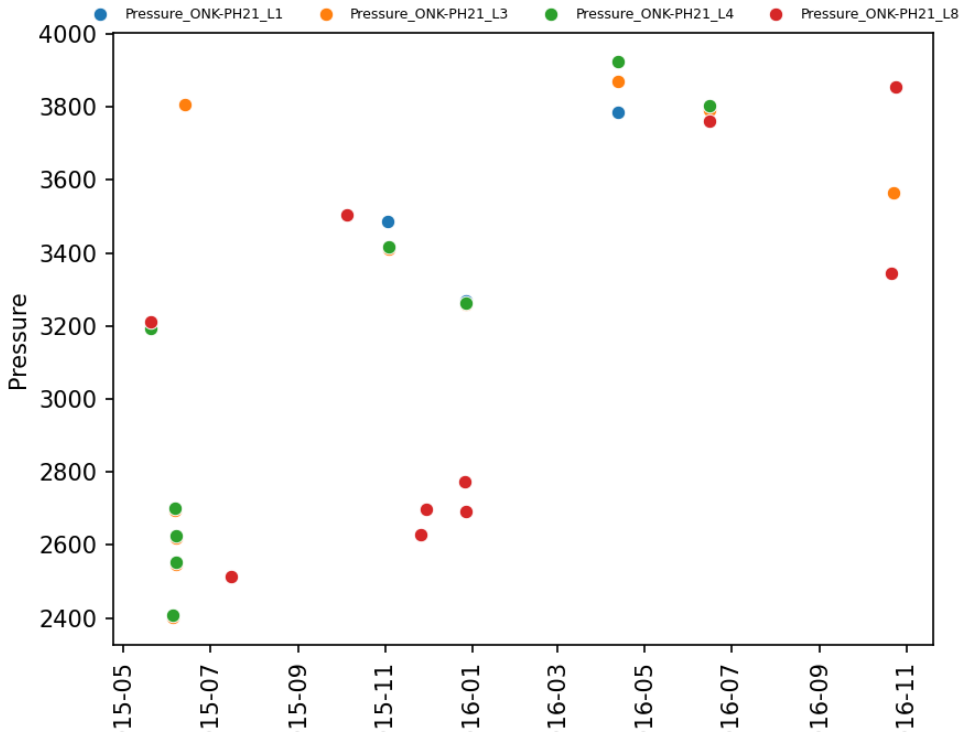


Figure 4.12 Locations of the peaks of detected spikes.

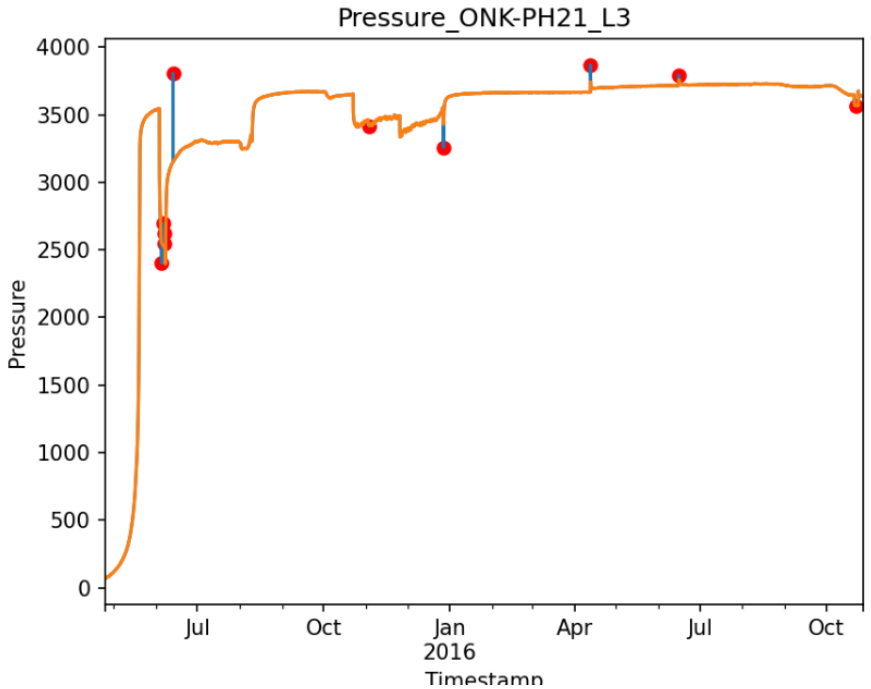


Figure 4.13 Detected and removed spikes for one sensor.

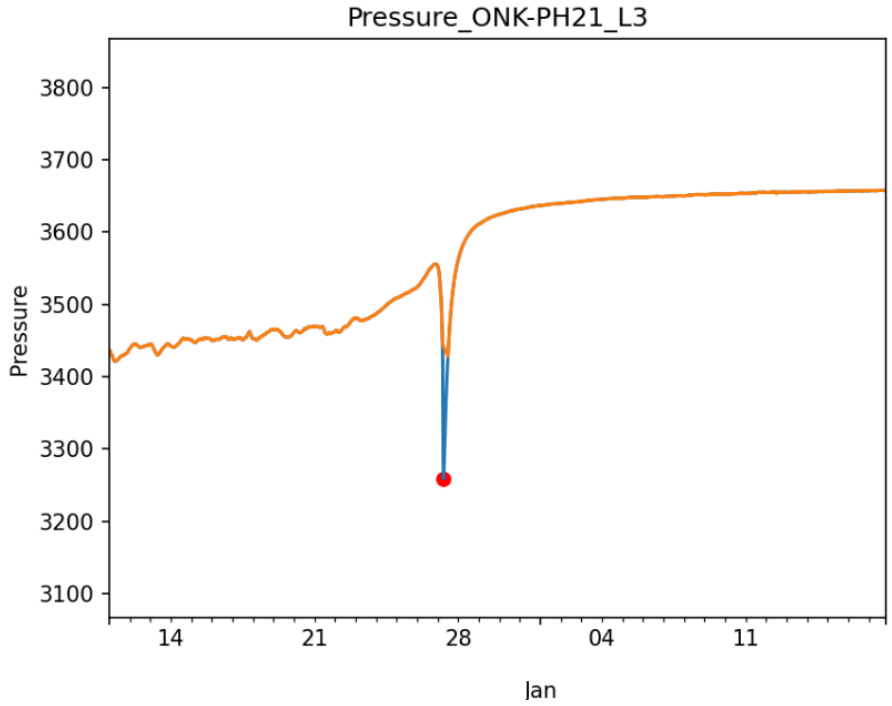


Figure 4.14 One spike zoomed.

4.5.5 Analysis case study: A tool for detecting sensor drift.

Sensor drift can be difficult to spot especially when the drifting happens slowly. The proposed approach uses nearby sensors to reveal potential drift in one or more sensors. It is assumed that in the long run the behaviour of nearby sensors correlates so in case there are changes in that correlation this may be caused by drifting. This case study is implemented in *Drifting.ipynb* and uses *detect_correlation_change* method from *cleaning.py*.

The setup of the case study includes a dataset from SKB Prototype Repository. The dataset is divided into original dataset and a test dataset (Figure 4.15 a and b). A small artificial drift was added into values of one sensor (PXPTR6021) in the test dataset. The results are shown as heatmaps in Figure 4.15 the darker the colour the lower the correlation. Figure 4.15c shows the correlation in the original dataset between included sensors, Figure 4.15d the test dataset, and Figure 4.15e the difference between those correlations. The correlation of the sensor PXPTR6021 with the other sensors is in the original dataset quite high with two of the other sensors but in the test dataset it becomes low as is expected. However, with comparing what was the situation earlier we can confirm that there may be a drift in that sensor. Induced artificial drift in this experiment was linear, but also non-linear drifts were added and tested and detected with this approach.

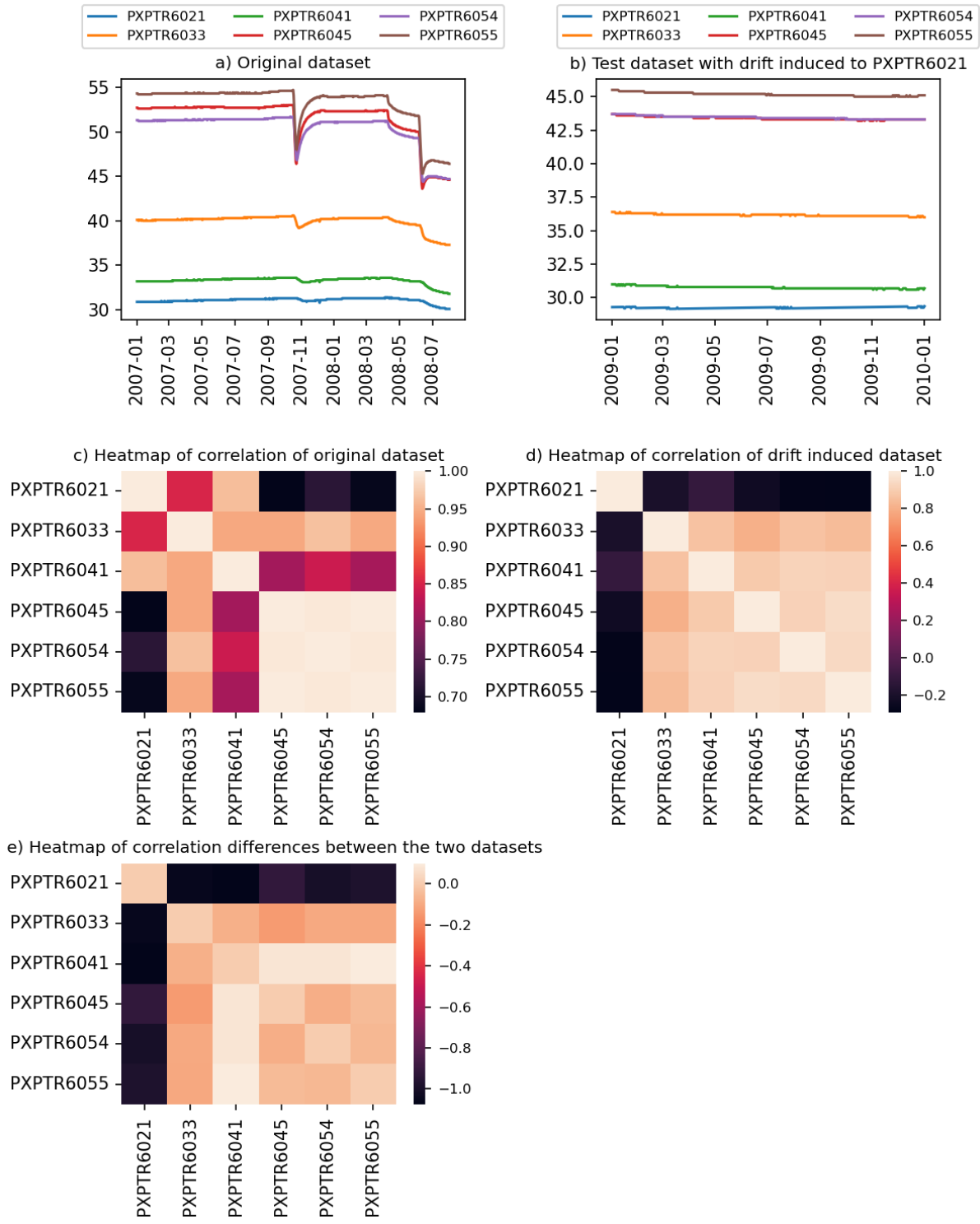


Figure 4.15 – Detecting sensor drift in SKB dataset.

4.7 Summary of Developed Tools

The developed tools are summarised in Table 6.2. Developed tools consist of the Python code modules and the Jupyter notebooks.

Code modules are divided into tools for preprocessing and formatting POPLU raw data and the cleaning module. Further, formatting is performed with source-specific code modules for each POPLU data source as they all have different formats. Formatting first changes the file-format as csv and preprocess-codes further combine the multiple raw files as one source-specific file.

The goal of cleaning is to prepare the used datasets for a specific purpose. Functionality in the cleaning.py can be used to configure different types of cleaning pipelines depending on the need of the user. The methods in the cleaning module have been used for both POPLU data and Prototype Repository data.

Python libraries that have been applied in the tools include *Pandas* [84] (McKinney, 2010) and *NumPy* [85] (Harris et al., 2020 for basic timeseries processing and *Scikit-learn* (Pedregosa et al., 2011), *SciPy* [86] (Virtanen et al., 2020), *statsmodels* (Seabold and Perktold 2010) and *BaselineRemoval* [87] (Haque 2022) in outlier and spike detection.

Jupyter notebooks in Table 6.2 are used as examples of how cleaning functionalities can be used in cleaning and analysis. In addition, there are examples on rock data analysis. Jupyter notebooks can be used both as templates and documentation for making specific types of analysis with changing input data and parameters.

4.8 Summary and Conclusions

Reliable data treatment and analysis requires lot of additional information, not just the acquired sensor values. Working with real-life sensor data poses questions on correct interpretation of the data, data anomalies, impacts of external events and correlations. Optimally datasets of several sensors need to be treated and analysed together making cross checking and adding information about known other events, even if at the time impacts are not seen in measured values. As data anomalies are many times due to other activities in the repository, applying domain knowledge of the whole environment and of the used sensors as well as data acquisition setup is also beneficial to ensure correct interpretations in further data analysis phase. In this documented work, discussions on the data, work processes and problems indicated such interdependencies and complexities of working with the data. However, only sensor data was available and used in these studies, when developing the tools and methods.

The study of the rock data prompted for data that can serve as a “control” in contrast to a “treatment/intervention”. If data is collected only during an experiment, we are unaware of typical uninterrupted phenomena. Thus, it would be important to start collecting data well before any intervention, e.g., POPLU or Prototype repository type of experiment, to establish well the initial state, baseline of the environment. Similarly, data collection should last long enough after the experiment to establish the final state. If possible, data collection could go on to record the normal uninterrupted state. Not all data needs to be processed or analysed, keeping data for future purposes has also value. Despite of the current artificial intelligence (AI) hype, also more serious AI is developing at a fast phase. As the tools and methods are evolving, new analysis can be performed also in the old datasets to find new phenomena.

URL experiments established in the past have provided data from the experiments. The tools to handle or analyse the data have been very limited. One of the most important tools has been the Microsoft Excel. Data curation and processing has required substantial amount of manual work. The work performed with the old data sets in MODATS has shown that all raw data is valuable to be kept and stored for later usage. As the tools and methods are evolving during time, new analysis can be performed also with the old data sets to find new phenomena and data events.

New data science concepts brought into repository data management frameworks can provide modular open-source tools to enhance and improve data treatment and analyses. These tools can be used to:

- Prepare data for further analysis.
- Give an overview of the quality of data.
- Be configurable to different kinds of needs.
- Allow traceability for what is cleaned and why.
- Automate laborious and monotonous tasks.

Benefits of the open-source tools increase if WMOs make some example data available for the research community to experiment the method and tool development. Further, from the research community point of view unified data format and meta data greatly facilitate the data usage. Instead of sharing larger amounts of repository monitoring data, WMOs could consider allowing computations on their proprietary data while ensuring data confidentiality.

Generic high-level common practices and standards initiated by international actors like IAEA or OECD NEA can lower the threshold to agree and adopt common standards to exchange data and information between WMO domain actors, a noteworthy example is the OECD NEA RepMet libraries, which outlines high level guidelines e.g. for the metadata definitions. Large actor group cocreated work can have faster larger impact and is easier to be agreed than a single actor proposal.

During the interviews as well as earlier work with the WMOs has also shown that confidentiality and information security versus openness need to be considered as in the nuclear field there are more tight restrictions concerning data and communication than in other domains.

5. PRACLAY (EURIDICE)

Authors: Jan Verstricht (EURIDICE)
Editor: Matt White (Galson Sciences)

5.1 Introduction

The HADES URL, managed by EURIDICE, has been operational since the early 1980's. Since then, the experimental data obtained have been managed by different database systems that allowed for basic retrieval, reporting and graphing of the sensor data. Further analysis of the data has usually been performed by the user, based on their specific needs (e.g., to follow-up experimental performance, and to compare experiment monitoring data with model predictions). Such analysis typically requires some basic data cleaning as the original data sets usually contain data that are not representative of the phenomena investigated owing to different reasons, ranging from data acquisition issues, power failures, planned interventions or accidental events. Owing to the lack of a uniform approach with each user typically undertaking their own data cleaning work, overlapping work has been undertaken and different data treatment methodologies followed. Furthermore, no standard approach has been available for data validation, which is particularly important for the supply of data to external users.

As part of the MODATS data management task, EURIDICE has investigated the development of a uniform data cleaning and validation process.

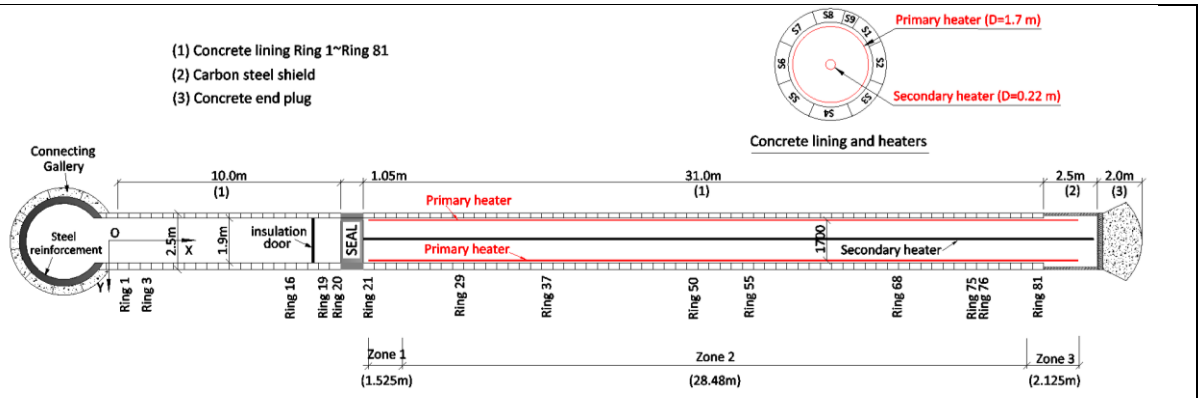
5.2 Description of the Reference Experiment

The data sets used in this work are generated from selected sensors used to monitor the PRACLAY Heater Test [88]. The experimental setup is designed to investigate the impact of a large-scale thermal load on the Boom Clay, which leads to perturbations in the clay that might affect its performance as a host rock. The Heater Test is intended to be representative of a generic disposal gallery for heat-emitting waste and was conceived to be as independent as possible of the final repository design and to be conducted under a well-controlled and a reasonably conservative combination of thermal, hydraulic and mechanical boundary conditions.

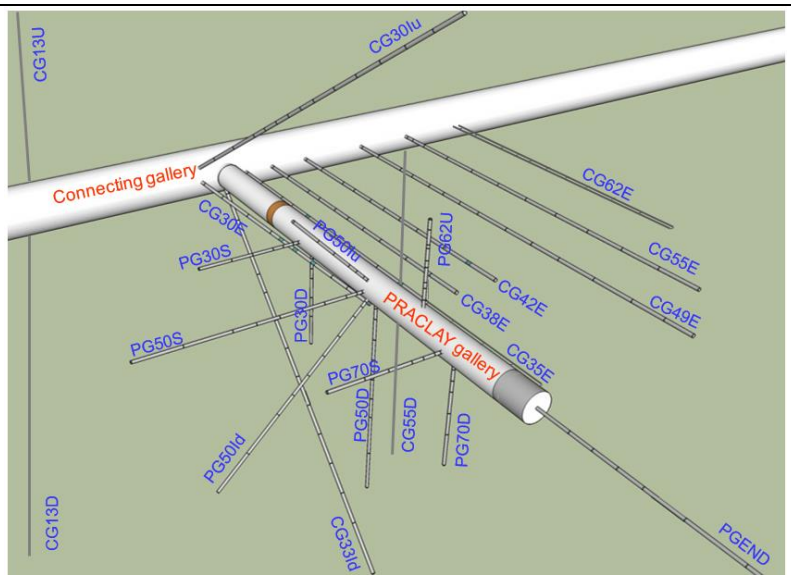
The test consists of a gallery (Figure 5.1a) in the clay host rock, accompanied by a network of observation boreholes (Figure 5.1b). Boreholes drilled and instrumented from the Connecting Gallery (CG) were drilled prior to the excavation and lining of the PRACLAY Gallery (PG). From the PG, additional instrumented boreholes were installed to extend the observed zone, in particular in the vertical direction.

The main test phases associated with the PRACLAY Heater Test are listed below in chronological order:

- 2001-2002: excavation of CG.
- 2006: Installation of instrumented boreholes from the CG.
- Oct – Nov 2007: Construction of PG.
- 2008 – 2009: Installation of instrumented boreholes from the PG.
- 2011-2014: Backfilling of the active PG with sand, installation of a seal for closure of the PG, saturation and pressurisation.
- Nov 2014: Start heating.
- Aug 2015: Stationary heating phase; heating with a constant temperature of 80°C on the inside of the PG lining.



(a) Setup in the PG



(b) Observation boreholes instrumented with pore pressure and temperature sensors around PG

Figure 5.1 - General view of the PRACLAY experimental setup in HADES URL.

5.3 Objectives

The objective of our work is to obtain a structured and uniform management of the PRACLAY experimental data. This includes several sub-objectives:

- Elaboration of a data cleaning and validation process that is more efficient, uniform and transparent than the range of approaches used to date, and results in validated data sets.
- Development and implementation of the supporting tools for:
  - Management of metadata.
  - Data visualisation (to support data validation).
  - Reporting of experimental data.
- A database structure supporting these functionalities.
- The generation of validated data files adapted to the needs of the different users, including:
  - Experimental follow-up.
  - Data modellers.
  - External expert users (e.g., other research organisations and PhD students).
  - Citizen stakeholders.

5.4 Methods

The development of the data treatment and management approach is based on a selected sub-set of the PRACLAY experimental data. These data consist of time series sensor signal values. For most sensors used in PRACLAY the signal values are stored in the main database. A range of sensors were selected for inclusion in this test case based on the following considerations:

- Inclusion of a range of sensor types.
- Inclusion of a range of data acquisition equipment.
- Inclusion of data with variable signal qualities such as noise and resolution.
- Inclusion of data containing various anomalies (e.g., artefacts, thermal disturbance, and sudden or gradual sensor failure)
- The relevance of the sensor data to analysis and comparison with model predictions (considered to be the primary use of monitoring data).

Within MODATS, the original data files were also uploaded to the Zenodo platform for optional use by the other teams within Task 2.

The actual process of data cleaning and validation consists of a machine-assisted screening of the sensor data, highlighting (labelling) the data of interest (e.g., anomalies, transients, outliers), followed by an expert assessment of the data. The expert assessment also uses metadata about the sensor (e.g., particular knowledge on the sensor behaviour) and about the experimental events (e.g., logbook information on incidents that have occurred). This might also lead to a future improved management of these metadata (e.g., improved logbook format).

To test the current ideas on data cleaning and validation, the related algorithms are coded in a Python integrated development environment. When the tests are successful (i.e., the algorithms can be applied to the real data sets), the code will be integrated in the LabTool environment. This is a database, which, for our purposes, acts as a laboratory information management system (LIMS), as it contains the metadata from the sensors. Recently, the sensor measurement data in the main database are being linked with the sensor definitions in LabTool, so that data retrieval and processing can be performed within this environment.

5.5 PRACLAY Data Set

5.5.1 Selection Criteria

To limit the amount of data for this exploratory research, whilst covering a range of challenges for monitoring data treatment and management, a selection of the PRACLAY data was made according to the following criteria:

- Different data acquisition systems: Most of the data has been acquired through automated data acquisition systems. Different systems are in use, resulting in varying characteristics, such as data resolution.
- Sensors with different performance: For different reasons, but mostly related to the harsh environmental conditions, a significant number of sensors have failed since their installation. Sometimes this failure did not happen suddenly, but was preceded by a gradual degradation of the sensor performance, which requires also an appropriate interpretation of the monitoring data.
- Different data quality: Sensor data quality characteristics such as signal-to-noise ratio, or more general signal stability, might also require additional data processing to allow a proper interpretation and reporting.
- Presence of various events: Sensor data with typical anomalies, such as those resulting from manipulation (e.g., calibration, sensor replacement), physical tests (e.g., permeability tests, geochemical sampling), artefacts (impact of unexpected thermal process on measurement of

temperature and pore pressure), and sensor failure, are selected to test the robustness of the cleaning and validation methods and algorithms.

- Scientific relevance: To include the feedback from the data users, data that have been mainly used in scientific analysis of the experiment have been selected.

5.5.2 Proposed Data Set

More than 1000 sensors have been installed in the context of the PRACLAY setup. Based on the selection criteria, porewater pressure and temperature sensors have been selected for the work in MODATS, as these data are most extensively used in the scientific analysis. Most of these sensors have been installed through instrumented boreholes. In addition, porewater pressures and temperatures are also measured in the PG and its lining, as they give essential data on the source term (simulated disposal gallery with heat source).

All these considerations have resulted in incorporation in the MODATS PRACLAY Data set of the following sensors:

- *Pressure transmitters connected to filters integrated in borehole casings installed from the CG instrumented boreholes*: These data include measurements from pressure transmitters connected to filters integrated in three borehole casings (CG35E, CG42E, CG55E (Figure *5.1*b)). The boreholes are drilled from the CG and are up to 45 m deep. All boreholes are horizontal and perpendicular to the CG axis. The boreholes are cased with a stainless-steel liner. Sintered stainless steel filters are integrated in this liner to collect the porewater of the Boom Clay. Each filter is connected through a twin-tube capillary tubing (inner diameter 2 mm) to a pressure transmitter, which is located in the CG (hence, the capillary tubing length can reach up to 45 m).

  All pressure transmitters included in this data set have a 4-20 mA signal output. The different setups are further read out by three different types of data-acquisitions systems for historical reasons. Depending on the data acquisition system, this current signal is then measured directly, or it is measured through a precision shunt resistance (typically 250 ohm) to convert the current signal into a voltage (in this case 1-5 V).

  The pressure transmitters are the only type of sensors that remain accessible. They are regularly calibrated and hence, accuracy and drift data are available for these sensors.

  Typical deviations in the measurement data are due to manipulations, which are mainly caused by calibration, porewater sampling, and permeability tests. Other simple operations such as disconnections can cause anomalies. Several pressure transmitters have also been replaced to adapt the measurement range to the actual conditions. This has caused a sudden change in the sensor signal, while the calculated signal should remain the same (through an adapted calibration formula).

- *Pressure transmitters connected to filters integrated in borehole casings installed from the PG*: This data set contains the signal values from the pressure transmitters that are connected to the filters integrated in the borehole casings PG30D and PG50D. Owing to problems with their installation (e.g., capillary tubing running along the boreholes open to the heated gallery and next to the heater elements), the pressure measurements are more prone to artefacts, for example those caused by unexpected thermal processes.
- *Thermocouples installed next to the filters of the CG instrumented boreholes*: These data have limited anomalies, but include limited features such as temporary increased noise, or disturbances owing to pumping in geochemical filters (with unstable temperature).
- *Thermocouples installed next to the filters of the PG instrumented boreholes*: The cabling for these thermocouples is routed through the heated gallery, and, as a consequence, these signals are prone to drift, and to sudden or gradual failure.

- *Pressure transmitters connected to filters inside the backfilled PG*: The gallery is backfilled with sand – a high permeable medium in order to create and maintain a source term with a well-defined boundary condition (uniform water pressure over the complete backfilled gallery length). In theory, the seven pressure transmitters should indicate exactly the same value, so these data have been included to allow for cross-comparison of monitoring data.
- *Thermocouples integrated inside the concrete lining of the PG*: Since the start of the heating phase, about half of these sensors have failed after functioning for varying periods; two different data acquisition systems are being used to acquire the sensor signals.
- *Thermocouples attached to the heater cables inside the PG*: These sensors typically detect the periodic switching of the heater cables; about half have failed, and exhibit different trends in the sensor signal prior to failure (sudden or gradual drift, unstable signal).

5.5.3 Current Data Management Practices

Most sensors are connected to data acquisition front ends or data loggers. The measured sensor values are then transferred to the database server. This server stores the monitoring data, undertakes preprocessing (e.g., calculation of engineered data) and makes the data available. Direct consultation of the data currently uses a web-based GUI. This whole system is called Daily, see *Figure 5.2* for the structure of this system.

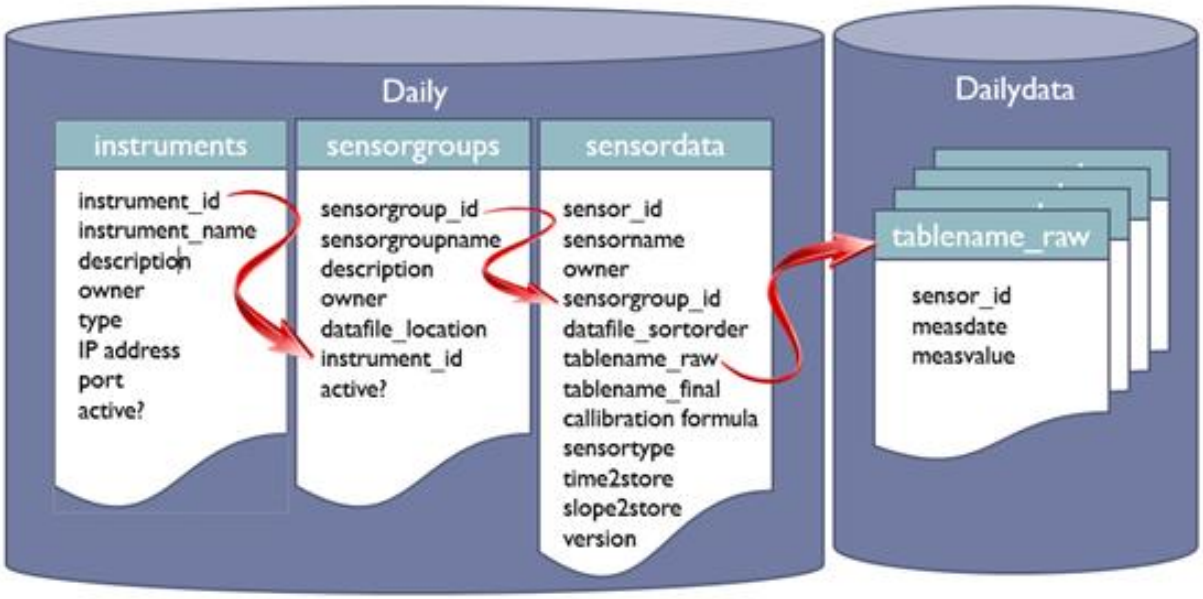


Figure 5.2 - Structure of the Daily database system used for the HADES (and thus PRACLAY) data.

Daily currently collects monitoring data from approximately 3,500 sensors. The data is stored in a central database in standard folders, files and tables. Additional functionalities of Daily are graphing, export and compression (mainly averaging) of data. Further, to assist in the operation of the various experiments being undertaken in HADES, an alarm system is included. It checks if the sensor data are regularly updated, if they do not exceed limits, and if the monitoring devices are still online.

To provide more efficient data management (e.g., periodical automated reporting), commercial software (DiaDEM from National Instruments) has been implemented for the PRACLAY data managed in Daily. In particular, it is used to generate a safety report each day (for a quick daily check of safety-relevant variables), weekly data reports (frontline scientific follow-up) and quarterly reports for the main client. In addition, it allows for customised graphs.

These tools do not incorporate a tool to support the management of monitoring equipment (e.g., sensor inventory, follow-up of calibrations and maintenance). Therefore, the LIMS tool, which is called LabTool, is being developed by the nuclear research centre SCK CEN. In this tool, all kinds of laboratory items

(e.g., sensors, data acquisition equipment, cabinets and integrated setups, and chemicals) can be defined and structured in a hierarchical way. The complete life cycle of the laboratory items (e.g., during commissioning, maintenance, updates, and out-of-service periods) can be recorded.

In parallel, a more geographically based GeoScientific Information System (GSIS) was developed by the Belgian WMO ONDRAF/NIRAS. Currently, it mainly manages all borehole related information, such as survey data (coordinates), borehole logs, core logs, and analyses of borehole filter water samples.

Some overlap between the data from GSIS and LabTool exists (e.g., pressure transmitters). The current idea is to merge LabTool and GSIS into one tool, and to extend it with data cleaning and validation functionalities.

5.5.4 Structure Of Metadata

In addition to the actual sensor signals, a proper interpretation of the signal values requires the knowledge of the metadata linked to this setup. Metadata include sensor specifications, installation records (e.g., sensor coordinates, data acquisition chain) and logbooks listing all events relevant to the setup chronologically.

Currently, most of these data are not structured yet in a uniform way, as it is typically up to the experiment responsible person or principal investigator to keep track of all metadata. Currently, a structure is defined in LabTool in which all metadata will be stored in a more uniform way. This will also serve as a guideline for the persons operating the setup.

5.6 Data validation

5.6.1 Definition

In the context of the operation of an experimental setup, data validation can be defined as the assessment of the measured data (or other observations), by adding an indication (or label) to each data point about its validity. The concept of “validity” is not binary (valid or not valid), but, depending on the user of the data, more labels indicating the validity should be considered. The labels should take into account the interest of different users and state if the data are valid for their purpose or not. For example, sensor signals which are of interest to the instrumentation engineer (e.g., when checking for sensor malfunctions), might not be of interest for a modeller who wants to check a long-term model prediction with the experimental data.

The end result of the data validation process is therefore a data set with these labels. By filtering the labels, the user then obtains the wanted data set. This data set can then be further optimised by e.g., data reduction (e.g., daily averages) or smoothing.

5.6.2 Development of a Data Validation Methodology

To establish a validated data set from the original signal data, the following phases can be distinguished:

- Preparation of the data file (data query / file import) into a standard format.
- Query of the relevant metadata which are usually specific for each sensor type, e.g. the range of sensor signals (including optional tolerances), the physical range – e.g., no negative temperatures in PRACLAY Heater test, and the resolution.
- Screening of the data based on metadata and modelling data to detect deviations from a “normal” evolution:
    - Out of limit sensor signals.
    - Out of physical range.
    - Discontinuities, including value (outliers) and time (missing data).
    - Sudden changes.
    - Transients.

- Expert interaction (principal investigator, data responsible) to confirm and/or add data labels (typically with reference to logbook information):
    - Labelling of sensor performance issues, such as sensor failure and sensor drift.
    - Labelling of intended operations, such as scientific interventions (e.g., sampling and permeability tests), technical maintenance/interventions, and calibrations or other sensor diagnostics.
    - Labelling of other disturbances, such as data acquisition system glitches, power failures, borehole drilling, abnormal temperature variations, and other incidents (e.g., cable or tubing disconnected inadvertently).
- Generation of validated data set:
    - Adapted database structure with one or more label(s) attached to each data point.
    - Dedicated file format for external users.
    - Data treatment (e.g., averaging and/or smoothing) as required.

5.6.3 Preliminary Implementation

Owing to the large amount of experimental data that have been/are being gathered, and as most of these data are acquired in an automated way, it is evident that many tasks of the data validation process should be automated through scripts. Currently, these are developed in a Python integrated development environment, and include:

- Query of Time-Series Data and Metadata: Most sensor time-series data are stored in Daily. A simple query allows the user to load the data into a temporary structure. Further, the metadata – relevant for the data set considered – are read.
- Screening of Time-Series Dataset: Anomalies can result from various events; in order to detect and label them correctly, the specific features of each event should be uniquely characterised. The specific features of the events are then quantified by some efficient indicators of the relevant data such as:
    - Basic statistics (minimum – maximum – average – linear trend) over the related period of the events (e.g., daily or weekly).
    - Noise levels, standard deviation (time window to be defined by expert).
    - Resolution.
    - Gradient over time.

  A dedicated Python script allows this screening to be undertaken, anomalies to be detected, and data to be labelled. The script also produces a visualisation of the data.

- Expert Interaction and Interpretation: The expert (e.g., principal investigator or data scientist) checks the result of the script and confirms, modifies, details, dismisses the labels. As interpretation requires understanding of the experimental conditions, other data sets can be accessed to check for correlations, e.g., the influence of temperature on other sensor data. Also, the logbook will be consulted to explain anomalous behaviour.
- Update of the Database: The result of the previous actions is then confirmed and uploaded in the database.
- Generation of Data Files: From the validated data, data files can be generated according to the needs of the data user by querying with the appropriate selection criteria based on the data labels. Further data treatment can be considered.

5.7 Results

Data cleaning and validation results for the different sensor types are detailed below, illustrating how the Python scripts developed in MODATS address a sub-set of the data challenges identified in Section 5.5.1.

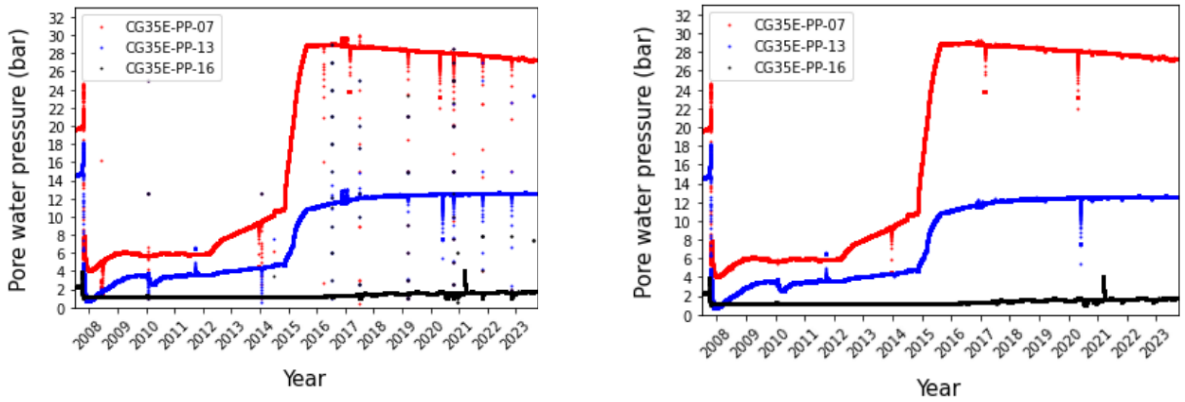
5.7.1 Pressure Transmitter Data from the CG Multifilter Piezometers

As these pressure transmitters are accessible, malfunctions of the transmitters themselves are almost non-existent. Disturbances or anomalies in the data can be attributed to interventions, such as sensor maintenance (calibration, replacement, or simple check of connection between quick-connect ends) and experimental manipulations (permeability tests or porewater sampling). Each event-induced anomaly has specific features which can be quantified by specific indicators of the associated time-series data, for example:

- Permeability tests start with a sudden pressure change followed by a nearly constant pressure (with low standard deviation) for more than 1 day.
- Porewater sampling starts with a sudden pressure drop to atmospheric pressure and remains nearly constant for a couple of hours.
- Sensor calibration can be characterised by a sudden drop to atmospheric pressure followed by quick recovery within an hour.

Figure 5.3 shows the graphs corresponding with porewater pressure data from three sensors located in borehole CG35E before and after a data cleaning and validation process. The initial data contain many anomalies, including outliers, calibration and other interventions. Only the data points resulting from the permeability tests have been maintained in the processed set (at the right), these data points are useful for users who want to derive the permeability of the Boom Clay close to this filter. Depending on the choice of the data user, different anomalies can be kept or excluded from this data set by filtering the labels available in the Screening of Time-Series Dataset script.

Expert judgement is sometimes essential for correct anomaly detection, and this is the case for the time-series data from these three sensors. These sensors are located close to the PG. When the excavation front passed by these sensors, significant pressure changes between two consecutive data points were observed. The magnitude of such change is physically normal. Therefore, the criteria to detect anomalies should be adapted for this excavation phase, otherwise, it is possible these data points are mistaken and labelled as a different event (e.g., calibration).

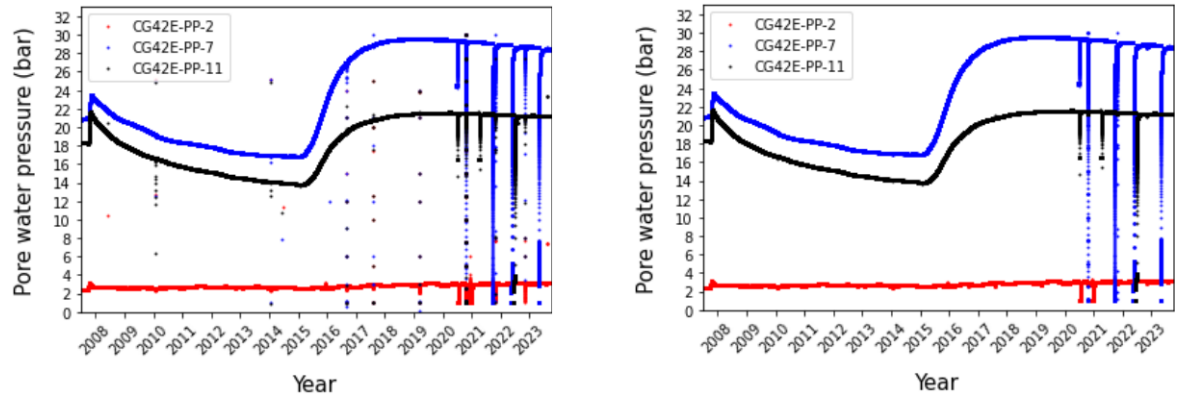


(a) All data without cleaning

(b) Data after cleaning but keeping permeability test data sets

Figure 5.3 - Time-series data from three pressure sensors in CG35E:CG35E-PP-07, CG35E-PP-12, CG35E-PP-16.

Another example is shown in Figure 5.4 for data from the pressure transmitters in borehole CG42E. These data illustrate disturbances owing to regular porewater sampling since 2020. The data points associated with these disturbances during this period are labelled as a separate category (i.e., geochemical sampling), these data are also useful for users who want to derive the Boom Clay permeability.

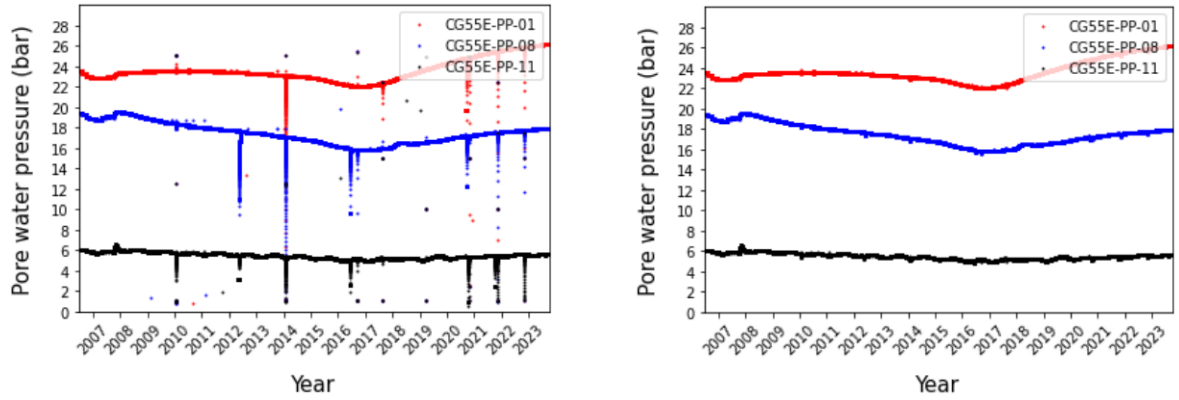


(a) All data without cleaning

(b) Data after cleaning but keeping permeability test and sampling data sets

Figure 5.4 - Time-series data from three pressure sensors in CG42E: CG42E-PP-2, CG42E-PP-7, CG42E-PP-11.

Finally, data from the CG55E pressure transmitters are shown in Figure 5.5. The figure at the right shows the cleaned data set, with all disturbances (including permeability tests, sampling and calibrations) removed, which is generally useful to the majority of end users.



(a) All data without cleaning

(b) Data after cleaning and without anomalies

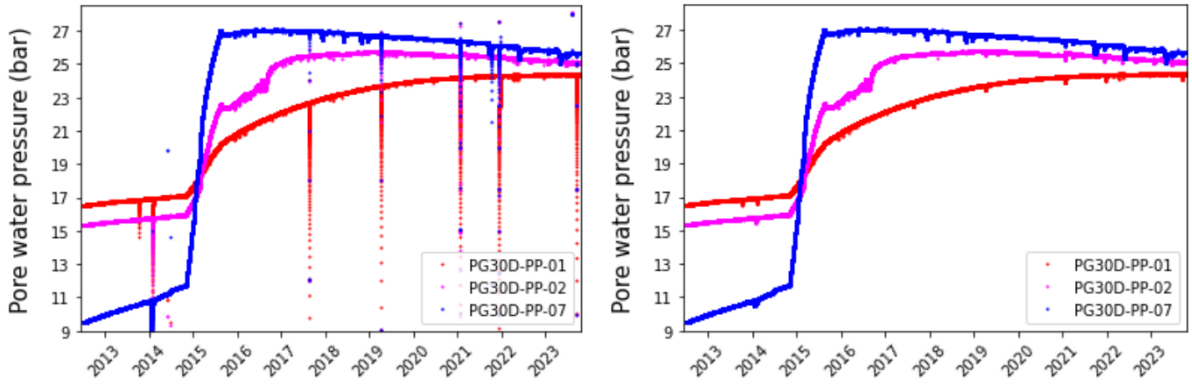
Figure 5.5 - Time-series data from three pressure sensors in CG55E: CG55E-PP-01, CG55E-PP-08, CG55E-PP-11.

5.7.2 Pressure Transmitter Data from the PG Multifilter Piezometers

As pressure transmitters are accessible, malfunctions of the transmitters themselves are almost non-existent, but unexpected and abnormal data points are still observed in the time-series data from some PG borehole filters, which may be explained by the artefacts of the instrumentation setup. The connection tubes between the pressure transmitter and the filter are typically quite long, and with this tubing running along the borehole casing open to the backfilled gallery and in the heated gallery, the unexpected heat transfer in the open borehole (i.e., convection) and sudden temperature variations (e.g., heater failure, or switching of heater cables) can cause complex pressure evolutions and transients. As artefacts are therefore more likely in these data sets, modelling data are therefore essential as an auxiliary tool in the cleaning and validation of measurement data.

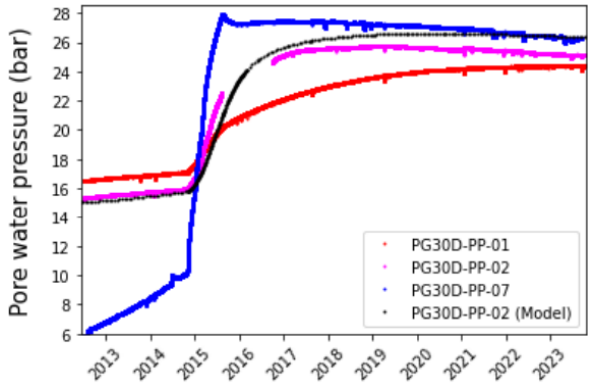
Figure 5.6a shows all data from three transmitters in the PG30D borehole, Figure 5.6b shows the cleaned data by removing data from permeability tests and calibration etc. PG30D-PP-02 shows some

unexpected non-smooth data, which are labelled as doubtful after comparing them with the modelled data (Figure 5.6c), and this segment of data is removed.



(a) All data without cleaning

(b) Data after cleaning but keeping the doubtful data



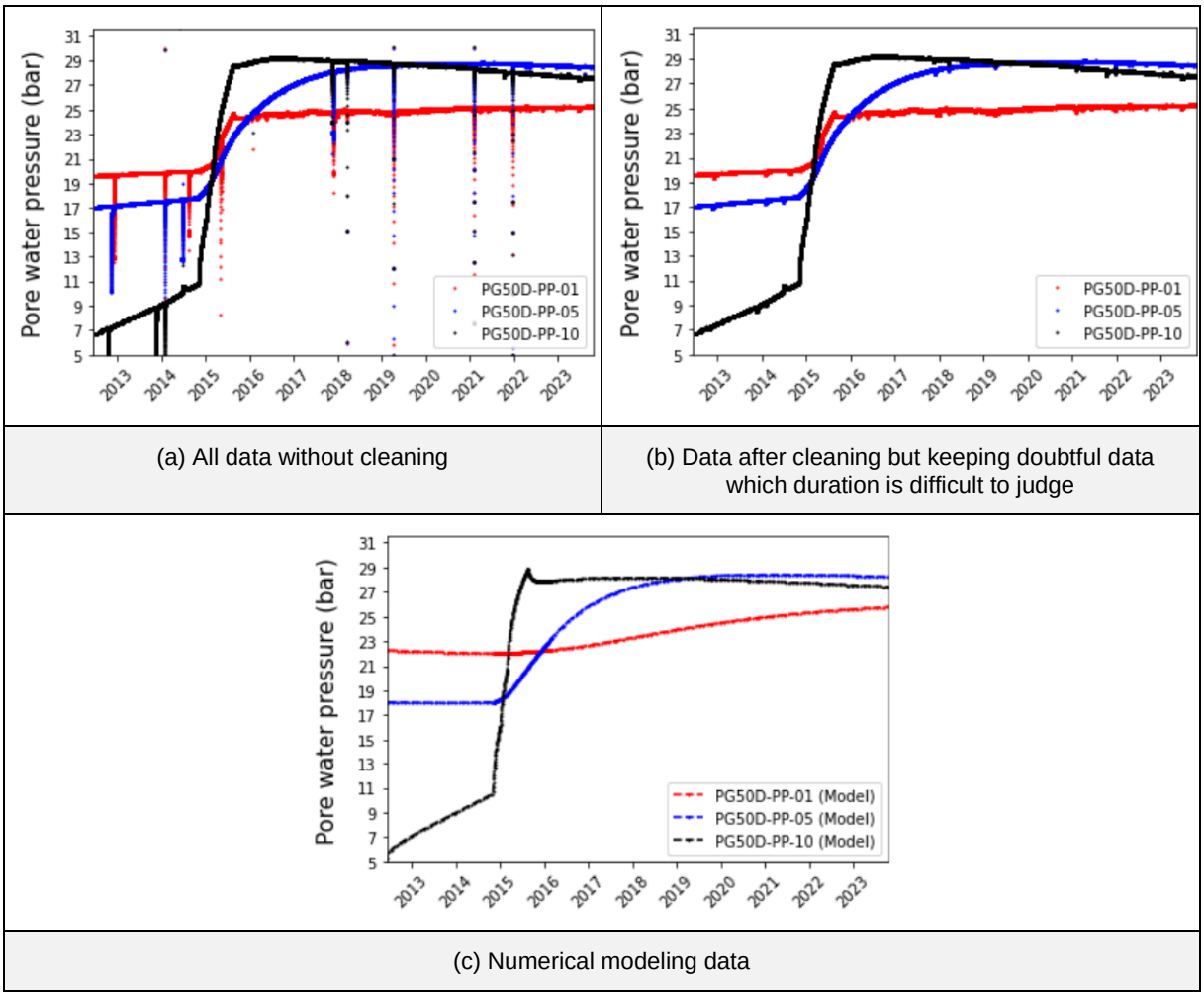
(c) Data after cleaning based on modelled data

Figure 5.6 - Time-series data from three pressure sensors in PG30D: PG30D-PP-01, PG30D-PP-02, PG30D-PP.

Another example is shown in Figure 5.7, where the results from the data cleaning and validation of some pressure transmitters connected to the PG50D filters are graphed. It is clear that the PG50D-PP-01 measurements are doubtful as they do not agree with the expected data from the model, although it is up to the expert judgment to indicate which data (and why) should be labelled as “doubtful”.

5.7.1 Pressure Transmitter Data from the PG Backfill Filters

Figure 5.8 shows the measurements as recorded by the pressure transmitters connected to the filters in the backfill of the PG. The disturbances are mainly due to (short time spans of) calibration activities, and also to short-term heater failures. The results of the cleaning, excluding both disturbances, are shown in the graph at the right (Figure 5.8b).

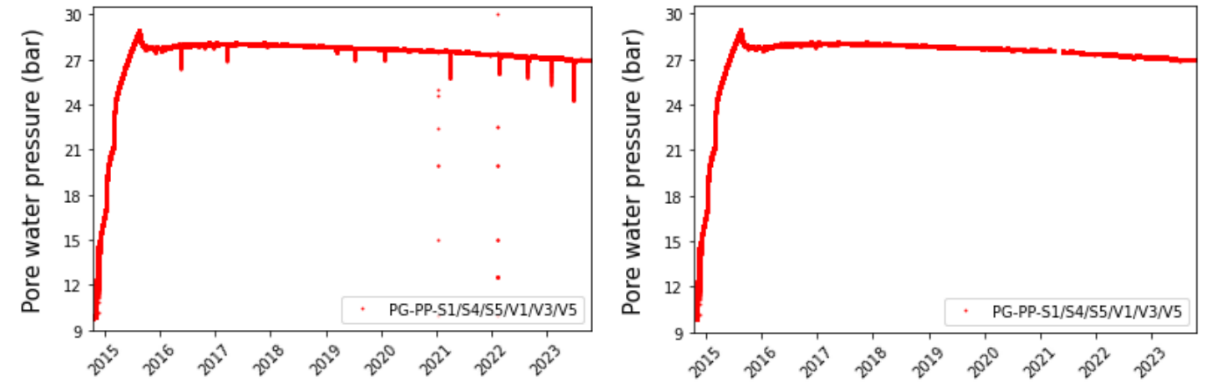


(a) All data without cleaning

(b) Data after cleaning but keeping doubtful data which duration is difficult to judge

(c) Numerical modeling data

Figure 5.7 - Time-series data from three pressure sensors in PG50D: PG50D-PP-01, PG50D-PP-05, PG50D-PP-10.



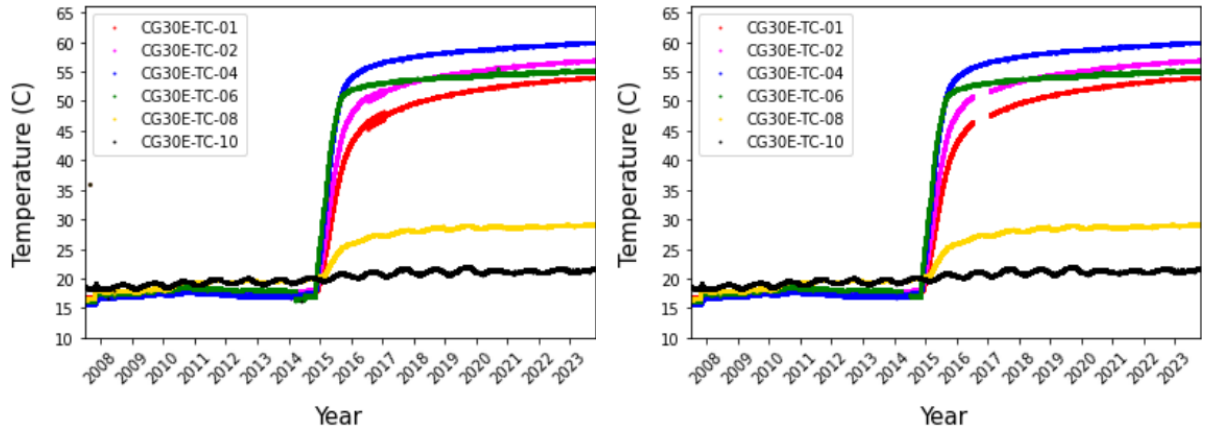
(a) Data without cleaning

(b) Data after cleaning with calibration and power failure related anomalies being excluded

Figure 5.8 - Time-series data from six pressure sensors in the backfilled PG: PG-PP-S1, PG-PP-S4, PG-PP-S5, PG-PP-V1, PG-PP-V3, PG-PP-V5.

5.7.2 Thermocouple Data from the CG Boreholes

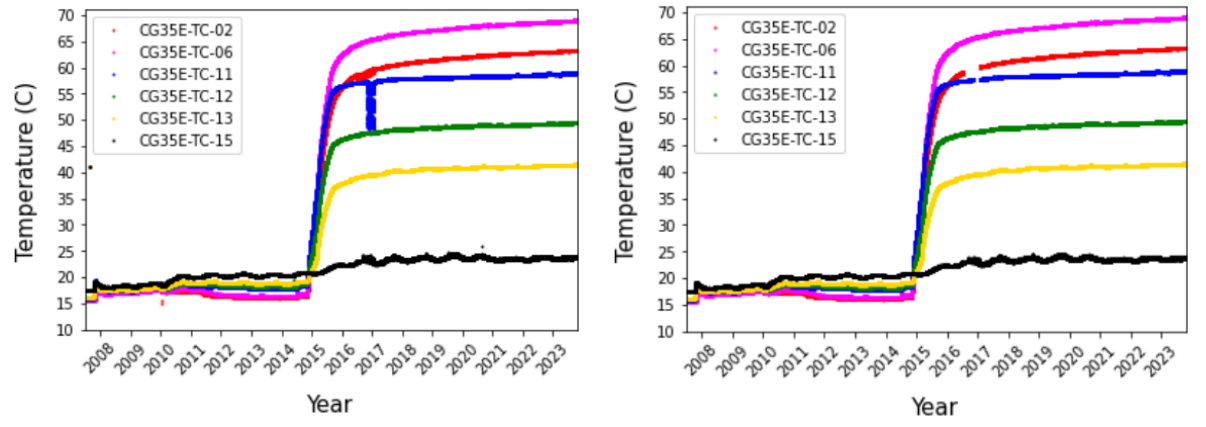
In general, anomalous data are rather limited in the measurements from the thermocouples mounted in the instrumented casings that have been installed from the CG, as their environmental conditions are quite favourable (dry, atmospheric pressure). Data from the thermocouples installed in the CG30E multifilter piezometer (Figure 5.9a) show limited outliers. Some thermocouples show disturbances owing to circulation of (hot) water in the capillary tubing of one of the filters used for geochemical on-line monitoring of the porewater. As this capillary tubing runs along some thermocouples, the circulation of hot water affects the thermocouple reading, which is no longer representative of the clay temperature next to the borehole casing. These measurements are then labelled and can be excluded from the final data set (Figure 5.9b). The seasonal variations in sensor CG30E-TC-10 TC (close to borehole mouth) reflect yearly and normal temperature evolution in the HADES URL. The CG35E thermocouples show a similar behaviour to the CG30E thermocouples (Figure 5.10).



(a) All data without cleaning

(b) Data after cleaning by removing the data segment with disturbance from pumping

Figure 5.9 - Time-series data from six pressure sensors in the backfilled PG: PG-PP-S1, PG-PP-S4, PG-PP-S5, PG-PP-V1, PG-PP-V3, PG-PP-V5.



(a) All data without cleaning

(b) Data after cleaning by removing the data segment with disturbance from pumping

Figure 5.10 - Time-series data from six thermocouples in CG35E: CG35E-TC-02, CG35E-TC-06, CG35E-TC-11, CG35E-TC-12, CG35E-TC-13, CG35E-TC-15.

5.7.3 Thermocouple Data from the PG Boreholes

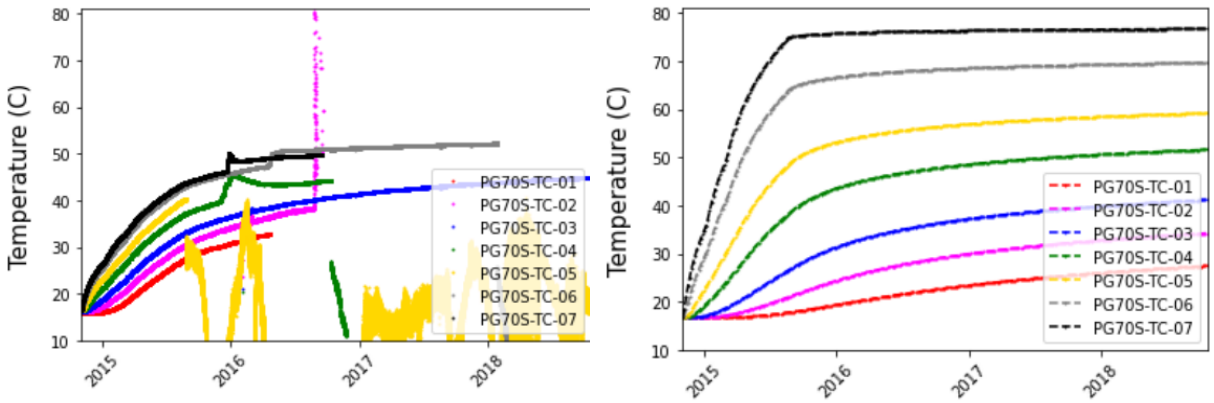
Thermocouples mounted in the PG multifilter piezometers are much more prone to showing disturbed measurements and failures. This is because of the harsher conditions (heated and saturated conditions, pressurized to more than 2.5 MPa), and also owing to the more complex installation (in particular, owing to extension wiring with the additional connections made in the field).

5.7.3.1 Thermocouples of PG70S Multifilter Piezometer

This instrumented casing, installed radially and horizontally from the PG, contains seven thermocouples, with PG70S-TC-07 being closest to the heated gallery (and hence with the highest temperature). From the measured data (Figure 5.11a), it is observed that most of the thermocouples failed in the first years of the heating phase (which started in November 2014). In addition, the temperature range shown by these thermocouples before their failure is much smaller than what was expected from the modelled data (Figure 5.11b).

This difference is most probably owing to the unexpected effect of local heat convection inside the borehole casing, creating an artefact, which was not taken into account in the model. The modelling only considers the heat conduction inside the clay formation.

Data cleaning would consist of indicating the measurements obtained after failure (which can be defined as a sudden deviation from the smooth initial evolution), while the data validation would consist of rejecting the whole data set – in particular after the start of the heating.

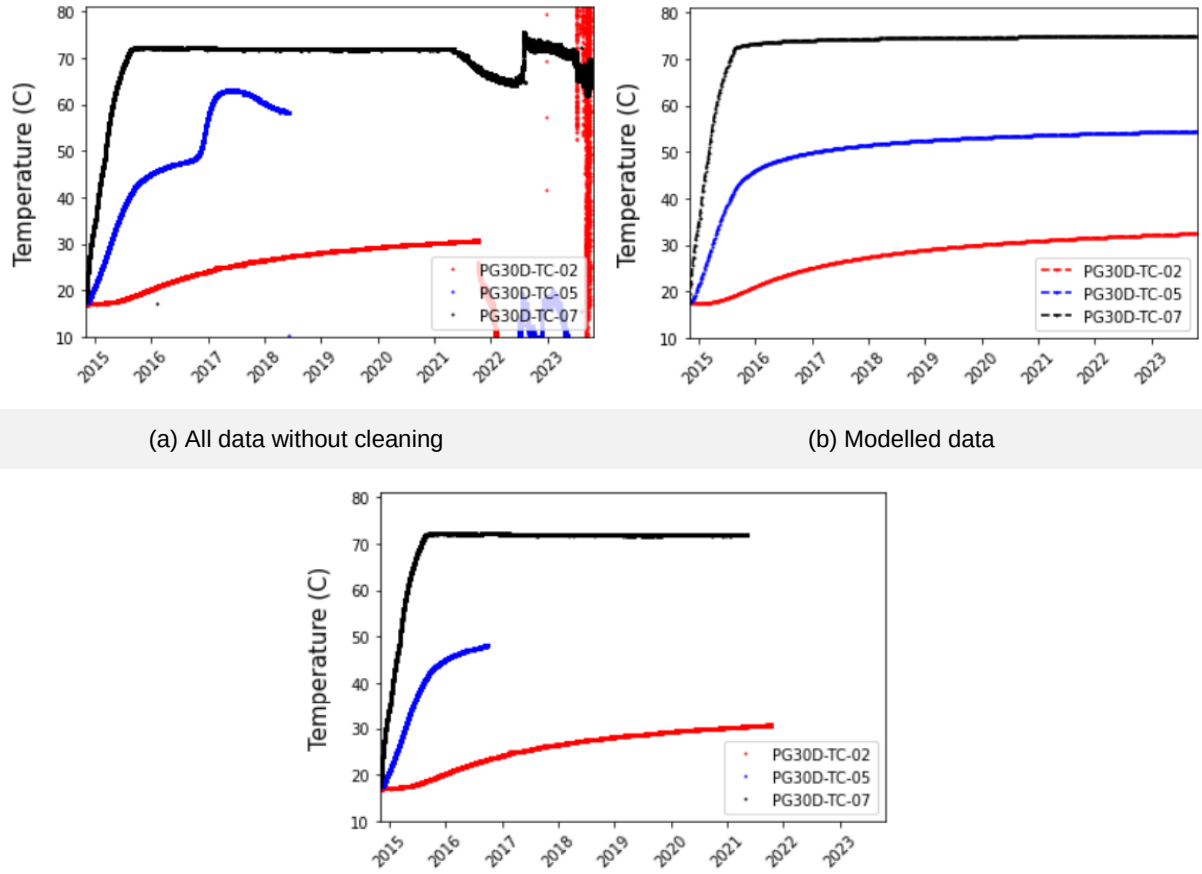


(a) All data without cleaning (b) Modelled data

Figure 5.11 - Both measured and modelled time-series data for seven thermocouples in PG70S: PG70S-TC-01, PG70S-TC-02, PG70S-TC-03, PG70S-TC-04, PG70S-TC-05, PG70S-TC-06, PG70S-TC-07.

5.7.3.2 Thermocouples of PG30D Multifilter Piezometer

Figure 5.12 presents the results from the data cleaning and validation performed on three thermocouples in PG30D (PG30D-TC-02, PG30D-TC-05 and PG30D-TC-07). The measurement time-series data present an obvious failure for PG30D-TC-05 and PG30D-TC-02 (Figure 6.12a). If comparison is made between measurement data and the corresponding modelling data at the three sensors, a different trend of evolution can be clearly observed for PG30D-TC-07 after 2021 and for PG30D-TC-05 after 2017 (Figure 6.12b). Since the change of the evolution trends observed in Figure 5.12a is unexpected, the unexpected data should be labelled as doubtful data. Figure 5.12c shows the cleaned data by removing the data after sensor failure and the doubtful data.



(a) All data without cleaning

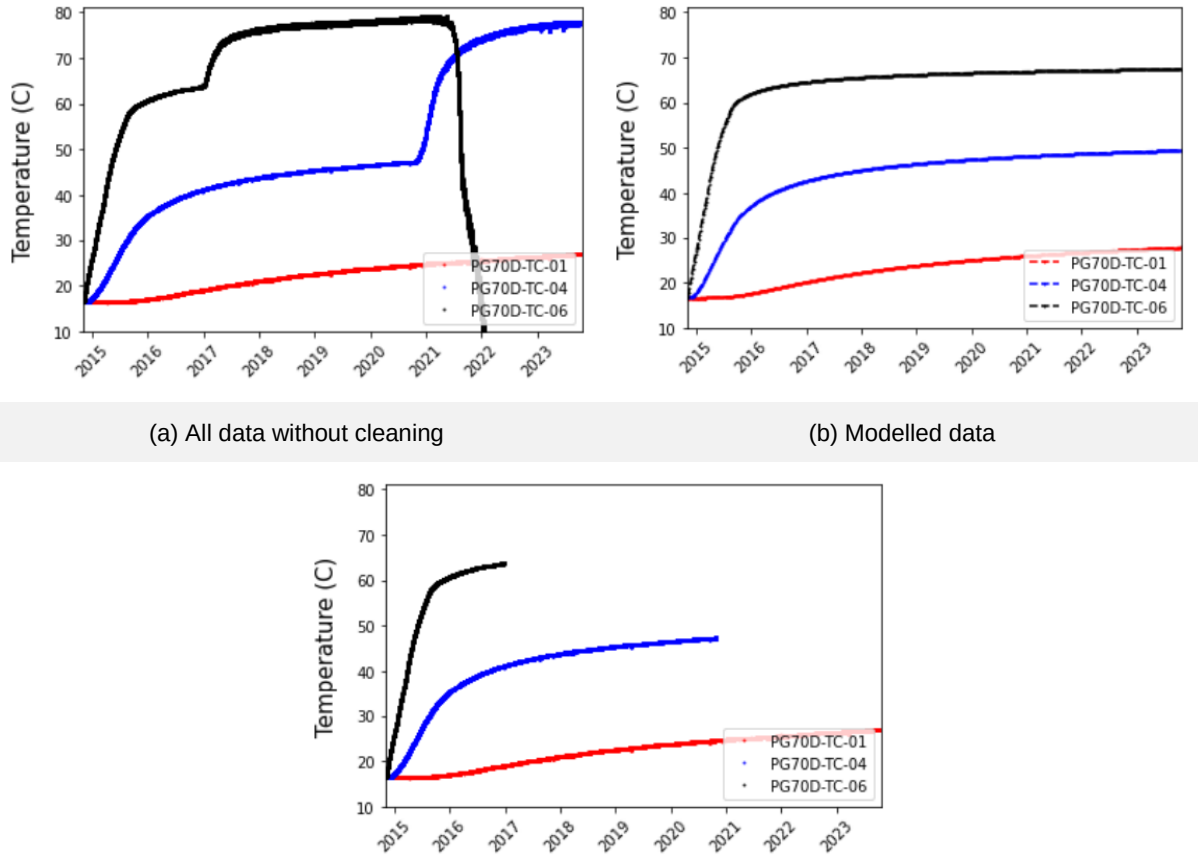
(b) Modelled data

(c) Data after cleaning with suport from modelling

Figure 5.12 - Time-series data from three thermocouples in PG30D: PG30D-TC-02, PG30D-TC-05, PG30D-TC-07.

5.7.4 Thermocouples of PG70D multifilter piezometer

Similar cleaning and validation results have been obtained for three sets of thermocouple data from the PG70D borehole (Figure 5.13).



(a) All data without cleaning

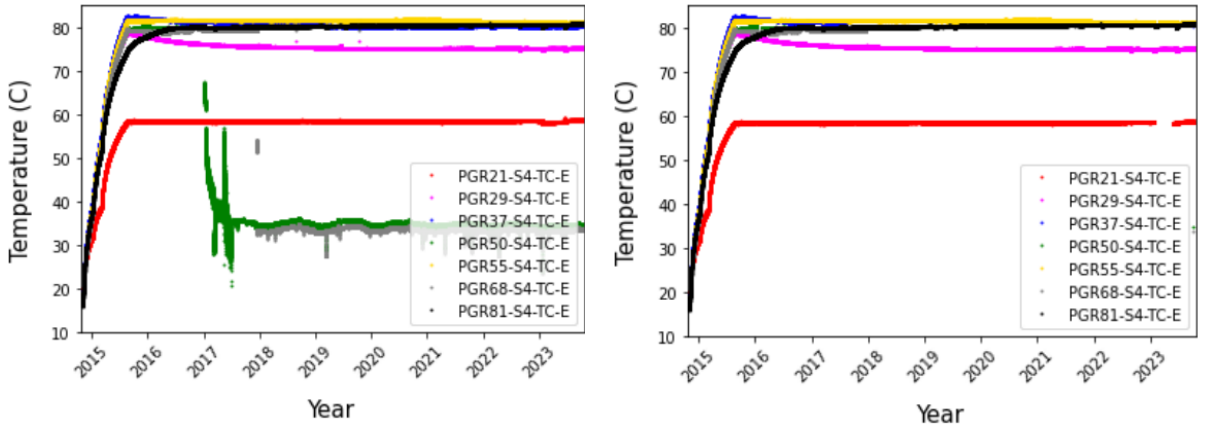
(b) Modelled data

(c) Data after cleaning with suport from modelling

Figure 5.13 - Time-series data from three thermocouples in PG70D: PG70D-TC-01, PG70D-TC-04, PG70D-TC-06.

5.7.5 Thermocouples Integrated in the PG Lining Segments

Sensor failures are also present in the thermocouples integrated in the concrete segments of the PG. Figure 5.14 shows the data from some sensors; the failed ones show temperatures reflecting the environmental conditions in the HADES tunnel as the thermocouple inputs on the data acquisition system have failed. Evidently, these values are not representative for the actual experimental conditions, and they are also excluded during the data cleaning process.

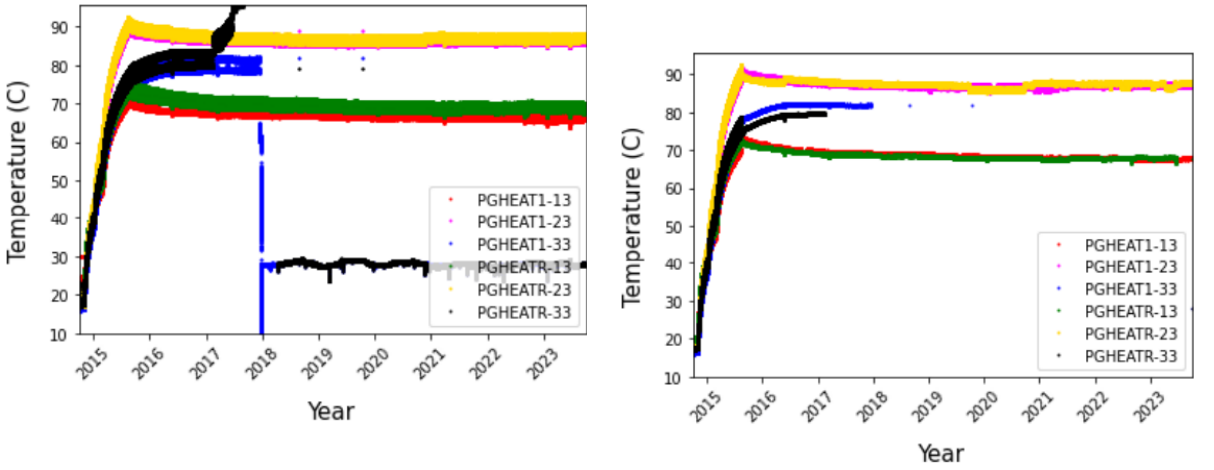


(a) All data without cleaning (b) Data after cleaning

Figure 5.14 - Time-series data from seven thermocouples at the extrados of segment 4 (S4) of seven rings.

5.7.6 Thermocouples Attached to the Heater Cables Inside the PG

A similar approach as for the previous case (thermocouples in the lining segments) has been applied for the thermocouples that are attached to the heater cables. Their original objective was to check these cables against overheating. Owing to the harsh conditions, about half of the installed thermocouples failed. Figure 5.15 shows the monitoring data from the thermocouples mounted in Sector 3 for both the normal cables and redundant cables, Figure 5.15a presents all data without cleaning, and Figure 5.15b graphs the data after cleaning by excluding outliers and data after sensor failure.



(a) All data without cleaning (b) Data after cleaning

Figure 5.15 - Time-series data from seven thermocouples at the extrados of segment 4 (S4) of seven rings.

5.8 Discussion and Conclusions

Monitoring data is often prone to errors and perturbations such as sensor manipulation, calibration disturbance, data transfer problems, and sensor malfunction. To guarantee reliable and high-quality data to be delivered to the users, and to ensure high-quality long-term data management, all data need to be carefully curated and managed by means of cleaning, validating, filtering, and labelling. With respect to labelling, each time series can be affected by different temporary events, namely normal functioning, sensor failure, sensor calibration or specific physical test, which can result in complex data values. Identifying in each recorded time series such possible sub-periods (that is, performing 1D data segmentation) is of primary importance for further use of the collected data.

In this data management process, an important part of the work is performed through algorithms. They prepare the work for the human expert by identifying suspect data and labelling them. In addition, they can provide relevant statistical information, as well as the adapted visualisation tools, to assist the expert judgment process.

The application of the method developed has shown that it has to be adapted to the particularities of each sensor type or measurement variable. For example, the approach used to screen thermocouple data is different to the approach used to screen porewater pressure measurements.

The data cleaning and validation method developed in MODATS for the PRACLAY test case is not optimised. The next steps in developing the method will include:

- Good knowledge and understanding of the experiment (different phases, boundary conditions) is essential, for example, the excavation phase exhibits fast pressure changes (even identification of individual excavation steps is possible), requiring specific attention during this phase for sensors close to PG.
- Each event has its specific characteristic in terms of porewater pressure, these features have to be characterised and quantified with some human effort.
- Correct detection and labelling of anomalous experimental data require a systematic documenting of all relevant events (which are part of the metadata); hence, metadata on the experimental setup (and by extension, on the repository) need a well-designed framework to make sure that these metadata are as complete as needed, and can be retrieved in a uniform way. In this context, the work performed in the RepMet Initiative [78], as well as, more in general, the concept of FAIR data [89] could be incorporated within the method.

The requirements on human effort can be optimised by applying data-driven ML techniques, which can monitor information from critical systems and flag anomalies as they occur [90]. Human experts can then view these warnings, and decide to deal with them on a case-by-case basis. This would decrease the amount of work that human experts have to do, thereby decreasing the burden to hire more and more experts. Building ML systems that can process information and identify anomalies is more cost-effective than creating human-based teams. It can also detect changes in the signal that are too subtle for humans to identify. The present data sets can serve as training data for such ML algorithms.

In order to perform the time series data segmentation in a more automated way, different types of state-of-the-art deep NN models suited to sequential data processing will be explored in future work, namely, recurrent NNs (e.g., [91]), and 1D convolutional networks (e.g., [92]).

6. Discussion: Advances in Data Management and Digital Twins in MODATS

Authors: Matt White (Galson Sciences), and, for Section 6.5, Martin Schoenball and Edgar Manukyan (Nagra), Arto Laikari (VTT), Jan Verstricht (EURIDICE) and Guangjing Chen (SCK CEN)

In this section, the outcomes from the six test cases are considered together to identify key integrated messages on data management. These key messages do not reflect specific tools or methods used in the test cases, but rather set out common principles that apply in all repository monitoring contexts. The discussion is structured as follows:

- Section 6.1 identifies integrated messages on data processing.
- Section 6.2 identifies integrated messages on data storage.
- Section 6.3 identifies integrated messages on data modelling and visualisation.
- Section 6.4 discusses the use of digital twins in the context of monitoring during repository operation.
- Section 6.5 presents a generic workflow focused on the use of monitoring data to enhance system understanding.

6.1 Data Processing

In MODATS, data processing was considered in most detail in the Nagra FE experiment test case, the VTT POPLU and Prototype Repository test case and the EURIDICE PRACLAY experiment test case.

As noted in the introduction (Section 1.3), monitoring data are prone to anomalies. These anomalies are caused by a range of internal and external (to the sensor) processes and events, including the influence of calibration and maintenance activities, sensor fouling, calibration drifts, variations in electrical inputs, data transfer problems, and configuration errors in the data acquisition system. To facilitate reliable analysis, all measurement data have to be processed prior to use, data errors have to be detected and handled, and data has to be in a harmonised form to allow cross-comparison of data from different sources. Data processing is the manipulation of raw data to make it suitable for its intended use.

Prior to undertaking data processing, user needs have to be identified as multiple actors have an interest in using data from repository monitoring. The actors within the monitoring process include scientists, engineers, decision makers, technicians, software developers, regulators and members of civil society. Processing does not have one end goal and data sets after processing will differ according to the user requesting the data. For example, an instrumentation engineer might have a focus on data anomalies for each specific sensor, whereas a modeller may want a reduced harmonised data set with all erroneous values removed. As has been highlighted in the POPLU and Prototype Repository, and PRACLAY test cases, a modular approach is useful in maximising flexibility in data processing (see also the generic workflow presented in Section 6.5).

The nature of data acquired from sensors varies. Some sensors automatically output parameter values, whereas some output raw data (e.g., in the form of voltages). For all sensor types, the original raw data should be stored and kept available for the period of the monitoring programme. Data processing should not delete data from the data set, as there may be a need to review and revise interpretations at a later stage in the programme. Instead, data processing should identify and flag data considered to be an anomaly.

The test cases in MODATS have provided an opportunity to identify and categorise the different types of anomalies that can occur in monitoring data (Table 6.1). These anomalies can be identified through consideration of the characteristics defined in the table. These characteristics are defined in a qualitative sense in Table 6.1 because the data processing approach of (time series) measurement data needs to be customised according to the sensor type (e.g. characteristics of thermocouple data are different from those of porewater pressure transmitters).

Table 6.1 – Identification of the main types of monitoring data anomalies and identification of options for their management.

Category	Characteristic (and example, if available)	Processing Options
NaN	No value is recorded for a particular timestep or for a specific period within the monitoring data set	Options include averaging of the adjacent values, or leaving the timestep as NaN and labelling the timestep so that the value can be left out from some uses
Null Value	Value recorded for a particular timestep is zero	Options include averaging of the adjacent values or leaving the timestep value as zero and labelling the timestep so that the value can be left out from some uses
Duplicate Values	The data file includes more than one value for a time step in a specific location or for a specific sensor	Options include averaging of the values, labelling of the values as uncertain to ensure they are not used in further analyses, or selection of a preferred value based on a pre-determined method
Non-Physical Value	Value recorded is not possible (e.g., negative relative humidity)	These data would be flagged and removed from data sets before use
Implausible Value	Value recorded is not reasonably expected (e.g., negative temperature)	These data would be flagged and removed from data sets before use based on a pre-defined approach for each parameter (this might include use of uncertainty ranges based on modelling, or definition of plausible parameter value ranges defined by expert judgement using formal elicitation methods)
Unexpected Constant Values	The values returned by a sensor do not change over time	These data would be flagged and removed from data sets before use
Spikes	A sharp change in measured value, followed by a sharp change in opposite direction for the subsequent value(s). Spikes can be single values or occur over a short period relative to the monitoring period.	These data would be flagged and removed from data sets before use, the value of the gap in data caused by removing the spike could be left as NaN or could be recreated by averaging adjacent values
Temporary Step Change	Data records show a sharp change in values, before a similar sharp change in the opposite direction, and then progressing at a similar rate of change as previously	This type of anomaly could be caused by temporary malfunction of the monitoring system such as an increase in electrical current over a short period. Data might be flagged and removed from data sets used in analysis, or an algorithm might be developed to correct the affected data.
Permanent Step Change	Data records show a sharp change in values, before progressing at a similar rate of change as previously	It may be possible to correct for permanent step changes, or the flagged data might be removed from the data sets used for analysis
Noise	Noise is characterised by a scattering of values around a central trend	Noise that is not characterised by the features specified for other anomaly types is usually challenging to remove from data; therefore, noisy data need to be evaluated on a case-by-case basis and used in a manner suited to the end user needs

Category	Characteristic (and example, if available)	Processing Options
Outliers	The values recorded by one sensor are inconsistent with values recorded by close-by sensors measuring the parameter in the same way in the same medium	Clearly defined outliers would be removed from data sets, for example if one sensor in a group was shown to have behaviour inconsistent with other close-by sensors; however, removal could only be undertaken based on a pre-determined formal process
Unexpected Data Trends	Data trend is inconsistent with model prediction	Data that is inconsistent with modelling would have to be subject to root cause analysis to identify the reason for the discrepancy; this might involve testing of sensor performance (e.g., recalibration where possible) and revision to the modelling (e.g., consideration of the conceptual model implemented, and investigating the effect of changing parameter values); it would not be acceptable to remove data with unexpected data trends from analyses until the root cause of the trend was identified

Following identification of anomalies, the data must be treated prior to use. MODATS has developed and made available several tools for data processing (Table 6.2). Good practice in data processing is for software to provide good visualisation of time series data to give the expert optimal insight in the data, and basic statistical analyses to indicate signal quality (e.g. daily averages and standard deviations to indicate signal-to-noise ratio or repeatability).

Data processing is currently a hybrid process; one or many algorithms perform the screening and preliminary labelling of the measurement data, which is then confirmed, rejected or altered using expert judgement. Automatic checking of data is undertaken on a data-point-by-data-point basis, using *a priori* physics-based knowledge (e.g., identification of non-physical and implausible values) and statistical tests (e.g., standard deviation from median values to identify outliers). To make computed results traceable, all data processing steps should be based on a quality assurance procedure that defines how the process should be conducted, and recorded in a version control system.

In addition to identification and treatment of anomalies, data processing requires amalgamation of data from different sensors (potentially including sensors monitoring the same parameter using different technologies) into integrated data sets for analysis. This includes data reduction so that modelling and visualisation can proceed efficiently. Temporal sampling of sensor data should be as homogeneous across sensors and time as possible to allow easy comparison across several sensors. The temporal resolution must be high enough to cover all relevant phenomena (e.g. porewater pressure can change more rapidly than temperature, so sampling of pressure data should be performed at a higher frequency than sampling of temperature data).

Multidisciplinary domain expertise is needed. Data sets from several sensors should be processed together, alongside metadata, to cross-check conclusions. Applying domain knowledge of the whole environment and of the used sensors and data acquisition setup is also crucial to ensure correct processing. Expertise is applied at all stages of data management, not just at the analysis stage by the data users.

Table 6.2 – Tools developed in MODATS for data processing, including a description of how to access the tool.

Org.	Tool	Description	Availability
UFZ	OGS VisCoSiR	OGS VisCoSiR is a tool, that supports domain experts in combining and analysing multiple spatial simulation results, i.e. from different simulation software or conducted with different parameter setups. In contrast to other tools, OGS VisCoSiR focuses on the visual comparison in 3D. It consists of two modules: 1) A combiner, that allows to spatially and temporally combine (i.e., interpolate) the data sets and calculate metrics (like deviation), and 2) an analyser that presents the combination result and allows domain experts to explore the differences and similarities in an interactive, visual way.	The tool is available in a public git-repository: https://gitlab.opengeosys.org/vislab/non-unity-apps/viscosir-cxx
UFZ	Converter for Transforming Measurement Data to VTK Unstructured Grids	In order to make monitoring data available for visualisation, we implemented a python script, that accesses the FEIS database (provided by Nagra) and retrieves temperature measurements for a given list of sensors and dates. In this way, we provide infrastructure for making observation data compatible with OGS tools and scientific visualisation software such as ParaView.	The python script is available under the license GNU GPLv3, and can be accessed online: https://zenodo.org/records/10017852 Graebling (2023). Converter for Transforming Measurement Data to VTK Unstructured Grids. doi:10.5281/zenodo.10017851.

Org.	Tool	Description	Availability
UFZ	Interactive Visualizations: Digital Twin Prototype & Virtual Field Trip	The visualisation applications are prototypes of digital replicas of the Mont Terri URL, that integrate heterogeneous data from several different sources. Both, simulation results and observation data are displayed within the same system.	Because these prototype applications include non-public data sets from external contributors as well as access to non-public external databases, they cannot be made publicly accessible. However, the visualisations are available in form of public videos online: Digital Twin Prototype: https://www.youtube.com/watch?v=X71DF7SG5uc Virtual Field Trip: https://www.youtube.com/watch?v=kH34J9cZ3aI
PSI	# Step 1: Data collection	# Load the temperature data	import torch import torch.nn as nn import numpy as np import pandas as pd import random import matplotlib.pyplot as plt
PSI	# Step 2: Preprocessing and normalization	# 2.1 data prepocessing # define range for S_temperature and T_temperature	H_range = [0.1, 100] T_range = [1, 180]
PSI	# Step 2: Preprocessing and normalization	# 2.2 Split the data into 70% for training and 30% for validation	n_train = round(0.7 * T_temperature.shape[0]) n_val = T_temperature.shape[0] - n_train T_shape = T_temperature.shape T_shape_rounded = tuple(round(dim) for dim in T_shape)

Org.	Tool	Description	Availability
PSI	# Step 2: Preprocessing and normalization	# 2.3 data normalization # Normalize the data using the mean and standard deviation of the training set	x_val = (x_val - mean_x_val) / std_x_val y_val = (y_val - mean_y_val) / std_y_val z_val = (z_val - mean_z_val) / std_z_val t_val = (t_val - mean_t_val) / std_t_val H_val = (H_val - mean_H_val) / std_H_val T_val = (T_val - mean_T_val) / std_T_val

PSI	# Step 3: Build the Recurrent Neural Network (RNN) model	# Define the neural network architecture	n_inputs = 4
			n_outputs = 2
			seq_length = 1 # Sequence length for RNN input
			# Define the number of neurons in each layer
			n_neurons_rnn = 20 # LSTM layer neurons
			n_neurons_fc1 = 20 # First fully connected layer neurons
			n_neurons_fc2 = 20 # Second fully connected layer neurons

PSI	# Step 4: Train the model	# 4.1 Instantiate the model	import torch.nn.functional as F
			model = NN()
			# Define the loss function (criterion) and optimizer
			criterion = nn.MSELoss()
			optimizer = torch.optim.Adam(model.parameters(), lr=0.001)

Org.	Tool	Description	Availability

PSI	# Step 4: Train the model	# 4.2 Measure CPU and GPU usage after prediction	# Measure CPU and GPU usage after prediction
			cpu_percent_after = psutil.cpu_percent()
			gpu_usage_after = get_gpu_usage()
			# Measure time after prediction
			Y_pred = model(input_val)
			training_time = time.time() - start_time

PSI	# Step 5: Evaluate the ML model on the validation set	# Evaluate the model on the validation set	# Create a DataFrame with the predicted and true values of H and T
			results_df = pd.DataFrame({
			"H_pred_val": H_pred_val,
			"H_val": H_val,
			"T_pred_val": T_pred_val,
			"T_val": T_val
			})

PSI	# Step 6: Visualize the results	# Plot the predicted temperature T vs the true temperature T for the validation set # Plot the predicted relative humidity H vs the true relative humidity H for the validation set	import matplotlib.pyplot as plt

Org.	Tool	Description	Availability
PSI	# Step 7: calculate the indicators of MAE, MSE, r-squared	# Unnormalize the validation data sets # Compute the performance metrics for temperature T # Compute the performance metrics for relative humidity H # Print the performance metrics for temperature T # Print the performance metrics for relative humidity H	from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
VTT	preprocessing.py	**Preprocessing and formattind data** Method: create_raw_dataset Creates a unified raw dataset of given csv or xlsx files and selected columns. Informs for each data element from what file it was taken. Drops duplicates.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	format_*.py	**Preprocessing and formattind data** Changes source file-formats to csv, adds column 'timestamp' for timestamps using unix epoch time format. In some cases, modifies column names to avoid confusion later. Data sources: aitemin, campbell, datataker, fuktlog, extensometer, rock.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	preprocess_*.py	**Preprocessing and formattind data** Uses create_raw_dataset in preprocessing.py to create a unified source-specific dataset from the formatted files.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.

Org.	Tool	Description	Availability
VTT	cleaning.py	**Cleaning** Method: get_length_gaps Counts length of gaps in selected columns of data	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: resample_dataframe Changes the sampling rate of data to target sampling rate	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: clean_analysis_data Creates a dataframe without any gaps and resampled to preferred frequency.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: smooth_known_events Smooths out known (manmade) events from given data with interpolation.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: filter_resids Returns found trend and residuals in separate dataframes.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: detect_outliers Detects outliers of selected columns individually.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.

Org.	Tool	Description	Availability
VTT	cleaning.py	**Cleaning** Method: detect_outliers_group Detects outliers for selected columns as a group.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: identify_spikes Identifies the peaks in timeseries that are considered as spikes	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: peak_widths Detects start and end of identified spikes.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: remove_spikes Smooths the identified spikes. Uses peak_widths.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: remove_baseline Removes baselines from given timeseries.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	cleaning.py	**Cleaning** Method: detect_correlation_change Tests whether two datasets with the same attributes have differences in correlation.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.

Org.	Tool	Description	Availability
VTT	Cleaning_pipeline.ipynb	**Jupyter notebooks** Purpose: Cleaning pipeline Example on a cleaning pipeline for a preprocessed dataset utilizing the functionality in cleaning.py.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	Anomaly_detection.ipynb	**Jupyter notebooks** Purpose: Anomalies Example on how anomalies can be detected using the functionality in cleaning.py.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	Drifting.ipynb	**Jupyter notebooks** Purpose: Drifting Example on how drifting can be detected using the functionality in cleaning.py.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	Spikes_detection_removal.ipynb	**Jupyter notebooks** Purpose: Spikes Example on how spikes can be detected and removed using the functionality in cleaning.py	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	RockData_BeforeVsAfterPOPLU_PCAMethod.ipynb	**Jupyter notebooks** Purpose: Rock data analysis Example on how to compare two time periods of high dimensional data: reduction of dimensions, time evolution in reduced space, detecting change visually, separating dominant behaviour and noise.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.
VTT	RockData_BeforeVsAfterPOPLU_StatTests.ipynb	**Jupyter notebooks** Purpose: Rock data analysis Example on basic statistical visualisations and testing for comparing two time periods of sensor data.	Currently in the access restricted VTT GitLab repository. If related WMOs allow the publication, can be released via e.g. GitLab.

6.2 Data Storage

In MODATS, data storage was considered in most detail in the Nagra and UFZ FE experiment test cases.

Data storage is the organisation of data and metadata to make it available for its intended use (see [4] for a discussion of the repository monitoring context). As much data as possible should be stored in a structured way to prevent the loss of implicit knowledge and to allow for automatization of data interpretation in the future.

It is necessary and important to clearly establish the ownership and responsibility for different parts of the monitoring database. These include the data, the metadata, and the expertise required to use the data, including the understanding of how the database has been designed and constructed. Ownership and responsibility for these different aspects of the monitoring database might be vested in different roles within a WMO.

Owing to the range of types of information that need to be included in a monitoring database (e.g.: objects, time series, variables, parameters, activities and metadata), it is a challenge to organise and structure the database in a way that facilitates easy access for data users. In particular, there is a risk of double storing large amounts of data in the database (for example, the co-ordinates of tunnels or the occurrence of events relevant to the interpretation of monitoring data), which can lead to inconsistencies and accessibility issues. Design of the database should be undertaken from the perspective of the data users rather than the data providers.

Tools to access, compile, interpret and present monitoring data often have a high learning threshold making them inaccessible to many data users. This often requires additional processing/re-learning steps, often with help from an expert. User-friendly tools with low-learning thresholds are more likely to reduce costs and meet deadlines.

Harmonised ontologies[8] and metadata conventions would benefit the efficient and transparent storage of monitoring data. Work in MODATS has highlighted the application of outcomes from the OECD NEA RepMet initiative [78]; high-level guidance provided in the RepMet reports should be followed when planning for the storage of metadata in a repository monitoring programme. In order to facilitate the implementation and use of shared digital infrastructure as well as to exploit the full capacities of automatization, data standards and shared APIs should be agreed on. This standardisation should also cover metadata, which needs to be structured in a uniform and consistent way, through time and by the different data sources.

When interpreting monitoring data, it is necessary to distinguish between cause and effect. It is then paramount that the external influences (natural and anthropogenic) relative the monitored system are identified. During the operational phase of the repository there will be many activities on-going simultaneously that might generate response in the monitoring system. It is therefore necessary to have a system in place that will document all activities in support of data interpretation and root-cause analysis that might be warranted. This functionality, of booking and tracing activities/events, is part of the monitoring database.

WMO-owned data are strictly controlled. This includes limits on the use of cloud environments for data storage, and the physical locations of servers. This results in conflicts between confidentially and information security on the one hand, and a desire for openness on the other hand. Furthermore, it affects company procedures, information sharing/exchange and standards.

In MODATS, an API has been developed to allow access to data related to the FE experiment and stored in the FEIS (Section 3.2.1), and a visual tool has been developed for the comparative exploration

[8] An ontology is the classification and explanation of entities such as (in this context) the information in a database.

and analysis of simulation data was developed for use alongside OGS (Section 3.4.4, see further discussion in Section 6.3).

Monitoring data and metadata need to be available for decades to support repository operation and closure. As a consequence, there needs to be a plan for regular updates to hardware and software, as it is likely that databases and data management tools will change over the lifetime of repository monitoring programmes. To date, the main tools for data processing and analysis of URL experiments have been spreadsheet applications, and significant manual work has been necessary for data processing. In the future, approaches are likely to include ML and AI, as demonstrated by the test cases reported herein. Flexibility is required in the manner in which data is stored to allow for different ways of processing and analysing data in the future. There is a need for databases to be capable of handling different time systems, to adjust for daylight saving time, and to accommodate different spatial coordinate systems. Work in MODATS has demonstrated how standardised file formats such as VTK can be used to support sharing of information and, in this way, improve robustness against future software developments. The use of a common and widely used data format, such as VTK, improves the robustness compared to establishing a new and specialised file format used only in the domain, because established formats are founded on the technical support of a large number of users and are less likely to be replaced by new formats.

When observation data is displayed in visualisation tools, data cleaning methods must also be communicated transparently to end users. The creation of a shared library for such methods (including comprehensible descriptions on different knowledge levels) could reduce redundancy and create benefits.

Although monitoring data is often described as vast, it is not large with respect to other data science applications. Therefore, the size of the data sets should not be a problem for repository monitoring. For example, the input data for modelling of the PSI FE test case was a data file of 1.9 GB, which provided temperature and relative humidity data for a period of ~5 years. Extension of the monitoring period to ~50-100 years would similarly not result in an excessively large data set. However, although the data file is not large, the information contained within it is multi-faceted, and requires efficient and effective algorithms to model and visualise it. Investigation of the use of data for purposes such as forecasting and education, was considered in data use test cases, which are discussed in the next two sections.

6.3 Modelling and Visualisation

In MODATS, data use was considered in most detail in the Andra, Amvalor and ENSAM ALC1605 test case, the PSI FE experiment test case, and UFZ FE experiment test case.

Experience and expertise have been developed in coupled numerical modelling of THMC processes over the last four decades of RD&D in geological disposal. For example, work in the EURAD Mechanistic Understanding of Gas Transport in Clayey Materials (GAS) and Influence of Temperature on Clay-Based Material Behaviour (HITEC) WPs of EURAD used a combination of experimental and modelling approaches to increase the understanding and predictability of the impact of coupled gas and heat transport on clay barriers [93].

However, repository monitoring during operations introduces new requirements on modelling, most notably:

- The requirement for more rapid modelling based on monitoring data acquired continuously through the operating period.
- The requirement to communicate the results effectively and efficiently, in order to strengthen understanding and support decision making.

In addition, the emerging application of data science approaches allows for enhanced use of monitoring results by using the data acquired directly in the modelling approach.

In MODATS, the application of data science approaches to modelling of monitoring data has been investigated in the ALC1605 test case and the FE test case.

In the ALC1605 test case, it has been demonstrated that hybrid twin models are a promising approach for modelling the thermal evolution of a HLW disposal cell. In the hybrid twin approach, a physics-based (surrogate) model using only heat conduction is first applied to the model domain. Monitoring data are then used to quantify the uncertainty in the model (the ignorance), which, when applied to the surrogate model, provides the ability to rapidly and accurately model the data from the ALC1605 test case. The use of a hybrid twin, rather than a purely data-driven model has several advantages, particularly in modelling of the thermal evolution of the disposal system over the operational period, as it grounds predictions in well-established physics, enhancing the reliability of our results.

The PSI FE experiment test case came to a similar conclusion as the ALC1605 test case, i.e. that a hybrid modelling approach combining the physical model with a data-driven model provided the best modelling results. In the PSI test case, the preferred model was the PIML, which combines the kNN algorithm with data on the heat power or power density heat source.

In addition to using data in advanced modelling applications, data (and the models generated from the data) can be used to communicate the outcomes from repository monitoring. The UFZ test case demonstrated the power of visualisation tools to develop understanding by integrating geological, infrastructure, monitoring and simulation data. The VEIS system automatically updates the when underlying data change, and can therefore be considered a digital twin prototype with a focus on visualisation.

The application of the VEIS was extended for use by university students. To extend the use of visualisation systems for these actors, additional features are required to facilitate independent exploration of the information. These include contextual information and tasks to be performed at each viewpoint. The inclusion of tasks is particularly important, as undertaking these tasks provides the user with a more interactive experience, increasing their learning from the process. Evaluation of the use of the VEIS in this way in a trial demonstrated the effectiveness of the approach in communication and education.

The work of UFZ also demonstrated that visualisation can be of significant benefit when comparing simulation results from different numerical models of THMC processes. The approach adopted involved a two-step process. The first step is combining the data into a common file format, which allows spatial and temporal differences in the data to be reconciled. The second step is to view the data and allow for contouring, slicing, selecting and filtering (based on conditions) of timesteps.

6.4 Digital Twins: Use in Monitoring during Repository Operation

6.4.1 Context to Repository Digital Twins: The Characteristics of Repository Systems

Repository systems are multi-faceted. At a high level they can be considered as a combination of the active system that delivers the construction, operation and closure of the facility in a safe, secure, sustainable and publicly acceptable manner; and the passive system that delivers protection of people and the environment in the long-term.

These systems contain elements with widely differing functions, structures and processes. Some examples are as follows:

- The geosphere is comprised of rock, discontinuities (e.g., faults and fractures), fluids (e.g., groundwater and porewater) and gases. The function of the geosphere is dependent on the disposal concept but can include isolation of the waste, containment of the waste, and protection of the EBS.
- The EBS is comprised of cementitious materials, clay materials (e.g., pre-compacted bentonite blocks and bentonite pellets), waste containers and wasteforms. As with the geosphere, the

functions of the EBS are dependent on the disposal concept but can include containment and attenuation of releases.

- The excavated repository will contain significant infrastructure such as ground support systems (e.g., shotcrete), ventilation systems, materials handling systems, monitoring sensors and boreholes that support the emplacement of waste and closure of the repository.
- The open repository system will also be host to dynamic underground processes such as the movement of people and materials (including the waste and EBS), air flows and seepage of water and gases into and out from the tunnels.

All of these systems operate on different temporal and spatial scales, and contain different materials that are affected by different types of processes. Some of the systems can be characterised and understood in detail. For example, the materials and function of the ventilation system. Other systems are characterised by irreducible uncertainty. For example, there is no technology in existence now or in the foreseeable future that can map the intersections, tortuosity and mineral coatings of fractures in a deterministic sense. Similarly, the understanding of the microstructure of clay host rocks will always be subject to uncertainty, for example in the distribution of pore space and the mineral content of specific locations. The EBS is also subject to uncertainty such as the water content distribution and packing density of emplaced bentonite pellets, the orientation of waste products in intermediate-level waste packages, and the diffusion of radionuclides within spent fuel pellets and the resulting impact on instantaneous release fractions.

Uncertainty in geological disposal programmes has been recognised for decades and is addressed within the safety case through a variety of means. This uncertainty will have an impact on the extent to which digital twins can be applied in repository programmes.

6.4.2 Digital Twins and Monitoring in Support of Post-Closure Safety

Digital twins have the potential to support monitoring programmes in demonstration of compliance with requirements and conditions linked to post-closure safety. These requirements and conditions differ between each repository programme are yet to be fully developed in some cases. However, monitoring during repository operation is not expected to be based on extensive sensor networks as currently employed in URL experiments. The extensive networks used in URL experiments are deployed to develop understanding of coupled processes occurring in the EBS and geosphere, and this understanding is fed into the safety case. Monitoring during repository operation is anticipated to be more focused on support of limited modelling used to check system behaviour (for example to confirm the absence of any conditions that could affect the safety of the facility after closure), and has to be implemented such that it does not impact passive safety. Hence, approaches to developing digital twins for monitoring repository processes during the operational period will most likely have to be developed with much sparser data sets than digital twins of URL experiments. Likewise, measured gradients might be smaller since sensors will generally be placed at a significant distance from the waste packages.

Digital twins can support monitoring by developing surrogate models that are able to recreate spatially-distributed time series data without the need for resource-expensive and time-consuming coupled process modelling. An example would be the development of a 3D model of the temperature field over time, without the need for modelling of hydraulic and mechanical interactions.

To achieve this aim, surrogate models require a PBM that incorporates the processes of greatest significance to the objective of the model, and the use of a DDM that incorporates ML approaches such as neural networks to train the model to deliver accurate results.

6.4.3 MODATS Definition of Digital Twins

As noted in the introduction (Section 1.3.3), a key feature of a digital twin is automatic feedback of monitoring information to better meet the objectives. It is the feedback from monitoring information that distinguishes digital twins from other types of models (e.g., geological interpretations, building information modelling representations of infrastructure, and coupled process models).

Furthermore, as justified above, digital twins are built with a specific objective in mind and cannot replicate the full reality of the repository. Therefore, it is the opinion of MODATS that repository digital twins are not “one size fits all”, but come in different forms depending on the objective for which they are developed. Several digital twins might be created for one repository, each with a different purpose, but, potentially, all with a common data architecture to enhance interoperability.

MODATS has developed the following definition of a repository digital twin:

> *A repository digital twin is a virtual model of part of a repository that is updated automatically to address specific objectives.*

6.5 Generic Workflow for Conformance Verification

The preceding sections have provided a discussion of each step involved in data management (including processing and storage), modelling and visualisation. This sub-section presents a good practice workflow for the evaluation of monitoring data. This includes the verification of the monitoring system, uncertainties in the as-built-state of the monitoring system, as well as an assessment of potential influences and processes that were discarded during the setup of numerical models. Here we propose a general workflow that should guide the development of processes from the start of monitoring through quality control of monitoring data through to decision-making.

The workflow presented here builds on the test cases presented in this report, especially FE Experiment Test Case 1 (Section 3.2), the POPLU and Prototype Repository Experiments Test Case (Section 4) and the PRACLAY Experiment Test Case (Section 5). In addition, practical experience in monitoring URL experiments has been taken into account (see, for example, [94]).

The workflow is based on five principal steps, that will be discussed in this chapter:

1. Data acquisition
2. Data screening
3. Anomaly interpretation
4. Comprehensive data analytics
5. Decision support

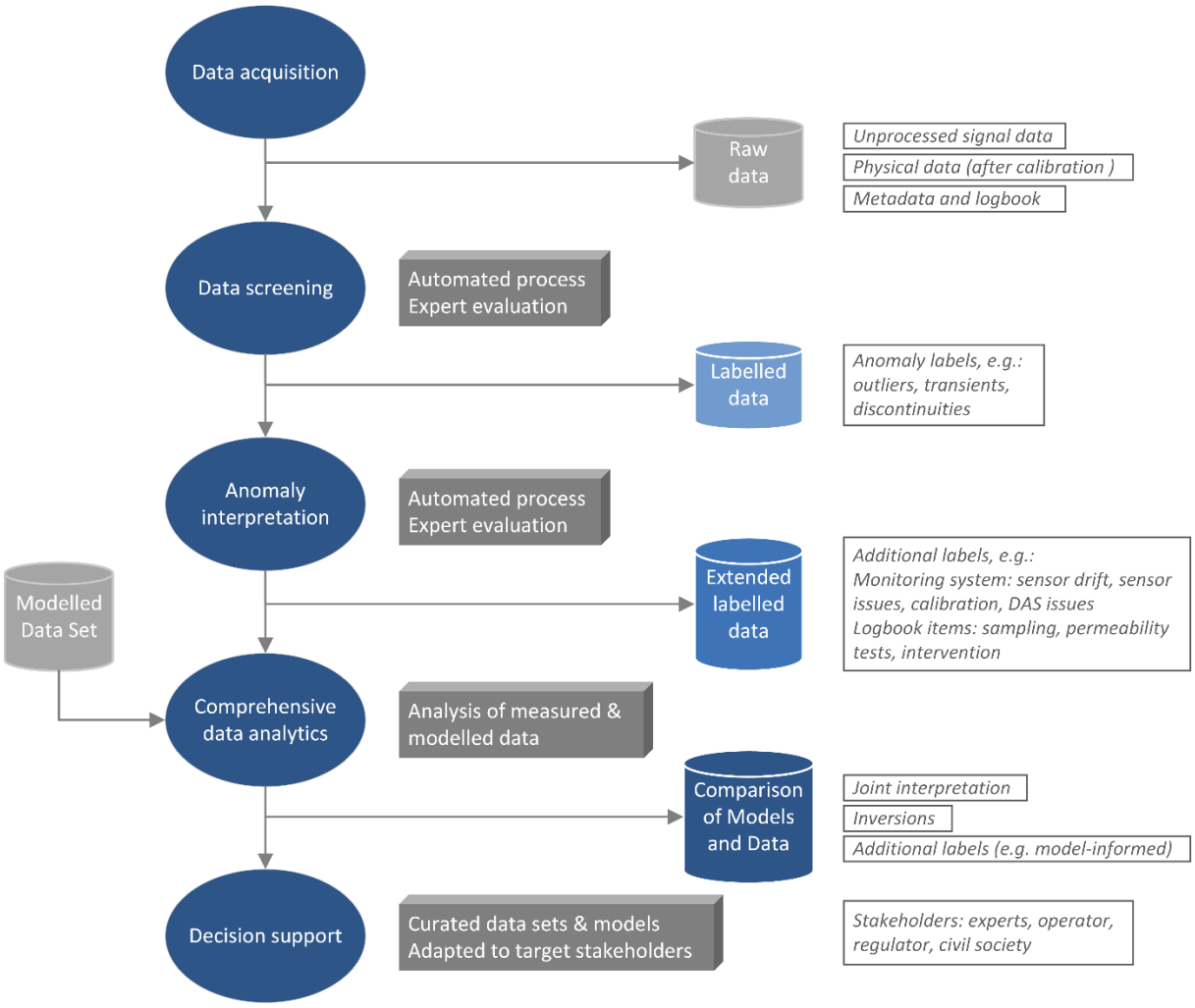


Figure 6.1 - Proposed workflow for data handling from acquisition to decision support.

6.5.1 Data Acquisition

Data acquisition needs to be conducted adhering to quality control standards and good practice. Important insights on monitoring good practice have been assembled in a survey of lessons learned from conducting URL experiments performed within MODATS [94]. Guidance on quality assurance can be found in [95].

Data needs to be stored in a database, that gets the most recent data in regular intervals or through live streams. Intervals of no greater than one day are suggested but shorter intervals will be generally preferable. The database needs to store the unprocessed signal data that are measured by the sensor (e.g. in units of millivolts) as far as reasonable, as well as converted data based on the sensor calibration certificate. Depending on sensor type and manufacturer it could be that the unprocessed signal data is not available, i.e. the output is already a product of a calibration performed in the laboratory. In both cases a complete definition of the sensor response characteristics including the calibration certificates needs to be stored as part of the metadata.

A complete set of metadata that describes the installed sensor network and any changes made to it during the monitoring period need to be maintained throughout and stored alongside the raw data. It is also important to store pertinent information on the overall system (e.g. repository operations affecting the monitored data) in a logbook for later reference. This logbook information is considered to be raw (meta) data and thus should be created and maintained close to real-time for consistency.

Raw data should be kept for at least the duration of the monitoring programme and not modified in any kind of way.

6.5.2 Data screening

Based on the raw data, a (periodical) basic data screening should occur. The goals of the data screening are to evaluate the general quality of the raw data. Specifically, anomalies such as sensor failure, outliers, or gaps and discontinuities in the data should be labelled. Because the type of anomalies to be identified in this step is fairly easy to identify, often an automated process may be suitable for this type of screening. In addition, expert evaluation could be employed to verify the automated processes or add additional labels.

The product of the data screening step is a set of labelled data that identifies anomalies in the monitoring data that are not associated with the system monitored but are spurious errors of the monitoring system. The premise of adding labels to the data is to identify sections of the data that are deemed reliably and to identify sections of the data for which confidence is reduced and which shall be excluded from further interpretation. Importantly, the process of labelling does not alter the data in any way and all data is maintained throughout the monitoring programme.

6.5.3 Anomaly interpretation

Further analysis steps are more involved to identify further anomalies that may not be as obvious as the ones identified in the data screening step. Again, the goal is to provide additional labels that help to classify the data and lay the foundation for further quantitative analysis. While the data screening step involved identification of easily recognisable issues with the data, now the goal is to verify the data more in-depth and identify potential issues with the data that are less obvious, that could still impact system understanding. While the potential for automated processing to aid in such tasks is great, currently this kind of labelling requires expert intervention.

Examples of issues that should be identified are issues with sensors such as drift, calibration issues or other unexpected behaviour. Effects related to the data acquisition system may be identified, and labelled accordingly, already in the previous step or may be detected through more in-depth analysis in this step.

Further, labels related to events from the logbook such as specific interventions that will impact the system behaviour, should be created. Applying a well-specified format for the metadata facilitates the implementation of automated processes that aid in the labelling of such events in the data.

6.5.4 Comprehensive data analytics

The labels created in the previous steps provides a quality-controlled data set (“extended labelled data”) for which confidence is created to conduct further analysis steps and ultimately lead into decision support. In addition to the monitoring data, modelled data sets need to be used in order to evaluate the monitoring data. Labelled data can then be treated properly, depending on the specific requirements of the analysis. While some methods are robust to outliers, others may not be. If labelled data points were to be removed for specific analyses, it needs to be decided whether to fill the gap or leave as a blank value. Again, not all analysis techniques are behaving the same way and a sensible filling of data gaps might be necessary.

Modelled data should provide the baseline of our understanding of the monitored system. Hence, a meaningful comparison of those with the monitored data would lead to an understanding whether the system behaviour is in conformance with predictions. In order to get to a meaningful comparison, it is important to understand the quality of the data and the limitations of the numerical models. The first three steps described above ensure that the data has been adequately quality controlled.

Numerical models are always an approximation of the real system. Differences may be due to approximations in the input data such as simplified time histories of modelled loads, discretization of the

spatial domain and of the distribution of material parameters. Further, there may also be (coupled) processes that are not captured by the numerical model. Hence, there will always be some level of disagreement between the results of numerical modelling and monitoring data. This disagreement can also be used to create additional labels and potentially identify (or confirm) faulty sensors through physical reasoning. Comparisons of modelled and measured data thus need to consider possible sources of differences before determining conformance.

Advanced analysis may include computational methods such as inversions. They have their own sources of uncertainty and may include biases from the analyst. For example, the specific choice of parameters, e.g. damping parameters, used to control the convergence of a solver will lead to different results and probability distributions of the target solution. Typically, inversions deal with non-uniqueness problems, which means that there are several configurations in the solution space that equally fit the observation.

6.5.5 Decision support

The ultimate goal of monitoring efforts is to aid in decision-making. To that end, suitable, curated data sets and models are to be created. There are different decisions that will need to be made, by different stakeholders during the repository lifecycle. This will range from decisions associated with the way in which disposal is implemented, including modifications to initial plans, to major decisions to progress from one stage of the programme to the next. A range of stakeholders will be involved in, or interested in, these decisions, and will require different information to address the questions that they have. It is important to tailor the presentation of materials with the target audience or specific stakeholders in mind. However, it is stressed here that decisions will be informed by the safety case, and the monitoring programme can be considered as a component of the safety case. In this instance, monitoring data will feed into decision making through the safety case, rather than directly.

7. Conclusions and Further Work

Author: Matt White (Galson Sciences)

7.1 Conclusions

The conclusions from the work are summarised below.

Data Processing

- The work in MODATS has allowed the identification and categorisation of monitoring data anomalies, description of the characteristics that define each type of anomaly, and identification of options for their processing.
- MODATS has developed a range of tools for undertaking monitoring data processing and these have been made available online.
- Data processing should be undertaken with reference to user requirements; processing does not have just one end goal; data sets after processing will differ according to the user requesting the data.
- Currently, data processing requires expert user checking of outcomes.
- Data processing should not remove data from the database but flag data so that it can be identified and used as appropriate.

Data Storage

- Monitoring data should be stored with the user in mind, rather than being stored from the perspective of data acquisition.
- Repository monitoring will be undertaken over long periods, so flexibility is required in monitoring databases, and regular upgrades to software and hardware should be planned.

Modelling and Visualisation

- The work in MODATS has demonstrated the potential for hybrid models that combine physics-based and data-driven-modelling to provide a basis for analysis of monitoring data during repository operation.
- The work in MODATS has also demonstrated benefits of communication through visualisation, including benefits to expert users and to university students. Communication to non-expert actors using virtual tours should include provision of contextual information, task-based activities and independent control of the tour to maximise learning.

Digital Twins

- A repository digital twin is a virtual model of part of a repository that is updated automatically to address specific objectives.
- A digital twin of a repository cannot replicate the full reality of the underground (i.e. including all processes relevant to post-closure safety), as the repository system is subject to irreducible uncertainty. Instead, digital twins should focus on specific aspects of a repository (for example a twin of the underground infrastructure) defined by the objectives of the digital twin project.
- Repository digital twins are not “one size fits all”, but come in different forms depending on the objective for which they are developed.
- In MODATS, algorithms have been developed for surrogate models and for implementing physics-informed machine learning, representing prototypes of repository digital twins.
- Future digital twins might help in global parameter sensitivity analysis and related parameter uncertainty quantification, and the digital twins produced in MODATS have provided an illustration of how such digital twins might be developed.
- A comprehensive review of the potential applications of digital twins in repository programmes is required to identify the ways in which digital twins can practically enhance repository programmes.

- Further use of the MODATS Reference Experiments, including development of their digital twins and their underlying surrogate models is required to establish the manner in which digital twins can be applied during repository monitoring.

7.2 Further Work

MODATS has provided significant advancements in the understanding of how monitoring data can be managed, modelled and visualised. The advancements in MODATS are particularly relevant to monitoring of the near-field during repository operations, but the lessons can be applied more widely (i.e., monitoring during all phases of the repository lifecycle, and all data management and modelling work). Nonetheless, further development of approaches to data management and modelling are warranted, especially in the emerging field of the use of digital twins in repository monitoring. Possible further work is described below, with the discussion separated into:

- Application Theme: research into the use of digital twins in repository programmes to identify the benefits that could be delivered through their development and application.
- Technology Theme: further development of the capabilities of digital twins, especially the prototype digital twins developed in MODATS.

Possible work in these two themes is discussed below.

7.2.1 Application Theme

The application theme could include further consideration and a review of the potential applications of digital twins in repository programmes. This would provide a reference document for understanding how different types of digital twins could be developed and applied across the broad spectrum of activities required to support geological disposal of radioactive waste. It would also support a more evolved and more harmonised understanding of the digital twin concept.

This review could be coupled with specific research into the application of digital twins associated with monitoring programmes. The outcome would be to understand the needs of digital twins linked to repository monitoring programmes, e.g., exploring how much data is required by the digital twin, would the associated data requirements impact on the passive safety of a repository, and how would information supplied by digital twins support periodic updates to the safety case during operations (i.e., address uncertainties in the safety case recognised during the licence application)?

As digital twins are still an emerging technology, there is a lack of clear understanding about the value they can bring to the geological disposal field. There is a lack of case studies of successful practices. Extension of the MODATS Reference Experiments and development of digital twins and their underlying surrogate models could be used to communicate the manner in which digital twins can be applied during repository monitoring, and to identify the benefits and challenges of doing so.

There is also a need to communicate the potential for digital twins to different stakeholders. This would include developing guidance/understanding on the challenges and difficulties associated with using digital twins for discussions with civil society (for example, explaining the complex science associated with PBMs and surrogate models, and the different modelling approaches undertaken in safety assessment calculations used to support safety case arguments).

7.2.2 Technology Theme

Current digital twin development is looking at sub-systems of the repository, whereas the ultimate ambition of digital twins could be to develop representations of the entire underground system (even if such digital twins would not be able to include all processes occurring underground). This would require research into the aggregation of digital twins into a single entity (so-called meta digital twin), or in other words, a digital twin made up of other digital twins that represent various aspects of the system. For example, current work in MODATS is looking at digital twins of single drifts or deposition modules;

research is required to consider how to integrate hundreds of these individual models into a meta digital twin.

The digital twin concept has the opportunity to change how we view system design, manufacturing and operation, and to augment systems engineering approaches. In a repository context, digital twins have the potential to inform optimisation processes and support changes to design during the operational period. This would require code and algorithm development for computational efficiency leading to the deployment of Exascale Computing[9] (See [96] for a discussion of the term ExaScale Computing).

To apply digital twins in support of monitoring programmes requires further developments in the integration of DDMs and PBMs. The digital twin has to be calibrated, and an "ignorance" model will be crucial to assess the root cause of the divergence of data from values anticipated by the early versions of the PBMs and DDMs.

Deployment of digital twins also requires the development of data standards to improve the feasibility to develop more comprehensive systems, and the improvement of numerical models for computational efficiency. Example of the avenues through which computational efficiency of numerical models might be improved include improving parallelisation schemes and using novel hardware developments, especially to manage multiphysics couplings on increasingly integrated systems.

[9] Exascale computing refers to the use of technologies to accelerate computing by 1,000X.

References

1 M. White and S. Scourfield (2019). Modern2020 Project Synthesis Repository Monitoring: Strategies, Technologies and Implementation. Work Package 6, Deliverable No D6.5.

2 EC (2004). Thematic Network on the Role of Monitoring in a Phased Approach to Geological Disposal of Radioactive Waste. European Commission Project Report EUR 21025 EN.

3 MoDeRn (2013). Monitoring During the Stages Implementation of Geological Disposal: The MoDeRn Project Synthesis. MoDeRn Deliverable D6.1.

4 Bertrand J., Haines, T., White, M. (2023): Final State-of-the-Art on Monitoring in Radioactive Waste Repositories in Support of the Long-Term Safety Case. Deliverable D17.2 of the HORIZON 2020 project EURAD. EC Grant agreement no: 847593.

5 White M., Haines, T., Bertrand, J. (2024): MODATS WP Synthesis: Confidence in Monitoring Data. Final version as of 31.05.2024 of deliverable D17.9 of the HORIZON 2020 project EURAD. EC Grant agreement no: 847593.

6 International Atomic Energy Agency (2001). Monitoring of Geological Repositories for High level Radioactive Waste. IAEA-TECDOC 1208.

7 International Atomic Energy Agency (2011): Disposal of Radioactive Waste. IAEA Safety Standards Series No. SSR-5, IAEA, Vienna.

8 International Atomic Energy Agency (2011): Geological Disposal Facilities for Radioactive Waste. IAEA Safety Standards Series No. SSR-14, IAEA, Vienna.

9 International Atomic Energy Agency (2014): Monitoring and Surveillance of Radioactive Waste Disposal Facilities. IAEA Safety Standards Series No. SSG-31, Vienna.

10 MoDeRn (2013). Case Studies. MoDeRn Deliverable D4.1.

11 OECD NEA (2022). Repository Library - A Report of the Radioactive Waste Repository Metadata Management (RepMet) Initiative. NEA Report NEA/RWM/R(2019)4. https://oecd-nea.org/jcms/pl_66908/repository-library-a-report-of-the-radioactive-waste-repository-metadata-management-repmet-initiative.

12 M. White, J. Farrow, S. Scourfield, C. Vivalda, C. Espivent, B. Frieg, M. Jobmann, M. Morosini, A. Simeonov, S. Norris (2019): Responding to Monitoring Results. Modern2020 Deliverable D2.3.

13 J.T. Birkholzer and A.E. Bond (2022). DECOVALEX-2019: An International Collaboration for Advancing the Understanding and Modelling of Coupled Thermo-Hydro-Mechanical-Chemical (THMC) Processes in Geological Systems. International Journal of Rock Mechanics and Mining Sciences, Volume 154, 105097.

14 J.T. Birkholzer, C.-F. Tsang, A.E. Bond, J.A. Hudson, L. Jing, O. Stephansson (2019). 25 years of DECOVALEX - Scientific Advances and Lessons Learned from an International Research Collaboration in Coupled Subsurface Processes. International Journal of Rock Mechanics and Mining Sciences, Volume 122, 103995.

15 Development of Coupled Models and their Validation Against Experiments (DECOVALEX), https://decovalex.org/, accessed in 2023.

16 Kolditz O, Jacques D, Claret F, Bertrand J, Churakov SV, Debayle C, Diaconu D, Fuzik K, Garcia D, Graebling N, Grambow B, Holt E, Idiart A, Leira P, Montoya V, Niederleithinger E, Olin M, Pfingsten W, Prasianakis NI, Rink K, Samper J, Szöke I, Szöke R, Theodon L & Wendling J (2023). Digitalisation for nuclear waste management: predisposal and disposal. Environmental Earth Sciences, 82(1), 42 (11 pp.). https://doi.org/10.1007/s12665-022-10675-4.

17 D.C. Montgomery, E.A. Peck (1992). Introduction to Linear Regression Analysis. https://www.worldcat.org/title/23900933.

18 R.H. Shumway, D.S. Stoffer (2017). Time Series Analysis and Its Applications. https://doi.org/10.1007/978-3-319-52452-8.

19 R.J. Hyndman, G. Athanasopoulos (2018). Forecasting: Principles and Practice. OTexts. https://research.monash.edu/en/publications/forecasting-principles-and-practice-2.

20 I. Goodfellow, Y. Bengio, A. Courville (2016). Deep Learning. MIT Press.

21 S. Hochreiter, J. Schmidhuber (1997). Long Short-Term Memory. Neural Computation 9 (November): 1735–80. https://doi.org/10.1162/NECO.1997.9.8.1735.

22 F.A. Gers, J. Schmidhuber (2001) LSTM Recurrent Networks Learn Simple Context-Free and Context-Sensitive Languages. IEEE Transactions on Neural Networks 12 (November): 1333–40. https://doi.org/10.1109/72.963769.

23 J. Gamboa (2017). Deep Learning for Time-Series Analysis. January. https://arxiv.org/abs/1701.01887v1.

24 H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.-A. Muller (2018). Deep Learning for Time Series Classification: A Review. Data Mining and Knowledge Discovery 33 (September): 917–63. https://doi.org/10.1007/s10618-019-00619-1.

25 L. Breiman, (2001). Random Forests. Machine Learning 45 (October): 5–32. https://doi.org/10.1023/A:1010933404324/METRICS.

26 C. Cortes, V. Vapnik, L. Saitta (1995). Support-Vector Networks. Machine Learning 1995 20:3 20 (September): 273–97. https://doi.org/10.1007/BF00994018.

27 C.E. Rasmussen, C.K.I. Williams (2005). Gaussian Processes for Machine Learning. Gaussian Processes for Machine Learning, November. https://doi.org/10.7551/MITPRESS/3206.001.0001.

28 K. Murphy (2022). Probabilistic Machine Learning. Gaussian Processes for Machine Learning, 1–272. https://www.bibguru.com/b/how-to-cite-gaussian-processes-for-machine-learning/.

29 F. Chinesta, E. Cueto, E. Abisset-Chavanne, J.L. Duval, F. El Khaldi (2020). Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data. Archives of Computational Methods in Engineering 27 (January): 105–34. https://doi.org/10.1007/S11831-018-9301-4/METRICS.

30 T. West, M. Blackburn (2017). Is Digital Thread/Digital Twin Affordable? A Systemic Assessment of the Cost of DoD's Latest Manhattan Project. In: Complex Adaptive Systems, Chicago, USA.

31 L. Wright, S. Davidson (2020). How to Tell the Difference Between a Model and a Digital Twin. Advanced Modeling and Simulation in Engineering Sciences Volume 7, Article Number: 13 https://doi.org/10.1186/s40323-020-00147-4.

32 G. Hu, W. Pfingsten (2023). Data-Driven Machine Learning for Disposal of High-Level Nuclear Waste: A Review. Annals of Nuclear Energy 180 (2023) 10945215; https://doi.org/10.1016/j.anucene.2022.109452.

33 G. Hu, W. Pfingsten (2023). Machine Learning-Assisted Heat Transport Modelling for Full-Scale Emplacement Experiment at Mont Terri Underground Laboratory. International Journal of Heat and Mass Transfer 213 124290; https://doi.org/10.1016/j.ijheatmasstransfer.2023.124290.

34 J.F. Abel, M.S. Shephard (1979). An Algorithm for Multipoint Constraints in Finite Element Analysis. International Journal for Numerical Methods in Engineering 14 (January): 464–67. https://doi.org/10.1002/NME.1620140312.

35 C. Farhat, C. Lacour, D. Rixen (1998). Incorporation of Linear Multipoint Constraints in Substructure Based Iterative Solvers. Part 1: A Numerically Scalable Algorithm. International Journal for Numerical Methods in Engineering 43 (November): 997–1016. https://doi.org/https://doi.org/10.1002/(SICI)1097-0207(19981130)43:6<997::AID-NME455>3.0.CO;2-B.

36 M.W. Mahoney, P. Drineas (2009). CUR Matrix Decompositions for Improved Data Analysis. Proceedings of the National Academy of Sciences of the United States of America 106 (January): 697–702. https://doi.org/10.1073/PNAS.0803205106/SUPPL_FILE/APPENDIX_PDF.PDF.

37 R. Ibáñez, E. Abisset-Chavanne, A. Ammar, D. González, E. Cueto, A. Huerta, J.L. Duval, F. Chinesta (2018). A Multidimensional Data-Driven Sparse Identification Technique: The Sparse Proper Generalized Decomposition. Complexity 2018. https://doi.org/10.1155/2018/5608286.

38 K.E. Atkinson (1989). An Introduction to Numerical Analysis. John Wiley & Sons, New York 693p.

39 U.M. Ascher, L.R. Petzold. (1998). Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. Society for Industrial and Applied Mathematics 314p. https://doi.org/10.1137/1.9781611971392.

40 K. He, X. Zhang, S. Ren, J. Sun (2016). Identity Mappings in Deep Residual Networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9908 LNCS: 630–45. https://doi.org/10.1007/978-3-319-46493-0_38/TABLES/5.

41 K. He, X. Zhang, S. Ren, J. Sun (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December (December): 770–78. https://doi.org/10.1109/CVPR.2016.90.

42 P.C. Blaud, P. Chevrel, F. Claveau, P. Haurant, A. Mouraud (2023). ResNet and PolyNet Based Identification and (MPC) Control of Dynamical Systems: A Promising Way. IEEE Access 11: 20657–72. https://doi.org/10.1109/ACCESS.2022.3196920.

43 C. Ghnatios, X. Kestelyn, G. Denis, V. Champaney, F. Chinesta (2023). Learning Data-Driven Stable Corrections of Dynamical Systems—Application to the Simulation of the Top-Oil Temperature Evolution of a Power Transformer. Energies 2023, Vol. 16, Page 5790 16 (August): 5790. https://doi.org/10.3390/EN16155790.

44 A. Chatterjee (2000). An Introduction to the Proper Orthogonal Decomposition. Current Science 78: 808–17. http://www.jstor.org/stable/24103957.

45 G. Berkooz, P. Holmes, J.L. Lumley (1993). The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows. Annual Review of Fluid Mechanics 25 (November): 539–75. https://doi.org/10.1146/ANNUREV.FL.25.010193.002543.

46 H.R. Müller, B. Garitte, T. Vogt, S. Köhler, T. Sakaki, H. Weber, H., T. Vietor (2017). Implementation of the Full-Scale Emplacement (FE) Experiment at the Mont Terri Rock Laboratory. Swiss Journal of Geosciences, 110(1), 287–306. https://doi.org/10.1007/s00015-016-0251-2.

47 Nagra (2019). Implementation of the Full-scale Emplacement Experiment in Mont Terri: Design, Construction and Preliminary Results. Nagra Technical Report NTB 15-02.

48 B. Lanyon, B. Firat Lüthi, E. Manukyan (2019). Interpretation of the First 5 Years of the FE Experiment: a THM Synthesis. Nagra Arbeitsbericht (Working Report) 19-46.

49 D. Jaeggi, M. Iten, F. Fischli, P. Ruppert, P. Gisiger, P. Tabani, M. Herfort, J. Hansmann, L.P. Wymann, P. Bossart, P. (2020). Long-Term Performance of Fiber-Optic Sensors Subjected to In-Situ Conditions in the Opalinus Clay of the Mont Terri Rock Laboratory Synthesis Report. Mont Terri Technical Report 2017-03.

50 H.R. Fisch, B. Firat Lüthi, A. Reinicke, T. Sakaki (2019). Modern2020 WP4.4, Full-Scale Emplacement (FE), Experiment, Mont Terri (Switzerland) – Field Realisation. Nagra Arbeitsbericht (Working Report) NAB 19-32.

51 M.B. Hausner, F. Suárez, K.E. Glander, N.v.d. Giesen, J.S. Selker S.W. Tyler (2011). Calibrating Single-Ended Fiber-Optic Raman Spectra Distributed Temperature Sensing Data. Sensors 2011, 11, 10859-10879. https://doi.org/10.3390/s111110859.

52 M. Kim, S. Lee, J. Jeon *et al.* (2019). Sensitivity Analysis of Bentonite Buffer Peak Temperature in a High-Level Waste Repository. Annals of Nuclear Energy, 2019(123): 190-199.

53 T. Bergman, A. Lavine, F. Incropera *et al.* (2011). Fundamentals of Heat and Mass Transfer. John Wiley & Sons, April 12, 2011.

54 A. Andrés, M. Paul, P. Ivan et al. (2021). FE-Modelling Task Force/Task 1: Validation of Thermally Induced THM Effects in the Rock around the FE-Tunnel. NAGRA NAB 19-40.

55 G. Hu, H. Zhang, Q. Liu (2022). Design Optimization on Characteristics of Packed-Bed Thermal Energy Storage System Coupled with High Temperature Gas-Cooled Reactor Pebble-Bed Module. Energy Conversion and Management, 2022(257): 115434.

56 G. Hu, N. Prasianakis, S.V. Churakov, W. Pfingsten (2023 submitted). Performance Analysis of Data-Driven and Physics-Informed Machine Learning Methods for Thermal-Hydraulic Processes in Full-Scale Emplacement Experiment.

57 N. Graebling, Ö.O. Şen, L. Bilke, T. Cajuhi, D. Naumov, W. Wang, G. Ziefle, D. Jaeggi, J. Maßmann, G. Scheuermann, O. Kolditz, K. Rink (2022). Prototype of a Virtual Experiment Information System for the Mont Terri Underground Research Laboratory. Frontiers in Earth Science.

58 N. Graebling, M. Althaus Ö.O. Şen, T. Reimann, T. Cajuhi, G. Scheuermann, O. Kolditz, K. Rink (2023). Feels like an Indie Game – Evaluation of a Virtual Field Trip Prototype on Radioactive Waste Management Research for University Education. IEEE CG&A.

59 Buchwald et al. (2023). Design-of-Experiment (DoE) based History Matching for Probabilistic Integrity Analysis: A Case Study of the FE-Experiment at Mont Terri. Reliability Engineering and System Safety (submitted).

60 S. Bhatia, P. Cozzi, A. Knyazev, T. Parisi (eds.) (2017). glTF 2.0 Specification. in Tech. Rep. (Oregon, USA: Khronos Group).

61 O. Kolditz, S. Bauer, L. Bilke, N. Böttcher, J.O. Delfs, T. Fischer, U. J. Görke, T. Kalbacher, G. Kosakowski, C. I. McDermott, C. H. Park, F. Radu, K. Rink, H. Shao, H. B. Shao, F. Sun, Y. Y. Sun, A. K. Singh, J. Taron, M. Walther, W. Wang, N. Watanabe, Y. Wu, M. Xie, W. Xu, B. Zehner (2012). OpenGeoSys: An Open-Source Initiative for Numerical Simulation of Thermo-Hydro-Mechanical/Chemical (THM/C) Processes in Porous Media. Env Earth Sci 67, 589–599.

62 K. Rink, L. Bilke, O. Kolditz (2014). Visualisation Strategies for Environmental Modelling Data. Environ. Earth Sci. 72, 3857–3868. doi:10.1007/s12665-013-2970-2.

63 J. Ahrens, B. Geveci, C. Law (2005). ParaView: An End-User Tool for Large-Data Visualization. In Visualization Handbook, 717–731. doi:10.1016/B978-012387582-2/50038-1.

64 K. Rink, O.O. Şen, M. Hannemann, U. Ködel, E. Nixdorf, U. Weber, U. Werban, M. Schrön, T. Kalbacher, O. Kolditz (2022). An Environmental Exploration System for Visual Scenario Analysis of Regional Hydro-Meteorological Systems. Computers and Graphics, 103, 192–200.

65 K. Rink, O.O. Şen, M. Schwanebeck, T. Hartmann, F. Gasanzade, J. Nordbeck, S. Bauer, O. Kolditz (2022). An Environmental Information System for the Exploration of Energy Systems. Geotherm. Energy 10, 1–16. doi:10.1186/S40517-022-00215-5.

66 P. Bossart, F. Bernier, J. Birkholzer, C. Bruggeman, P. Connolly, S. Dewonck, (2018). Mont Terri Rock Laboratory, 20 Years of Research: Introduction, Site Characteristics and Overview of Experiments. In Mont Terri Rock Laboratory, 20 Years, eds. P. Bossart and A. G. Milnes (Birkhäuser, Cham), chap. 1.1 edn., 3–22.

67 L.W. Anderson, D.R. Krathwohl (2001). Taxonomy for Learning, Teaching, and Assessing, A: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Addison Wesley Longman.

68 J. Nielsen (1994). Enhancing the Explanatory Power of Usability Heuristics. In Proc. ACM CHI'94 Conf., Boston, MA, 1994, pp. 152–158.

69 J.R. Lewis, B.S. Utesch, D.E. Maher (2013). UMUX-LITE – When There's no Time for the SUS. In Proceedings of CHI 2013, ACM, Paris, pp. 2099–2102.

70 J.R. Lewis, B.S. Utesch, D.E. Maher (2015). Investigating the Correspondence Between UMUX-LITE and SUS Scores. In: Design, User Experience, and Usability: Design Discourse. Lecture Notes in Computer Science, Springer, Cham., pp. 204–211.

71 J. Brooke (1996). SUS: A Quick and Dirty Usability Scale. In Usability Evaluation in Industry, London, UK, Taylor & Francis, pp. 189–194.

72 J. Sauro, J.R. Lewis (2016). Quantifying the User Experience: Practical Statistics for User Research. 2nd ed. Cambridge, MA, USA: Morgan Kaufmann.

73 Graebling et al. (2023). Converter for Transforming Measurement Data to VTK Unstructured Grids. doi:10.5281/zenodo.10017851.

74 J. Buchwald, A.A. Chaudhry, K. Yoshioka, O. Kolditz, S. Attinger, T. Nagel (2020). DoE-Based History Matching for Probabilistic Uncertainty Quantification of Thermo-Hydro-Mechanical Processes around Heat Sources in Clay Rocks. Int J Rock Mech Min Sci, 134.

75 J. Buchwald, S. Kaiser, O. Kolditz, T. Nagel (2021). Improved Predictions of Thermal Fluid Pressurization in Hydro-Thermal Models based on Consistent Incorporation of Thermo-Mechanical Effects in Anisotropic Porous Media. Int J Heat Mass Transf, 172:121127.

76 E. Holt and P. Koho (2016). D4.5 POPLU Experimental Summary Report. DOPAS Project Deliverable D4.5.

77 C. Andersson, I. Bárcena, N. Bono, L. Boergesson, P. Cleall, T. Forsmark, D. Gunnarsson, L.-E. Johannesson, A. Ledesma, L. Liedtke, A. Luukkonen, K. Pedersen, I. Puigdomenech, R. Pusch, I. Rhén, T. Rothfuchs, T. Sandén, J.-L. Sineriz, Y. Sugita, C. Svemar, H. Thomas (2005). Full-Scale Testing of the KBS-3V Concept for the Geological Disposal of High-Level Radioactive Waste Prototype Repository. European Commission Report EUR 21924. https://cordis.europa.eu/docs/projects/files/FIKW/FIKW-CT-2000-00055/projrep-prototype-repository_en.pdf.

78 OECD NEA (2018). Metadata for Radioactive Waste Management (RepMet) Initiative. NEA Report 7378. https://www.oecd-nea.org/jcms/pl_15062.

79 OECD NEA (2022). Site Characterisation Library - A Report of the Radioactive Waste Repository Metadata Management (RepMet) Initiative. NEA Report NEA/RWM/R(2019)2. https://www.oecd-nea.org/jcms/pl_66894/site-characterisation-library-a-report-of-the-radioactive-waste-repository-metadata-management-repmet-initiative.

80 OECD NEA (2022). Waste Package Library - A Report of the Radioactive Waste Repository Metadata Management (RepMet) Initiative. NEA Report NEA/RWM/R(2019)3. https://www.oecd-nea.org/jcms/pl_66896/waste-package-library-a-report-of-the-radioactive-waste-repository-metadata-management-repmet-initiative.

81 OECD NEA (2022). RepMet Tools and Guidelines - A Report of the Radioactive Waste Repository Metadata Management (RepMet) Initiative. NEA Report NEA/RWM/R(2019)5. https://www.oecd-nea.org/jcms/pl_66909/repmet-tools-and-guidelines-a-report-of-the-radioactive-waste-repository-metadata-management-repmet-initiative?details=true.

82 S. Seabold and J. Perktold (2010). Statsmodels: Econometric and Statistical Modeling with Python. Proceedings of the 9th Python in Science Conference.

83 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, É. Duchesnay (2011). Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research.

84 W. McKinney (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference.

85 C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, T.E. Oliphant (2020). Array Programming with NumPy. Nature. https://doi.org/10.1038/s41586-020-2649-2.

86 P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272.

87 M.A. Haque (2022). Feature Engineering and Selection for Explainable Models: A Second Course for Data Scientists. Lulu Press, Inc.

88 Ph. Van Marcke, X.L. Li, W. Bastiaens, J. Verstricht, G. Chen, J. Leysen, J. Rypens (2013). The Design and Installation of the PRACLAY In-Situ Experiment. EURIDICE Report 13-129, 2013.

89 M. D. Wilkinson, M. Dumontier, I. Jan Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J-W. Boiten, L. Bonino da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J.G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S-A Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data, 3:160018. DOI: 10.1038/sdata.2016.18.

90 K. Shaukat, T.M. Alam, S. Luo, S. Shabbir, I.A. Hameed, J. Li, S.K. Abbas, U. Javed (2012). A Review of Time-Series Anomaly Detection Techniques: A Step to Future Perspectives. In book: Advances in Intelligent Systems and Computing, Publisher: Springer, DOI: 10.1007/978-3-030-73100-7_60.

91 S. Gaugel, M. Reichert (2023). PrecTime: A Deep Learning Architecture for Precise Time Series Segmentation in Industrial Manufacturing Operations. Engineering Applications of Artificial Intelligence Volume 122, 106078. https://doi.org/10.1016/j.engappai.2023.106078.

92 T. Wen, R. Keyes. Time Series Anomaly Detection Using Convolutional Neural Networks and Transfer Learning. https://www.zurich.ibm.com/AI4IoT/2019/AI4IoT-19_Wen.pdf.

93 S. Levasseur, X. Sillen, P. Marschall, J. Wendling, M. Olin, D. Grgic, J. Svoboda (2022). EURADWASTE'22 Paper Host Rocks and THMC Processes in DGR. EPJ Nuclear Sci. Technol. 8, 21. https://doi.org/10.1051/epjn/2022021.

94 T. Haines and M. White. (2022): Lessons for Repository Monitoring from Underground Research Laboratory Experiments. Deliverable D17.3 of the HORIZON 2020 project EURAD. EC Grant agreement no: 847593.

95 M. White, T. Haines, Y. Caniven and J. Verstricht (2024): Guidance on Quality Assurance Project Plans for Repository Monitoring. Deliverable D17.4 of the HORIZON 2020 project EURAD. EC Grant agreement no: 847593.

96 P. Kogge, K. Bergman, S. Borkar, D. Campbell, W. Carlson, W. Dally, M. Denneau, P. Franzon, W. Harrod, K. Hill, J. Hiller, S. Karp, S. Keckler, D. Klein, R. Lucas, M. Richards, A. Scarpelli, S. Scott, A. Snavely, T. Sterling, R.S. Williams, K. Yelick (2008): ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems.