Model Order Reduction methods for sensor data assimilation to support the monitoring of embankment dams

by

Christina Nasika

Supervisors:
Prof. Pedro Díez
Prof. Pierre Gerard
Prof. Thierry J. Massart
Prof. Sergio Zlotnik

A thesis submitted to the

Building, Architecture and Town Planning Department
École Polytechnique de Bruxelles
Université libre de Bruxelles

and

Departament d'Enginyeria Civil i Ambiental
Universitat Politècnica de Catalunya

in partial fulfilment of the requirements for the degree of
Doctor in Engineering Sciences

Barcelona, March 2022

Abstract

The latest monitoring and asset management technologies for large infrastructures involve digital representations that integrate information and physical models, exist in parallel to the real-life structures, and are continuously updated based on assimilated sensor data, in order to accurately represent the actual conditions in the structures. This type of technology is often referred to as Digital Twin. The implementation of such cutting-edge technology in monitoring assets like tailings dams, or embankment dams in general, and other large structures, implies the development of highly efficient numerical tools that, combined with sensor data, may support rapid, informed decision making.

For the particular case of embankment dams, enabling this type of technology requires an efficient numerical model that describes the coupled hydro-mechanical phenomena, pertinent to a dam functioning and safety. This may for instance be a Finite Elements (FE) model, describing thde groundwater flow through unsaturated porous geomaterials.

The process of updating and calibrating a model, such as the above mentioned FE model, based on sensor data is typically referred to as data assimilation. Often, this is achieved via an optimization approach, where a specific problem is solved multiple times for various parametric values, in search for the values that best describe the sensor data. The bottleneck in this type of application is typically the cost of multiple evaluations of the model, that may become prohibitive when the underlying FE model is large. In order to enable such applications, the present work proposes Model Order Reduction (MOR) methods tailored to the hydro-mechanical nonlinear problem at hand.

MOR aims at the creation of a surrogate model that seeks an approximation of the FE solution in a reduced-order space. This is achieved by applying an offline-online strategy. In the offline stage, the solution manifold of the

full-order problem is sampled, in order to identify a low-order affine subspace, where an accurate approximation of the full-order solution can be captured. To tackle the nonlinearities related to partially saturated conditions in the soil, a similar strategy must be employed in order to define reduced-order spaces where an affine system approximation may be recovered. The resulting Reduced Order Model (ROM) may be used as an efficient surrogate to the FE model in any problem that requires fast and/or repetitive solutions.

In this work, MOR techniques are implemented to solve the coupled nonlinear transient problem under consideration. ROMs are created to solve problems that pertain to tailings dams and embankment dams monitoring. The efficiency and the accuracy of these models are demonstrated by solving inverse problems for parametric identification. MOR is found to be a reliable tool, significantly accelerating the inverse identification process while resulting to accurate solutions.

Acknowledgements

Firstly, I would like to acknowledge the financial support I received from the European Union's Hor'izon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 764636, for the research project ProTechTion.

I would like to express my gratitude to my supervisors, Pedro Díez, Sergio Zlotnik, Thierry J. Massart and Pierre Gerard for their continuous support and guidance through these years. They were always there when I needed their help, and they could always offer valuable insights and propose great ideas. Their trust, encouragement and flexibility were a great relief through the challenging times of Covid. Thanks to their supervision, my PhD journey has been an enjoyable and very positive experience. I would also like to thank Péter Berke for his support and the time he spent offering his insights when I was stuck and needed external input. Finally, I thank Professor Sebatian Olivella and Núria Pinyol for their much appreciated assistance in this work.

I would like to thank Fabiola for being the best colleague and friend one could ask for. Moreover, I will never forget the time I spent with many dear colleagues - Giacomo, Jonatha, Alex, Caterina, Alice, Luca, Davide, Ygee, Arash, Chaimae, Rohith, Roland, Batoma, Karim. They turned the working place into a place of fun, a place of friends, shared lunches in the sun, games, laughter, and there is no price for that. I would also like to thank all my friends that supported me one way or another through this journey, Maria P., Maria X. Anna Maria S., Anna Maria G., Vaso, Eleni, Stella, Dorina, Giorgos, Khaoula, Marco, Theo, Aggelos. Some of them I knew before, and I've had the luck to meet some of them during this journey. This thesis would surely have been impossible without them.

Lastly, I dedicate this thesis to my family, who have provided me with

endless love and support through this PhD and always. I thank my parents, Eleni and Thanos, and my sister Asimina, for listening to me, encouraging and believing in me.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and state of the art

Tailings are a common by-product of the process of extracting valuable minerals and metals from mined ore. They usually take the form of a liquid slurry made of mineral particles, created as mined ore is crushed, ground and processed. The volume of tailings is normally far in excess of the liberated resource and the tailings often contain potentially hazardous contaminants. It is usual practice for tailings to be stored in isolated impoundments under water and behind dams [60]. These are called tailings dams, and they are usually earthfill embankment dams.

Tailings dams are some of the largest structures built by geotechnical engineers [68].The current rate of major tailings dam failure is much higher than that of major failures of water-retaining earthfill dams [26] [68]. The amount and the toxic nature of materials held within tailings dams means that their failure can have devastating consequences on the surrounding river systems, water aquifers, aquatic and human life, for potentially hundreds of kilometers downstream. One of the most catastrophic tailings dams collapses that caused 270 deaths and the release of about 12 million m^3 of tailings, occurred in 2019 in Brazil. In Figure 1.1, the large volume of toxic material that was released in the surrounding area in the form of liquid mud is shown.

Figure 1.1: Slurry released in Brumadinho dam collapse (25/1/2019, Brazil)

One of the reasons for the high failure rate of these structures may be the fact that mine waste facilities do not represent a source of profit for mine owners, which places constraints to the resources allocated to impoundment monitoring and maintenance [54]. This also implies that all the failure modes in tailings dams have not been as extensively investigated as in water-retaining earthfill dams [58].

The mining industry, along with many industrial sectors, is turning to digital technologies to seek solutions that may upgrade the functioning and monitoring of its facilities [13, 8]. This trend drives the exploration of new technologies that make use of the increasing data availability to transform mining practices. As part of this trend the Digital Twin approach has emerged as a technology that may offer benefits in the optimization of a mining facility functioning throughout its life-cycle.

A Digital Twin is a virtual counterpart of a real-life object or asset, such as a tailings dam. It is a comprehensive digital representation, consisting of information regarding the properties, materials, conditions of the asset, and models that can simulate its real-life behaviors [44].

The physical asset and its digital representation exist in parallel and influence each other in a reciprocal manner. All the changes that occur to the physical asset and the information regarding its condition, that is, sensor observations of important physical quantities, is broadcast and integrated into its digital counterpart. The Digital Twin in turn, incorporates numerical tools that permit simulating different scenarios, obtaining predictions of crucial physical quantities that cannot directly be measured, and analyzing and comparing incoming data to previously collected observations. Moreover, the input parameters of the underlying model may be calibrated against the incoming sensor data, in a process that is called data assimilation. The insight obtained by the Digital Twin results into informed decision making about the way the asset is managed, and eventually it leads to changes and interventions carried out on the actual structure. This concept of mutual feedback that forms the central idea of Digital Twins technology is illustrated in Figure 1.2.

This work is motivated by, and oriented toward, this type of integrated technology. The essential goal of this work is not to design a Digital Twin per se, but rather to develop numerical tools that aim to enable an integrated safety monitoring system, that combines sensor data and a numerical model of the monitored dam.

Tailings dams technical characteristics and failure modes

There are key differences between tailings dams and conventional water dams. Tailings dams are usually constructed by readily available local materials and they are often built of and/or on tailings material [101]. Compacted tailings material is used to build the embankment, behind which, uncompacted tailings may be stored. Tailings grain size is highly variable and dependent

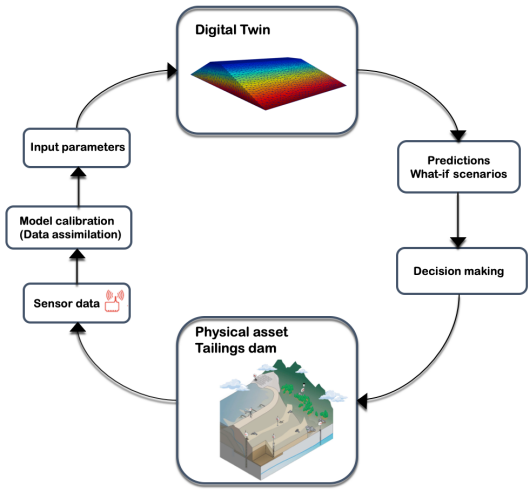


Figure 1.2: Digital Twin technology for tailings dams monitoring. A continuous feedback is implemented between the physical embankment and its digital representation throughout the life cycle of the asset.

on the parent rock and the method of extraction. They tend however to be largely gravel-free and clay-free, with sand being more common than silt [60]. Their chemical composition depends on the mineralogy of the ore body, the degree of weathering during storage and the extraction process. Silica, iron and oxygen display an almost universal presence and seem to be the most abundant elements in tailings [60]. Tailings material is pumped from the mill to the impoundment and is often size differentiated during deposition. The coarser and more porous material settles close to the discharge point, near the embankment, and may be used to extend the structure itself, while the finer fraction (slimes) is carried further away forming an impermeable barrier. This size-differentiated dispersal contributes to the integrity of the dam [60].

Rather than constructed at once, tailings dams are gradually raised, as the mining activity results to larger capacity demands for the storing reservoir. Three different methods are used for embankment level raise, namely the upstream, downstream, and centerline method, as illustrated in Figure 1.3. The brown-colored part in the Figure represents the part of the structure that has to be constructed from coarse and compacted material, and therefore the most expensive part to be built. The upstream method, requiring the smallest volume of processed fill material, and therefore being the most cost effective, is most widely chosen [68], but it is associated to many major failures. In

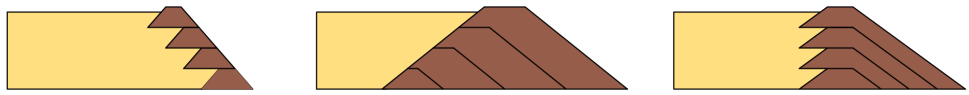


Figure 1.3: Tailings dam construction method: conceptual illustration. Left to right: Upstream, Downstream, Centreline level raise. Yellow color: Tailings. Brown color: Fill material.

upstream dams, the additional layers of structure are placed on top of tailings depositions, therefore the construction's safety depends on the integrity of tailings for stability [26]. This type of dam requires tighter ongoing scrutiny [72].

Impoundment failure can be categorized into four basic mechanisms. These main mechanisms are, overtopping, often occurring in inactive structures after a flooding event; localized failure, caused by the presence of a shear band; piping, caused by internal erosion due to seepage; and diffuse failure, triggered by liquefaction of loose tailings material [45, 64]. The common factors in most cases are the importance of the stress and seepage fields on the dam. The seepage field might directly induce instability by erosion, or cause the pore pressure to rise, leading to a reduction of effective stress and shear strength.

Slope instability -that corresponds to the second failure mechanism local failure - is one of the most common failure modes [74]. In upstream tailings dams, the risk of instability of the downstream slope should be particularly investigated during design and monitoring. The stability of upstream dams may depend on the stability of the tailings depositions upstream, that support the overlying layers of the embankment. In case the tailings deposited have a high water content, a fast loading rate that allows no time for consolidation may lead to excess pore pressure and loss of resistance to shear failure. This may occur during the fast construction of new layers of compacted tailings, in the stage of raising the dam level.The new layers of structure do not provide a stabilizing force to the downstream slope, as is the case with centerline and downstream raising methods. The timescales of loading, i.e. raising the dam level and consolidation of the underlying material, are governing the hydromechanical response of the system.

Obviously, the stability problem in tailings dams should be treated as a *hydro-mechanical problem*, as it is the excess of pore pressure that causes the loss of shear strength of the material and consequently its local failure. High pore pressures, and high hydraulic gradients in the case of erosive failures (piping), are omnipresent elements in tailings dams failures. Pore pressure monitoring is therefore crucial in the case of tailings dams.

Sensor data monitoring of embankment dams

The usual monitoring practice in tailings dams involves periodic visits to the site and manual collection of instrument readings. These visits may occur as rarely as once or twice per year, depending on site conditions and allocated resources.

The most important monitoring instruments in tailings dams are those that measure pore pressure and water level. Among different types of piezometers, vibrating-wire ones are compatible with most remote sensing technologies. They are typically placed in multiple locations in the structure and measure positive pore pressure with an accuracy of $\pm 0.1 - 0.5\%$ FS. Pressure monitoring is essential, as it can help spot potential risks while they are still developing and manageable, that is, before deformation and damage occurs [116].

Displacements in tailings dams are measured on the surface with GPS type systems, and in depth with inclinometers. Vertical inclinometers are installed in boreholes and measure angle (tilt) with respect to initial placement, that can be translated to horizontal displacement. A typical accuracy is limited to 0.01mm/m. In-place inclinometers might be combined with settlement systems that measure vertical displacement in the same position where horizontal displacement is measured, with an accuracy of millimeters. Displacement per se, may be non-conclusive for integrity assessment, and is often used to calculate quantities such as displacement rate and shear strain [116].

The technology developments in remote sensor data acquisition have introduced novel possibilities for monitoring [25]. The use of Internet of Things (IoT) technologies is becoming a preferred solution for the assessment of tailings dams safety. In the context of IoT, interconnected devices are used for the continuous collection, sharing and visualization of data that are broadcasted in real time from sensors that are installed in the monitored system -in this case, the dam or mining installation. Real-time sensor monitoring proves to be a key tool for reducing the risk related to these ever-evolving earthfill structures, that exhibit a high rate of sudden and hazardous failures. However,

even though data acquisition has now become easier, the interpretation of said data in real time remains challenging [58, 54].

The context of this work is related to the employment of such a technology in an integrated monitoring system, that would incorporate data analysis and decision making tools. The models developed in this work, estimate as elementary quantities the pore water pressure and displacement. These are related to some of the most essential monitoring instruments in tailings dams and embankment dams in general.

The need for modeling to support sensor data-based decision making in tailings dams

Despite the increasing availability of sensor data, their interpretation remains a challenging task. The health assessment of tailings dams based on data is a particularly daunting task due to the ever-evolving nature of their design, an element that is absent in conventional water dam engineering. In water dams, sensor data can be evaluated by means of comparison with previous observed trends. Alarms can easily be set up for observed pressures, deformations, temperatures etc that differ drastically from previously recorded values that are within the range of serviceability. In addition, in the case of water dams, the geometry of the dam is usually stable and pre-decided. Therefore the ranges of "safe" values of pressure, stress, deformation can be pre-decided once and for good.

In tailings dams the following points render continuous integrated monitoring necessary:

- A common approach in embankment monitoring is to assess data by comparing them to historical data. Expected behaviors are based on previous trends [58, 116]. This is an appropriate method for monitoring sudden changes that indicate potential trouble. However, it is not always an appropriate option for non-static structures like tailings dams. During periods when the dam crest level is raised, these structures undergo significant changes, therefore the resulting stress and pressure states cannot be compared to previous ones. Nonetheless, the induced pore pressures and deformations must remain in the serviceability range. This ever-evolving nature and geometry of these structures gives rise to the need for constant redefinition of the acceptable values for a the stress and pressure fields, and for the identification of existing conditions in the dam.

- It has been shown that failure is often initiated in the beach of depositions rather than in the embankment structure itself [106, 101]. This can be the case both for localized failure (slope instability) and for diffuse failure (liquefaction) of loose and high water content tailings. This implies that the constantly evolving beach of tailings depositions must be monitored, alarm thresholds must be established and updated regularly [116, 64]. The existence of a numerical model that is constantly upgraded and updated to represent the actual dam conditions would immensely facilitate the monitoring [74].

- The material that is stored in tailings dams is newly formed soil material that typically undergoes fast transformations once it is deposited, altering its mechanical and hydraulic properties [116]. Moreover, changes in stress and strain conditions can induce changes in soil properties, due to physical processes that are often not described in numerical models. This may lead to unexpected behavior of the dam. For example, a change in permeability may lead to less favorable drainage conditions in the dam and eventually to failure [68]. Considering these sources of uncertainty, data assimilation, that is, regular calibration of the model based on data becomes crucial.

- There are other sources of uncertainty featured in tailings dams, related to loads and natural phenomena like rainfall, evaporation, and seismicity. Combining sensor data and numerical modeling can help to understand better the state of the structure, in the presence of these uncertainties.

Combining the sensor network with a numerical model that simulates the physical processes that are pertinent to the safety of an earthfill dam, can facilitate the interpretation of large volumes of sensor data [108]. In the case of earthfill dams, a coupled hydro-mechanical problem that describes the groundwater flow of water through soil governs the dam integrity. The methods that are developed in this work, are related to the establishment of a numerical model that can evaluate in quasi real time stress, strain and water pressure states of the embankment dam. Moreover, this work is oriented toward the development of data assimilation techniques that may allow the combination of sensor data and numerical model for the continuous updating of the model, so that the prediction accuracy increases through the life of the structure, and manages to stay up to date with the evolving conditions in the real-life dam.

1.2 Model Order Reduction

One of the bottlenecks for sensor monitoring in a quasi real-time context is the lack of a numerical model providing real-time or close to real-time responses. A model would be required with a response that can be obtained faster than the physical evolution of the phenomenon to be analyzed.

The need for high efficiency is even more relevant in the framework of sensor data assimilation for model optimization. The reason for this is that data assimilation implies the use of methods that involve repetitive model evaluations, that is, they require the solution of many-query problems.

When the model evaluation is based on a Finite Element (FE) model that requires the solution of a problem with a potentially very large number of unknowns, the computational cost of these methods becomes prohibitive. This is why methods that are aimed to significantly decrease the model response time are often used in the solution of many-query problems [100, 88].

Model Order Reduction has been a fast developing scientific field over the last decades, due to the increased need for computational efficiency in numerical models. High dimensional problems in engineering and scientific computing remain challenging despite the increase of computational resources. The hardest scenarios to tackle involve problems that require repetitive direct solutions, often referred to as many query problems (i.e. in optimization, inverse parameter identification, uncertainty quantification), and large-scale problems that require very fast solution (i.e. close to real-time simulations). The former group of applications implies the definition of a parametric problem, where a number of input parameters such as the geometric configuration of the underlying domain, the material properties or the boundary conditions can vary within a range of possible values.

The central idea in Model Order Reduction for mesh-dependent numerical methods (Finite Elements (FE), Finite Differences etc) lies in the identification of a low order space where an accurate approximation of the full-order solution of a parametric problem lives.

Considering a mesh discretization that yields N degrees of freedom, by associating an approximation function to each degree of freedom, an approximation space where the discrete solution of the problem lives is defined. This is referred to as full-order or high-fidelity solution space. The problem is solved by computing N values. If the set of equations is *transient* and *nonlinear*, as it is in the case that is studied in this work, then a N-sized system of equations must be solved at least once for every time step.

Model Order Reduction is an umbrella term for several techniques that

acknowledge the fact that in many cases, the solution of a numerical model can be approximated in a space of much smaller dimension than N. The objective is then to identify this low order space and construct a Reduced Order Model (ROM) that seeks for an approximation of the high-fidelity solution in that low-order space. The low order approximation unavoidably features some error with respect to the full-order solution, since the low-order space cannot possibly capture all the information related to the full-order approximation space.

For the construction of the ROM, the full-order solutions must be obtained either for a small period of time in the case of a transient problem or for a small number of parametric values. In any case, the constructed ROM should then allow to solve similar problems (i.e. for parametric values that have not been sampled) with high efficiency.

Proper Orthogonal Decomposition

Proper Orthogonal Decomposition (POD) based methods, that will be used in this work, are an instance of Model Order Reduction methods that requires the solution of the full-order problem at least once. The Reduced Order Model (ROM) is extracted from this solution or set of solutions, that will eventually allow the solution of many similar problems with a much smaller cost. Due to the requirement of a number of full-order solutions to construct the ROM, POD-based methods are often called *a posteriori* [24].

Approaching a problem with a POD-based ROM is performed in two stages:

- Offline stage: An expensive pre-process, where the low-order approximation space is identified, by "sampling" the full-order solution space.
- Online stage: A much cheaper solving process, where approximate solutions can be computed. This stage is repeated as many times as a real-time response is required in the context of a many-query problem.

Specifically, in the present work, the Reduced Basis method (RB) [70, 93] will be used. This is an instance of MOR that utilizes POD to identify a low order space where an approximate solution can be sought for.

Discrete Empirical Interpolation

In the presence of nonlinearities in the system of PDEs at hand, the POD technique reduces the problem dimension in the sense that far less unknowns

are present in the system solving process. However, the cost of evaluating nonlinear terms still requires a projection to the full order space, and is therefore still related to the size of the original, high-fidelity problem [21]. In the FE context, this limitation refers to the high computational complexity related to the repetitive calculation of inner products required to evaluate weak forms that contain general nonlinearities.

The Discrete Empirical Interpolation Method (DEIM) [21] follows a POD-based strategy to reduce the cost of evaluating such nonlinearities. It defines low-order approximation spaces for each nonlinear term, in addition to the POD reduced space where the solution of the problem is approximated. The nonlinear terms approximations are constructed by *projection* onto the POD basis and *interpolation* based on selected grid points.

To further reduce the cost of assembly Localized Discrete Empirical Interpolation (LDEIM) [88] proposes the creation of multiple POD bases, each corresponding to a particular set of system states. These bases are expected to be of even smaller dimension than the global POD bases introduced in DEIM, thus resulting in an even more significant dimension reduction, and an even faster assembly.

There have been several model reduction methodologies applied to the simulation of porous media flow, such as data-driven models [12, 40], Proper Orthogonal Decomposition (POD) based methods [40, 114, 61, 11], as well as POD paired with DEIM (Discrete Empirical Interpolation Method) where the nonlinear terms are approximated by some form of interpolation, ensuring a large reduction of computational cost [41, 39, 29]. The work in all the papers mentioned above is motivated by reservoir and petroleum engineering and often refers to multiphase fluid flow. In the aforementioned papers, POD-based reduction is applied to the hydraulic problem alone, and does not concern the hydro-mechanically coupled problem [40, 41, 114, 119]. A POD-DEIM approach, has not been applied so far to the coupled flow/geomechanics problem to the knowledge of the author of the present.

The selection of the RB and DEIM methods for the work in this thesis was based on the identified gap in the related literature. The specific problem that is described in this thesis, which is a coupled non-linear transient hydro-mechanical problem (as is discussed in Chapter 2) had not been previously approached with this combination of methods, to the knowledge of the author of the present work. However, there are various Model Order Reduction methods that may be employed for this problem.

One commonly used method is Proper Generalized Decomposition (PGD) [23, 24, 133] that offers the advantage that it does not require an offline

preprocessing stage in order to compute the approximate solution. It is therefore referred to as an *a priori* method, because it does not require the prior knowledge of any full-order solutions to the problem. Moreover, it is possible in PGD to compute a general solution, once for life, that includes all the approximate solutions for every possible value of the input, that is, a sort of computational vademecum. After having computed the vademecum, any approximate solution can be obtained in real-time. This is an important advantage of the method in the context of many-query problems.

Another method that has been gaining popularity is the use of Artificial Neural Networks (ANN) to replace the full-order FE model. In that case, the cost of training the model can be large, but once it is trained, the model can output real-time solutions. The main disadvantage of this method with respect to POD and PGD methods is that after the model is trained, the relationship to the physical problem is lost.

1.3 Inverse problem solving and data assimilation

A common challenge in developing realistic models is related to highly uncertain parameters, mainly representing material properties [49]. In the application to tailings dams, the hydraulic and mechanical properties of the structural and stored materials are often unknown and may vary over time [120]. Real soil is a highly nonlinear, heterogeneous material, with many of its mechanical and hydraulic properties depending on its strain, stress and pressure state [17, 79]. In the context of modern technologies like digital twins for monitoring and control of large infrastructures, this problem becomes particularly important, since the value of these new technologies is dependent on the degree of accuracy of prediction of the underlying models. The idea of monitoring data assimilation to automatically update and calibrate a numerical model is essential to the concept of digital twins [17]. In engineering practice, the choice of appropriate parameters for geomechanical models is either based on laboratory or in-situ experimental explorations, or left to the judgment of experienced engineers. These methods have some obvious limitations, as experimental campaigns may fail to provide realistic in-situ values. There is often a lack of sampling, the effect of changes in the mechanical state of the material is not accounted for, and the heterogeneity and temporal uncertainty of material properties are neglected.

To treat such problems efficiently, data assimilation may prove useful. In this context, field measurements are used for back-analyses in order to

identify realistic values for the properties used in the model. Data assimilation, implies the solution of an inverse problem, that is, a problem that examines multiple solutions for different values of the parameters, and identifies the most realistic parametric values by driving the model output as close to the reality as possible.

Inverse problem solving for soil properties identification has been extensively explored in the literature. In geotechnical engineering, the inferred parameters are usually mechanical and strength properties, and the inversion is based on displacement or deformation, field or experimental data [102, 130, 97, 27, 111]. Hydraulic parameter identification using inverse modeling has also been explored, mainly in the context of agricultural engineering. The applications are based on either real [55, 96, 113, 28, 43] or synthetic experimental data [59, 126, 110, 134, 65].

There is extensive discussion in the literature on the subject of the inherent ill-posedness of the inverse problem for hydraulic parameter identification, that may result in non-unique and/or unstable solutions [110, 59, 113, 43, 71, 134]. The work in these publications is not motivated by earthfill dam applications, rather they generally treat the problem of water flow through soil. Specifically, authors have investigated the uniqueness of the inverse problem solution, when hydraulic soil properties must be identified simultaneously.

In this work, a deterministic approach is adopted for inverse problem solving. Parameter identification is performed by minimizing a function that measures the discrepancy between sensor data and model output. The minimization process is a many-query problem, since solutions for various parametric values must be computed until the values that yield minimum discrepancy are identified. Therefore, the computational cost of this process is very high, and depending on the complexity of the underlying problem, may be prohibitive. To remedy this issue, one method that has been examined by several authors is the use of artificial neural networks (ANN), trained to approximate the relation between a permeability field and the pore water pressure distribution of the domain for a specific problem [35, 112, 132]. Nonetheless, the use of ROMs as surrogates to full-order FE models in the process of parameter identification may as well present a solution to the issue of computational efficiency.

The research gap that this thesis aims to address, is a comprehensive study of parameter identification via inverse problem solving based on sensor data from actual earthfill dams. A study that addresses the various issues that may arise in a real-life application based on actual sensor data, such as the problem's ill-posedness when solving for multiple parameters simultaneously,

and the computational efficiency issue, remains to be done, to the knowledge of the authors of the present.

1.4 Scope and objectives

In the scope of creating models of embankment dams with close to real-time responses, the main objective pursued in this work is related to the application of Model Order Reduction strategies for coupled nonlinear hydro-mechanical problems. To achieve this, this thesis is oriented toward answering the following research questions:

- Is it possible to obtain accurate ROMs using POD-based reduction methods for solving transient nonlinear problems in geomechanics?
- What is the order of magnitude of the time speedup achieved with POD-based reduction for hydro-mechanical problems?
- What is the additional gain that can be obtained by introducing hyper-reduction with DEIM?
- Does the inverse problem for identification of mechanical and hydraulic parameters via objective function minimization based on sensor data have a unique and stable solution?
- Can the inverse problem be solved with ROMs with high accuracy?
- Are the developed methods applicable to inverse problems based on actual pore water pressure data measured on existing embankments?

Creating a model to support decision making in real-time for embankment dams is a complex task, some aspects of which are addressed in this work. There is a wide range of issues that are open and require further examination. Some are mentioned in the concluding chapter. The scope of this thesis is limited to the study of certain areas, defined by the following assumptions and simplifications:

- A coupled problem that considers mechanics and the hydraulic flow of groundwater through soil is considered. These physical processes were deemed most pertinent to the safety of an embankment dam. Other possibly interesting processes, like thermal effects and transfer of pollutants are neglected in this work.

- The nonlinearity investigated here is related to the description of the partially saturated soil. The author focuses on the investigation of Model Order Reduction for this particular source of nonlinearity. For the mechanical constitutive modeling, linear elasticity is therefore considered.
- The reported results that pertain to the efficiency and accuracy of POD-based ROMs are problem-specific. Even though they demonstrate an expected behavior and showcase the capabilities of the studied methods, different results may be recovered when different hydro-mechanical problems are tackled.
- The above comment also applies for the results reported in inverse problem solving.

It is worth mentioning that even though the engineering application that motivates this work is related to the safety monitoring of tailings dams using real-time remote sensing, the methodologies that have been developed can be applied to earthfill embankment dams in general, and some of the problems that are studied in the present work, are therefore related to conventional water earthfill dams, for which experimental measurements are readily available.

1.5 Contributions and outline of the thesis

The main contributions of this thesis are related to achieving drastic reduction in the computational costs in solving coupled hydro-mechanical problems of groundwater flow through unsaturated soil. Furthermore, a parametric identification method is developed based on inverse problem solving and using a ROM.

In Chapter 2, the equations governing the coupled hydro-mechanical groundwater flow through unsaturated soil are presented, and the Finite Element model created for this problem is described. Moreover, the development of the Reduced Order Model is described and its performance is evaluated. The reduction method is applied to a 2D and a 3D illustrative problems, designed to simulate realistic critical conditions in the safety monitoring of tailings dams. This example illustrates the reduction in time that is achieved by solving the problem with a ROM instead of FEM. The novel element introduced in this chapter is the implementation of the Reduced Basis Method for coupled hydro-mechanical nonlinear problems.

In Chapter 3, the methodology is extended to the specific treatment of nonlinear problems with DEIM and LDEIM methods. The methods are

described and implemented and the resulting model performance is examined. A 3D problem is used to illustrate the methods advantages in terms of computational efficiency and accuracy of the resulting model. Several details regarding the implementation of the method are discussed. The ROMs that result from applying various combinations of the RB, DEIM and LDEIM methods and varying the size of reduced approximation spaces are compared. The methods are employed for the solution of parametric problems featuring one, and three parameters, in order to illustrate the ROMs performance with scaling of the parametric domain. Moreover, the one-parameter problem is solved on a significantly denser mesh, in order to examine the methods scaling capabilities with increasing full-order problem size. Similarly to Chapter 2, the novelty introduced here is related to the implementation of DEIM and LDEIM in the coupled hydro-mechanical problem of groundwater flow through unsaturated soil. A significant boost in computational is observed in most cases, with no compromise in accuracy.

In Chapter 4, parameter identification based on sensor data for earthfill dams is described and some results are presented. This chapter discusses the feasibility of parametric identification with optimization algorithms and a pressure-based objective function. A ROM is employed as a surrogate for the commonly used FEM that proves to be impractical in the context of most optimization algorithms, due to its high computational cost. The novel contributions introduced in this chapter are firstly, the use of an objective function that includes a normalization term based on the time derivative of the pore water pressure, and the implementation of objective function optimization-based parameter identification, for both mechanical and hydraulic soil properties, on nonlinear coupled problems using ROMs. The method is employed to identify soil properties based on actual measured pore pressure sensor data recorded on an existing earthfill water dam during a rapid drawdown event.

Finally, in Chapter 5 the conclusions of this study are drawn and some observations and comments related to the topics that have been studied are discussed. The limitations of this work are summarized, and the topics that remain to be explored in future works to address the limitations, and extend the work that has been completed here, are outlined.

Chapter 2

The Reduced Basis method for coupled hydro-mechanical problems

---

This chapter is based on:
C. Nasika, P. Díez, P. Gerard, T.J. Massart, S. Zlotnik, Towards real time assessment of earthfill dams via Model Order Reduction, Finite Elements in Analysis and Design. 199 (2022) 103666. https://doi.org/10.1016/j.finel.2021.103666.

2.1 Introduction

In this Chapter the adopted physical model, the full-order FE model and the Model Order Reduction methodology are presented. A ROM is created and employed in solving an illustrative problem that simulates critical conditions in a tailings dam.

The original contribution proposed in this Chapter, lies in the implementation of POD-based reduction for the coupled nonlinear problem, considering partial saturation of the porous medium. Coupling of the two equations that govern the mechanical and hydraulic part of the problem, yields a nonlinear transient system of equations, that is discretized using the Finite Element Method. Nonlinearities are introduced in order to describe partially saturated states for the soil. Partial saturation must be considered in the study of earthdams, as it is a common occurrence in these structures that the water table is located below the dam crest, and therefore only part of the material is saturated. Solving the discretized system of equations, one can obtain the full-order or high fidelity approximation of the solution to the PDEs. The Reduced Basis method [31, 70, 69, 100] represents an instance of model order reduction techniques in which the parametric dependence of the PDE solution is explored by solving the high-fidelity problem a number of times. The resulting set of solutions is explored in a POD framework, in order to find a set of basis functions, hopefully fewer in number than the dimension of the full-order problem, the linear combination of which can provide a satisfactory approximation of the high-fidelity solution.

To adapt this technique to problems in partially saturated soils, this Chapter is structured as follows. Sections 2.2 and 2.3 contain a detailed description of the methodologies used for developing the forward FE model and the low-order approximation model using RB. In Section 2.4 the accuracy and computational efficiency achieved with ROM are demonstrated solving a problem related to the construction of tailings dams. The results are discussed and an outline of future developments that could add value to the present work is given in section 2.5.

2.2 The physical model describing groundwater flow through unsaturated porous media

In this Section, the equations that govern the hydro-mechanically coupled problem of groundwater flow through an unsaturated soil are presented.

The equations written here describe various geomechanical problems that feature water flow through unsaturated porous media, and the methodologies developed in this work can be used to solve these problems.

Constitutive relations and governing equations

The equation of mechanical equilibrium reads

$$\tilde{\boldsymbol{\nabla}}^{\intercal}\boldsymbol{\sigma} + \rho(p)\mathbf{g} = \mathbf{0}, \tag{2.1}$$

where $\boldsymbol{\sigma} = [\sigma_x, \sigma_y, \sigma_z, \tau_{xy}, \tau_{xz}, \tau_{yz}]^{\intercal}$ is the vector of total stresses, the differential operator $\tilde{\boldsymbol{\nabla}}$ for the general 3D case is defined as,

$$\tilde{\boldsymbol{\nabla}}^{\intercal} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 & \frac{\partial}{\partial y} & 0 & \frac{\partial}{\partial z} \\ 0 & \frac{\partial}{\partial y} & 0 & \frac{\partial}{\partial x} & \frac{\partial}{\partial z} & 0 \\ 0 & 0 & \frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{bmatrix}, \tag{2.2}$$

$\mathbf{g} = [0, 0, -g]^{\intercal}$ is the gravity acceleration vector, and ρ is the density of the multiphase medium, comprised of soil particles and water, evaluated as a function of pore water pressure p, and related to the density of soil particles and water (ρ_{s}, ρ_{w}) according to the relation

$$\rho(p) = (1-\eta)\rho_{\mathrm{s}} + \eta \mathrm{S_e}(p)\rho_{\mathrm{w}} = (1-\eta)\rho_{\mathrm{s}} + \Theta(p)\rho_{\mathrm{w}}, \tag{2.3}$$

where η denotes the soil porosity. The volume water content (VWC) $\Theta(p)$ and effective degree of saturation, or dimensionless water content $\mathrm{S_e}(p)$ [79] are evaluated according to a hydraulic model detailed in Section 2.2

In this work the air pressure is considered equal to the atmospheric pressure, as commonly assumed in geotechnics. The constitutive stress is defined as

$$\boldsymbol{\sigma}' = \boldsymbol{\sigma} - \mathrm{S_e}(p)p\tilde{\mathbf{I}}, \tag{2.4}$$

where $\boldsymbol{\sigma}' = [\sigma'_x, \sigma'_y, \sigma'_z, \tau'_{xy}, \tau'_{xz}, \tau'_{yz}]^{\intercal}$ is the vector of effective stresses, and $\tilde{\mathbf{I}} = [1, 1, 1, 0, 0, 0]^{\intercal}$ is a column vector with 1 at normal stress entries and 0 at shear stress entries.

Linear elasticity is assumed for the soil skeleton's response. In that framework, the constitutive stress-strain relation reads,

$$\boldsymbol{\sigma}' = \mathbf{D}_{\mathrm{el}}\boldsymbol{\varepsilon}(\mathbf{u}), \tag{2.5}$$

where $\mathbf{u}$ is the displacement vector, $\varepsilon = [\varepsilon_x, \varepsilon_y, \varepsilon_z, \gamma_{xy}, \gamma_{xz}, \gamma_{yz}]^\intercal$ denotes the infinitesimal strain vector, calculated as $\varepsilon(\mathbf{u}) = \frac{\boldsymbol{\nabla}\mathbf{u}+\boldsymbol{\nabla}^\intercal\mathbf{u}}{2}$,E is the Young's modulus and ν the Poisson's ratio. $\mathbf{D}_{\text{el}}$ is the elastic stress-strain matrix, which for the general 3D case is defined as,

$$\mathbf{D}_{\text{el}} = \frac{\text{E}}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix}. \tag{2.6}$$

Introducing the stress-strain relation, the mechanical equilibrium can be written in the form

$$\tilde{\boldsymbol{\nabla}}^\intercal \left(\mathbf{D}_{\text{el}}\varepsilon(\mathbf{u}) + \text{S}_{\text{e}}(p)p\tilde{\mathbf{I}} \right) + \rho(p)\mathbf{g} = \mathbf{0}. \tag{2.7}$$

Let $\Gamma_{\text{D}}^{\text{u}}$, $\Gamma_{\text{N}}^{\text{u}}$ and be two partitions of the boundary $\partial\Omega$ of the domain Ω on which Dirichlet and Neumann boundary conditions are applied respectively. The boundary conditions are,

$$\mathbf{u} = \hat{\mathbf{u}} \quad \text{on } \Gamma_{\text{D}}^{\text{u}}, \tag{2.8}$$

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \hat{\mathbf{t}} \quad \text{on } \Gamma_{\text{N}}^{\text{u}}, \tag{2.9}$$

where $\mathbf{n}$ is the outward pointing unit normal vector along $\partial\Omega$, and $\hat{\mathbf{t}}$ is the imposed surface traction.

Considering the mass balance of pore fluids leads to the continuity equation for flow, stating that the water outflow from a representative elementary volume is equal to the changes in mass concentration. Neglecting the deformations of solid particles due to effective stress and pore pressure, as well as the density gradients of water, introducing the Darcian definition for fluid velocity, the strong form of the continuity equation reads

$$\boldsymbol{\nabla}^\intercal \left[\frac{\text{k}(p)}{\gamma_{\text{w}}} (\boldsymbol{\nabla} p + \mathbf{b}_{\text{w}}) \right] + \left(\text{C}(p) - \frac{\Theta(p)}{\text{K}_{\text{w}}} \right) \dot{p} = \Theta(p)\tilde{\mathbf{I}}\boldsymbol{\nabla} \cdot \dot{\mathbf{u}}, \tag{2.10}$$

where γ_{w} is the specific weight of water, $\mathbf{b}_{\text{w}} = \rho_{\text{w}}\mathbf{g}$ are the body water forces and K_{w} is the water bulk modulus. The hydraulic conductivity $\text{k}(p)$, the

specific moisture capacity $C(p)$, and the volumetric water content (VWC) $\Theta(p)$ are estimated using the soil water retention relations that are presented in Section 2.2.

Typical boundary conditions that arise in the case of earthfill dams may be either of Dirichlet, Newmann or Robin type. For the flow equation, Dirichlet conditions may be used to prescribe a known hydraulic head, Neumann conditions for a known outflow, inflow or a hydraulically closed (impervious) boundary.

A particular case arises in the description of a seepage face, which occurs when a water table touches an open downstream boundary [91, 37]. The length of the seepage surface is pressure-dependent [3], and can be prescribed as a non-linear Robin condition.

Let Γ_D^p, Γ_N^p and Γ_R^p be three partitions of the boundary $\partial\Omega$ of the domain Ω on which Dirichlet, Neumann and Robin boundary conditions are applied respectively, for the flow part of the problem. The boundary conditions are

$$p = \hat{p} \quad \text{on } \Gamma_D^p, \tag{2.11}$$

$$\mathbf{q} \cdot \mathbf{n} = \hat{q} \quad \text{on } \Gamma_N^p, \tag{2.12}$$

$$\mathbf{q} \cdot \mathbf{n} = \langle \beta p \rangle \quad \text{on } \Gamma_R^p \tag{2.13}$$

where $\mathbf{n}$ is the outward pointing normal vector and $\hat{q}$ is the fluid flux on the boundary. Equation (2.13) refers to the seepage condition, where the nonlinear function of p that is denoted with angular brackets prescribes a flux that is equal to βp, when $p > 0$ and vanishes for negative pressure [37]. The coefficient β depends on the hydraulic conductivity and geometry of the domain and defines the water runoff on a boundary in seepage conditions. This is a nonlinear Robin type condition.

Soil water characteristics

The most commonly used hydraulic model for the water content - pore water pressure relation in unsaturated soils is the one proposed by Van Genuchten [115]. The effective saturation S_e -or dimensionless water content- is given by

$$S_e(p) = \begin{cases} \frac{1}{[1+(\alpha|\frac{p}{\gamma_w}|)^{\frac{1}{1-m}}]^m} & p < 0 \\ 1 & p \geq 0 \end{cases}, \tag{2.14}$$

where α is a parameter related to the air entry value of the soil and m is a curve fitting parameter. The upper branch of this equation describes a sigmoid

curve which is called a water-retention curve. The VWC is then given by

$$\Theta(p) = \mathrm{S_e}(p)(\Theta_\mathrm{s} - \Theta_\mathrm{r}) + \Theta_\mathrm{r}, \tag{2.15}$$

where $\Theta_\mathrm{s}, \Theta_\mathrm{r}$ are soil characteristics: the VWC for fully saturated conditions, and the residual VWC. Differentiation of equation (2.15) with respect to pore water pressure gives

$$\mathrm{C}(p) = \frac{\partial\Theta(p)}{\partial p} = \frac{-\alpha\mathrm{m}(\Theta_\mathrm{s} - \Theta_\mathrm{r})}{1-\mathrm{m}}\mathrm{S_e}(p)^{1/\mathrm{m}}(1 - \mathrm{S_e}(p)^{1/\mathrm{m}})^\mathrm{m}. \tag{2.16}$$

The relation between the hydraulic conductivity of the soil-water system and the pore water pressure as proposed by van Genuchten [115] reads,

$$\mathrm{k}(p) = \mathrm{k_s}\sqrt{\mathrm{S_e}(p)}\left[1 - \left(1 - \mathrm{S_e}(p)^{1/\mathrm{m}}\right)^\mathrm{m}\right]^2, \tag{2.17}$$

where $\mathrm{k_s}$ is the hydraulic conductivity for saturated conditions.

In some cases in the following, an alternative hydraulic conductivity-pore water pressure relation is adopted. That is, a commonly used [91, 20] cubic law that reads,

$$\mathrm{k}(p) = \mathrm{k_s}\mathrm{S_e}(p)^3. \tag{2.18}$$

The reason for selecting this alternative relation, is that in one case in the following, an illustrative problem is studied, that has been examined in a previous publication by different authors [91]. In that case, all the details concerning the modeling and physical assumptions were taken in accordance to that work, so that comparison between results and validation may be enabled.

2.3 Model reduction methodology

Finite Element Method for hydro-mechanical groundwater flow problems in unsaturated conditions

A Taylor-Hood element is used to create a stable scheme that satisfies the Ladyshenskaya-Babuska-Brezzi (LBB) condition [10, 16]. Given a domain Ω with boundary $\partial\Omega$ the following function spaces are introduced,

$$L^2(\Omega) = \left\{\mathrm{p} : \int_\Omega |\mathrm{p}^2|\mathrm{dx} < +\infty\right\},$$
$$H^1(\Omega) = \left\{\mathrm{p} : \mathrm{p} \in L^2(\Omega), \mathrm{Dp} \in L^2(\Omega)\right\},$$

$$\boldsymbol{H}(div,\Omega) = \left\{\mathbf{u} : \mathbf{u} \in L^2(\Omega)^d, \boldsymbol{\nabla}\cdot\mathbf{u} \in L^2(\Omega)\right\},$$

where d denotes space dimension. Water pressure p should be sought in function space $H^1(\Omega)$ and displacement $\mathbf{u}$ in $\boldsymbol{H}(div,\Omega)$. The following subset spaces are also defined,

$$W = \left\{\mathrm{w} \in H^1(\Omega) : \mathrm{w}\Big|_{\Gamma_{\mathrm{D}}^{\mathrm{p}}} = 0\right\},$$
$$\boldsymbol{V} = \left\{\mathbf{v} \in \boldsymbol{H}(div,\Omega) : \mathbf{v}\Big|_{\Gamma_{\mathrm{D}}^{\mathbf{u}}} = 0\right\},$$

where $\Gamma_{\mathrm{D}}^{\mathrm{p}}$ and $\Gamma_{\mathrm{D}}^{\mathbf{u}}$ are partitions of the boundary $\partial\Omega$ where Dirichlet boundary conditions are applied for pressure and displacement respectively.

Equation (2.7) is multiplied with a vector test function $\mathbf{v} \in \boldsymbol{V}$, integrated over the domain Ω, and applying the Green-Gauss theorem is discretized applying the Galerkin approach. The discretized equation reads

$$\mathbf{K}\mathbf{u} - \mathbf{Q}\mathbf{p} = \mathbf{f}_{\mathbf{u}}, \tag{2.19}$$

where,

$$\mathbf{K} = \int_\Omega \mathbf{B}_{\mathbf{u}}{}^{\intercal}\mathbf{D}_{\mathrm{el}}\mathbf{B}_{\mathbf{u}}\mathrm{dx},$$
$$\mathbf{Q} = \int_\Omega \mathbf{B}_{\mathbf{u}}{}^{\intercal}\mathrm{S_e}\tilde{\mathbf{I}}\mathbf{N}_{\mathrm{p}}\mathrm{dx},$$
$$\mathbf{f}_{\mathbf{u}} = \int_\Omega \mathbf{N}_{\mathbf{u}}\mathbf{b}\mathrm{dx} + \int_{\Gamma_{\mathrm{N}}^{\mathrm{u}}} \mathbf{N}_{\mathbf{u}}\hat{\mathbf{t}}\mathrm{ds}.$$

$\mathbf{N}_{\mathbf{u}}, \mathbf{N}_{\mathrm{p}}$ are displacement and water pressure shape function matrices respectively, and $\mathbf{u}$ and $\mathbf{p}$ are unknown nodal value vectors for displacement and pressure, such that $\mathbf{u} \approx \mathbf{N}_{\mathbf{u}}\mathbf{u}$ and $\mathrm{p} \approx \mathbf{N}_{\mathrm{p}}\mathbf{p}$, $\mathbf{B}_{\mathbf{u}} = \tilde{\boldsymbol{\nabla}}\mathbf{N}_{\mathbf{u}}$ is the gradient matrix relating displacements to strains.

Similarly, multiplying equation (2.10) by scalar test function $\mathrm{w} \in W$, the variational form of the water flow equation is obtained, assuming no inflow or outflow in the domain. It reads

$$\int_\Omega \frac{\mathrm{k}}{\gamma_{\mathrm{w}}}\boldsymbol{\nabla}p\cdot\boldsymbol{\nabla}\mathrm{w}\mathrm{dx} + \int_{\Gamma_R^p}\langle\beta p\rangle\mathrm{w}\mathrm{dx} + \int_\Omega \Theta\tilde{\mathbf{I}}^{\intercal}\boldsymbol{\nabla}\cdot\dot{\mathbf{u}}\mathrm{w}\mathrm{dx} - \int_\Omega\left(\mathrm{C} - \frac{\Theta}{\mathrm{K_w}}\right)\dot{p}\mathrm{w}\mathrm{dx} =$$
$$= \int_\Omega \frac{\mathrm{k}}{\gamma_{\mathrm{w}}}\mathbf{b}_{\mathbf{w}}\cdot\boldsymbol{\nabla}\mathrm{w}\mathrm{dx}. \tag{2.20}$$

The space discretized form is written as

$$\mathbf{H}\mathbf{p} + \mathbf{C}\dot{\mathbf{u}} - \mathbf{S}\dot{\mathbf{p}} = \mathbf{f}_{\mathrm{p}}, \tag{2.21}$$

where,

$$\begin{aligned}
\mathbf{H} &= \int_{\Omega} \boldsymbol{\nabla}\mathbf{N}_{\mathrm{p}}^{\intercal} \frac{\mathrm{k}}{\gamma_{\mathrm{w}}} \boldsymbol{\nabla}\mathbf{N}_{\mathrm{p}} \mathrm{dx} + \int_{\Gamma^{\mathrm{p}}_{\mathrm{R,p>0}}} \mathbf{N}_{\mathrm{p}}^{\intercal} \beta \mathbf{N}_{\mathrm{p}} \mathrm{ds}, \\
\mathbf{C} &= \int_{\Omega} \mathbf{N}_{\mathrm{p}}^{\intercal} \Theta \tilde{\mathbf{I}}^{\intercal} \mathbf{B}_{\mathbf{u}} \mathrm{dx}, \\
\mathbf{S} &= \int_{\Omega} \mathbf{N}_{\mathrm{p}}^{\intercal} \left(\mathrm{C} - \frac{\Theta}{\mathrm{K}_{\mathrm{w}}}\right) \mathbf{N}_{\mathrm{p}} \mathrm{dx}, \\
\mathbf{f}_{\mathrm{p}} &= \int_{\Omega} \boldsymbol{\nabla}\mathbf{N}_{\mathrm{p}} \frac{\mathrm{k}}{\gamma_{\mathrm{w}}} \mathbf{b}_{\mathrm{w}} \mathrm{dx}.
\end{aligned}$$

To solve equations (2.21) and (2.19), time stepping is implemented by a generalized θ-scheme, which approximates $\mathbf{X}^{\intercal} = [\mathbf{u}\ \mathbf{p}]^{\intercal}$ at time $\mathrm{i} + \theta$ as

$$\dot{\mathbf{X}}^{\mathrm{i}+\theta} \simeq \frac{\mathbf{X}^{\mathrm{i}+1} - \mathbf{X}^{\mathrm{i}}}{\Delta \mathrm{t}}, \quad \mathbf{X}^{i+\theta} \simeq (1-\theta)\mathbf{X}^{i} + \theta \mathbf{X}^{i+1}, \tag{2.22}$$

where $\Delta\mathrm{t}$ is the time step and $\mathrm{i}+1$ denotes the current time step. Parameter θ takes values in $[0, 1]$. It has been proven that an implicit time integration scheme with $\theta \geq 0.5$ results in unconditionally stable solution and allows for the use of large time increments [15]. In this work, the value $\theta = 0.75$ was selected in accordance to [103].

Operators $\mathbf{Q}$, $\mathbf{H}$, $\mathbf{C}$, $\mathbf{S}$, all depend on the pressure state, therefore they must be re-evaluated at each time instance. The same applies for force vectors $\mathbf{f_u}$ and $\mathbf{f}_{\mathrm{p}}$. Solving equations (2.21) and (2.19) at time $\mathrm{i} + \theta$, the system reads:

$$\mathbf{K}\mathbf{u}^{\mathrm{i}+\theta} - \mathbf{Q}^{\mathrm{i}+\theta}\mathbf{p}^{\mathrm{i}+\theta} = \mathbf{f_u}^{\mathrm{i}+\theta}, \tag{2.23}$$

$$\mathbf{H}^{\mathrm{i}+\theta}\mathbf{p}^{\mathrm{i}+\theta} + \mathbf{C}^{\mathrm{i}+\theta}\frac{\mathbf{u}^{\mathrm{i}+1} - \mathbf{u}^{\mathrm{i}}}{\Delta \mathrm{t}} - \mathbf{S}^{\mathrm{i}+\theta}\frac{\mathbf{p}^{\mathrm{i}+1} - \mathbf{p}^{\mathrm{i}}}{\Delta \mathrm{t}} = \mathbf{f}_{\mathrm{p}}^{\mathrm{i}+\theta}. \tag{2.24}$$

The time stepping scheme as presented in equation (2.22) is used for the approximation of operators and vectors that depend on the pressure state. Hence, the operator $\mathbf{Q}$ at time $\mathrm{i} + \theta$ is approximated as,

$$\mathbf{Q}^{\mathrm{i}+\theta} \simeq (1-\theta)\mathbf{Q}^{\mathrm{i}} + \theta \mathbf{Q}^{\mathrm{i}+1}. \tag{2.25}$$

Other operators, $\mathbf{Q}^{i+\theta}$, $\mathbf{H}^{i+\theta}$, $\mathbf{C}^{i+\theta}$, $\mathbf{S}^{i+\theta}$, $\mathbf{f_u}^{i+\theta}$ and $\mathbf{f_p}^{i+\theta}$ are approximated similarly. Introducing these approximations to equations (2.23) and (2.24), the fully coupled discretized system using a monolithic approach reads,

$$\begin{bmatrix} \hat{\mathbf{K}} & \hat{\mathbf{Q}} \\ \hat{\mathbf{C}} & \hat{\mathbf{H}} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{p} \end{bmatrix}^{i+1} = \begin{bmatrix} \hat{\mathbf{f_u}} \\ \hat{\mathbf{f_p}} \end{bmatrix}, \tag{2.26}$$

where the components $\hat{\mathbf{K}}, \hat{\mathbf{Q}}, \hat{\mathbf{C}}, \hat{\mathbf{H}}$ of the global stiffness matrix and the force vector are evaluated as,

$$\hat{\mathbf{K}} = \theta \mathbf{K}, \tag{2.27}$$

$$\hat{\mathbf{Q}} = -\theta(1-\theta)\mathbf{Q}^{i} - \theta^2 \mathbf{Q}^{i+1}, \tag{2.28}$$

$$\hat{\mathbf{f_u}} = -\mathbf{K}(1-\theta)\mathbf{u}^{i} + [\theta(1-\theta)\mathbf{Q}^{i+1} + (1-\theta)^2\mathbf{Q}^{i}]\mathbf{p}^{i} + (1-\theta)\mathbf{f_u}^{i} + \theta \mathbf{f_u}^{i+1}, \tag{2.29}$$

$$\hat{\mathbf{C}} = \theta \mathbf{C}^{i+1} + (1-\theta)\mathbf{C}^{i}, \tag{2.30}$$

$$\hat{\mathbf{H}} = \Delta t(1-\theta)\theta \mathbf{H}^{i} + \Delta t \theta^2 \mathbf{H}^{i+1} - (1-\theta)\mathbf{S}^{i} - \theta \mathbf{S}^{i+1}, \tag{2.31}$$

$$\hat{\mathbf{f}}_p = -[\Delta t(1-\theta)^2 \mathbf{H}^{i} + \Delta t(1-\theta)\theta \mathbf{H}^{i+1}\theta + (1-\theta)\mathbf{S}^{i} + \theta \mathbf{S}^{i+1}]\mathbf{p}^{i} + \\ + [\theta \mathbf{C}^{i+1} + (1-\theta)\mathbf{C}^{i}]\mathbf{u}^{i} + \Delta t(1-\theta)\mathbf{f}_p^{i} + \theta \mathbf{f}_p^{i+1}. \tag{2.32}$$

In this work, the nonlinear system of equations is solved using a Picard iterative scheme, a method that is often used in this type of problem [19] [66]. The method is considered to be more robust compared to the Newton-Raphson method, though it has a slower convergence rate.

The model that is described above implies certain simplifying assumptions. The constitutive equations that are incorporated may not be adequate to realistically describe some problems that arise in earthfill dams. In particular, the hydraulic and mechanical properties of the soil may vary in space and time as a result of physical processes that are not described here. For example, applying mechanical loads to the soil may lead to changes in its mechanical properties [105], and straining the material modifying its porosity may have an effect on hydraulic properties [81]. At a future stage of this study, more involved constitutive relations may be considered.

The model was developed in FEniCS platform [1], a collection of free and open-source software components with the common goal to enable automated solutions of differential equations.

Model Order Reduction: The Reduced Basis Method

In data assimilation problems, where the value of some parameters needs to be determined based on the available information, many queries have to be done to the numerical model. If it is based on a full-order approach (e.g. FE) the computational cost might become prohibitive.

The Reduced Basis (RB) [93, 50, 31] method tries to create a small basis that is able to represent the family of the solutions spanned by the parameters variations. The simplest method to create the RB is to solve the full-order FE problem at a set of parametric values that capture the overall behavior of the solution. Each one of these samples is usually called a snapshot.

Once the parametric space has been sampled, the RB is constructed by orthonormalizing the set of snapshots and discarding those with amplitudes smaller than a certain threshold. This process is usually called off-line, as it is done once, i.e. it can be seen as a pre-process of the data-assimilation.

Once the RB is ready, the solution of the problem for any point in the parametric space can be obtained as a linear combination of the members of the RB. Therefore, the computational cost is largely reduced as the number of unknowns to determine (i.e. the coefficients of this linear combination) is usually several orders of magnitude smaller than the size of the original FE problem. The small problem size allows for a very fast solution that can be done repetitively within the assimilation of data. This fast solution is usually called the on-line phase of the RB method.

In the following, the parameter vector is denoted by $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^P$ where the parameter space $\mathcal{P}$ represents a closed and bounded subset of the Euclidean space $\mathbb{R}^P$, $P \geq 1$. The field variable given by the Finite Element solution of a parametrized PDE can be seen as a map $\mathbf{x} : \mathcal{P} \rightarrow V$, that to any $\boldsymbol{\mu} \in \mathcal{P}$ associates the solution $\mathbf{x}(\boldsymbol{\mu})$ belonging to a suitable functional space V.

The full-order approximation of a PDE for a given $\boldsymbol{\mu} \in \mathcal{P}$ can be represented in the generic form

$$\mathbf{A}(\boldsymbol{\mu})\mathbf{x}(\boldsymbol{\mu}) = \mathbf{f}(\boldsymbol{\mu}), \tag{2.33}$$

where $\mathbf{A}(\boldsymbol{\mu}) \in \mathbb{R}^{N_h \times N_h}$ and $\mathbf{f}(\boldsymbol{\mu}) \in \mathbb{R}^{N_h}$ are a $\boldsymbol{\mu}$ -dependent matrix and vector respectively, representing the stiffness matrix and the force vector. The system has N_h degrees of freedom.

The key idea of RB, is to replace this system with another one, of lower dimension $N_r < N_h$ [93] . For any given $\boldsymbol{\mu} \in \mathcal{P}$, the solution field is approximated as $\mathbf{x}(\boldsymbol{\mu}) \approx \mathbf{B}\boldsymbol{\alpha}(\boldsymbol{\mu})$ and the low-order system reads

$$\mathbf{B}^\intercal \mathbf{A}(\boldsymbol{\mu})\mathbf{B}\boldsymbol{\alpha}(\boldsymbol{\mu}) = \mathbf{B}^\intercal \mathbf{f}(\boldsymbol{\mu}), \tag{2.34}$$

where $(\mathbf{B}^\intercal \mathbf{A}(\boldsymbol{\mu})\mathbf{B}) \in \mathbb{R}^{\mathrm{N_r}\times\mathrm{N_r}}$, $(\mathbf{B}^\intercal \mathbf{f}(\boldsymbol{\mu})) \in \mathbb{R}^{\mathrm{N_r}}$ and $\boldsymbol{\alpha}(\boldsymbol{\mu})$ is the reduced vector of degrees of freedom. The form $\mathbf{B}\boldsymbol{\alpha}(\boldsymbol{\mu})$ represents the approximation of the high-fidelity solution $\mathbf{x}(\boldsymbol{\mu})$, in the low-order space $\mathbb{R}^{\mathrm{N_r}}$, where $\mathbf{B} \in \mathbb{R}^{\mathrm{N_h}\times\mathrm{N_r}}$ is a $\boldsymbol{\mu}$-independent transformation matrix, the columns of which collect the reduced basis vectors.

For time-dependent problems, like the one at hand, the full-order PDE approximation is written in a general form as

$$\mathbf{M}(t;\boldsymbol{\mu})\dot{\mathbf{x}}(t;\boldsymbol{\mu}) + \mathbf{A}(t;\boldsymbol{\mu})\mathbf{x}(t;\boldsymbol{\mu}) = \mathbf{f}(t;\boldsymbol{\mu}), \tag{2.35}$$

where $\mathbf{A}(t;\boldsymbol{\mu}), \mathbf{M}(t;\boldsymbol{\mu}) \in \mathbb{R}^{\mathrm{N_h}\times\mathrm{N_h}}$ are time and parameter-dependent matrices and $\mathbf{f}(t;\boldsymbol{\mu}) \in \mathbb{R}^{\mathrm{N_h}}$ is a vector of $\boldsymbol{\mu}$ and time-dependent data. Considering the approximation of the time derivative $\dot{\mathbf{x}}(t;\boldsymbol{\mu}) \simeq \frac{\mathbf{x}^{i+1}-\mathbf{x}^{i}}{\Delta t}$, the reduced-order approximation of the PDE for any time level $t^i = i\Delta t$, ($\Delta t > 0$ being the time step) reads [50],

$$\mathbf{B}^\intercal\left(\frac{1}{\Delta t}\mathbf{M}(t;\boldsymbol{\mu}) + \mathbf{A}(t;\boldsymbol{\mu})\right)\mathbf{B}\boldsymbol{\alpha}^{\mathrm{i}+1}(\boldsymbol{\mu}) = \mathbf{B}^\intercal\left(\frac{1}{\Delta t}\mathbf{M}(t;\boldsymbol{\mu})\mathbf{B}\boldsymbol{\alpha}^{\mathrm{i}}(\boldsymbol{\mu}) + \mathbf{f}(t;\boldsymbol{\mu})\right). \tag{2.36}$$

The basis creation in the presence of non-homogeneous Dirichlet boundary conditions is treated here by isolating the known boundary degrees of freedom from the unknown values to determine [51]. Thus the reduced approximation of the solution $\mathbf{x}(\boldsymbol{\mu})$ reads,

$$\mathbf{x}(\boldsymbol{\mu}) \approx \begin{bmatrix}\mathbf{B}\\ \mathbf{0}\end{bmatrix}\boldsymbol{\alpha}(\boldsymbol{\mu}) + \begin{bmatrix}\mathbf{0}\\ \hat{\mathbf{x}}\end{bmatrix} \tag{2.37}$$

and the reduced problem becomes homogeneous. This guarantees the exact fulfillment of the Dirichlet boundary conditions. The reduced problem now reads,

$$\begin{bmatrix}\mathbf{B}\\ \mathbf{0}\end{bmatrix}^\intercal \mathbf{A}(\boldsymbol{\mu})\begin{bmatrix}\mathbf{B}\\ \mathbf{0}\end{bmatrix}\boldsymbol{\alpha}(\boldsymbol{\mu}) = \begin{bmatrix}\mathbf{B}\\ \mathbf{0}\end{bmatrix}^\intercal \mathbf{f}(\boldsymbol{\mu}) - \begin{bmatrix}\mathbf{B}\\ \mathbf{0}\end{bmatrix}^\intercal \mathbf{A}(\boldsymbol{\mu})\begin{bmatrix}\mathbf{0}\\ \hat{\mathbf{x}}\end{bmatrix}. \tag{2.38}$$

Thus the final solution necessarily respects the Dirichlet boundary conditions.

For the problem at hand, two separate low-order bases are built to approximate each unknown field [94, 82]. Transformation matrices $\mathbf{B_u}, \mathbf{B}_\mathrm{p}$ correspond to the displacement and pressure fields respectively.

In the following, the indicator of dependence of operators on the parameter vector ($\boldsymbol{\mu}$) has been omitted for clarity. The unknown vectors are approximated

as,

$$\mathbf{U} \approx \mathbf{B_u}\boldsymbol{\alpha}_\mathbf{u}, \tag{2.39}$$

$$\mathbf{P} \approx \mathbf{B}_\mathrm{p}\boldsymbol{\alpha}_\mathrm{p}, \tag{2.40}$$

and the reduced dimensional system to be solved, at time step i + 1 reads,

$$\begin{bmatrix} \mathbf{B_u}^\intercal \hat{\mathbf{K}} \mathbf{B_u} & \mathbf{B_u}^\intercal \hat{\mathbf{Q}} \mathbf{B_p} \\ \mathbf{B_p}^\intercal \hat{\mathbf{C}} \mathbf{B_u} & \mathbf{B_p}^\intercal \hat{\mathbf{H}} \mathbf{B_p} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_\mathbf{u} \\ \boldsymbol{\alpha}_\mathrm{p} \end{bmatrix}^{\mathrm{i}+1} = \begin{bmatrix} \mathbf{B_u}^\intercal \hat{\mathbf{f}}_\mathbf{u} \\ \mathbf{B_p}^\intercal \hat{\mathbf{f}}_\mathrm{p} \end{bmatrix}. \tag{2.41}$$

The unknowns in this new system are vectors $\boldsymbol{\alpha}_\mathbf{u}$ and $\boldsymbol{\alpha}_\mathrm{p}$ that contain the coefficients for linearly combining the elements in the reduced bases, to approximate the high-fidelity solution for any parametric value.

Constructing the Reduced Basis

We denote $\mathcal{M}$ the solution map, or solution manifold of the high-fidelity problem. $\mathcal{M}$ represents the set of solutions $\mathbf{x}(\boldsymbol{\mu})$ for all parameters $\boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^P$, defining a map:

$$\mathcal{M} = \{\mathbf{x}(\boldsymbol{\mu}) \in V \,:\, \boldsymbol{\mu} \in \mathcal{P} \subset \mathbb{R}^P\}. \tag{2.42}$$

The idea behind RB is to sample the solution manifold by taking snapshots, and use these snapshots to create the reduced space in which the reduced solution is sought. To achieve this, we start from a set of $\mathrm{N_s}$ high-fidelity solutions that are stored in a matrix $\mathbf{M} \in \mathbb{R}^{\mathrm{N_h} \times \mathrm{N_s}}$, as

$$\mathbf{M} = [\mathbf{x}^1, ..., \mathbf{x}^{\mathrm{N_s}}]. \tag{2.43}$$

That set of solutions, if well selected, contains the information necessary to describe the parametric dependency of the solution with an acceptable accuracy.

The Proper Orthogonal Decomposition (POD) will be used for the estimation of the reduced basis functions. The singular value decomposition of the matrix $\mathbf{M} \in \mathbb{R}^{\mathrm{N_h} \times \mathrm{N_s}}$ yields the product representation $\mathbf{M} = \mathbb{U}\boldsymbol{\Sigma}\mathbb{V}^\intercal$. The columns of the matrix $\mathbb{U}^{\mathrm{N_h} \times \mathrm{N_h}}$ are orthonormalized vectors that contain information on the parametric dependency of the snapshots. The matrix $\boldsymbol{\Sigma}$ is a diagonal matrix that contains the singular values $\sigma_1, \sigma_2 ..., \sigma_L$, $L = min\{\mathrm{N_h}, \mathrm{N_s}\}$. Extracting the $\mathrm{N_r}$ first columns of $\mathbb{U}$ will yield the transformation matrix $\mathbf{B}^{\mathrm{N_h} \times \mathrm{N_r}}$. The number $\mathrm{N_r}$ of vectors that are kept may be evaluated based on the singular value corresponding to each vector. The singular values provide a measure of the information of the matrix $\mathbf{M}$ that is captured by each vector.

Hence one might extract the first N_r columns of matrix $\mathbb{U}$ that correspond to the N_r largest singular values, and carry the most essential information.

Two separate snapshot matrices, each containing the degrees of freedom that correspond to each field are created. The problem is transient, so the snapshots consist of solution vectors for each time step, saved serially in the snapshot matrices as seen below,

$$\mathbf{M_u} = \begin{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_{N_t} \end{bmatrix}_1 \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_{N_t} \end{bmatrix}_2 \dots \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_{N_t} \end{bmatrix}_{N_s} \end{bmatrix} \tag{2.44}$$

$$\mathbf{M}_\mathrm{p} = \begin{bmatrix} \begin{bmatrix} \mathrm{p}_1 & \mathrm{p}_2 & \dots & \mathrm{p}_{N_t} \end{bmatrix}_1 \begin{bmatrix} \mathrm{p}_1 & \mathrm{p}_2 & \dots & \mathrm{p}_{N_t} \end{bmatrix}_2 \dots \begin{bmatrix} \mathrm{p}_1 & \mathrm{p}_2 & \dots & \mathrm{p}_{N_t} \end{bmatrix}_{N_s} \end{bmatrix}. \tag{2.45}$$

N_t is the number of time steps that constitute the snapshots, and N_s is the number of snapshots taken.

Singular value decomposition is applied to each matrix $\mathbf{M_u} = \mathbb{U}_\mathbf{u}\mathbf{\Sigma}_\mathbf{u}\mathbb{V}_\mathbf{u}^\intercal$, $\mathbf{M}_\mathrm{p} = \mathbb{U}_\mathrm{p}\mathbf{\Sigma}_\mathrm{p}\mathbb{V}_\mathrm{p}^\intercal$, thus obtaining matrices $\mathbb{U}_\mathbf{u}$ and $\mathbb{U}_\mathrm{p}$. These contain orthonormal basis vectors for the spaces spanned by the snapshots of each solution field. These vectors are sorted in decreasing order of their corresponding singular values. They are to be truncated based on singular values, as mentioned above, keeping the vectors that define spaces where most of the information of the snapshot sets is embedded. Thus, a low dimensional space, where one may seek an approximation of the snapshot matrices is found as a result.

The essential assumption that is made, is that, if an accurate enough approximation of the snapshot set can be found in this low order space, then, given that the snapshot set is representative of the solution manifold, an accurate approximation for any solution in the manifold may be found in that very space.

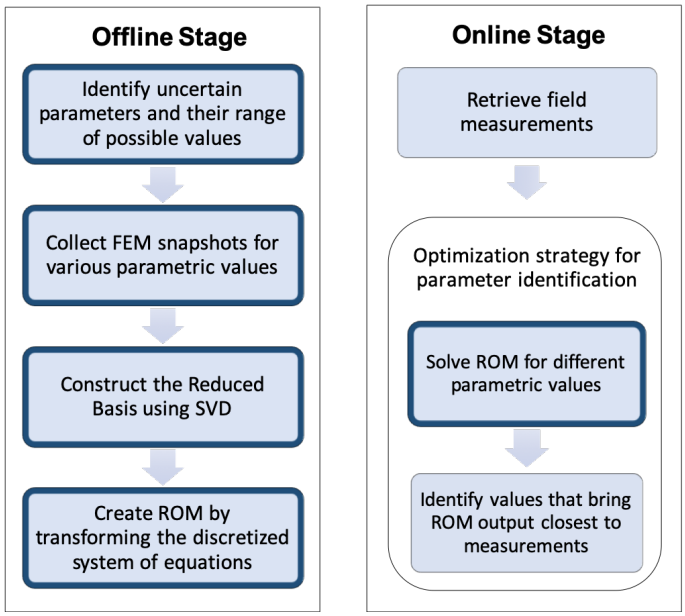


Figure 2.1: The online and offline stages of the Reduced Basis method in the context of data assimilation. Only the steps in the boxes with thick blue outlines are explained in this work.

The Reduced Basis method in a nutshell

In Figure 2.1 a clarifying summary of the Reduced Basis method stages in the context of data assimilation is displayed. The Figure illustrates the connection between the methodologies presented in this Chapter, and the motivating application, which is data assimilation for parameter identification, as mentioned in Section 4.1. It should be noted that the steps of the online stage that are related to the optimization method are not described in this work. The only part of the online stage that is examined here, is the solution of the ROM. Moreover, the one step of the online stage that is examined in this work, namely the solution of the ROM, may be part of other optimization applications that require repetitive solutions, and are not shown in this graph. The figure is only meant to encapsulate the RB method as part of the specific motivation of this work.

2.4 Reduced Order Model for predictive monitoring of Tailings Dam

Problem Setup

In this Section, a ROM of a tailings dam is created to simulate a problem that corresponds to the dam level raise. A complete study of the structure's integrity in such conditions, would imply modelling the tailings deposit, as well as the embankment, and considering a possible spatial variation in the mechanical and hydraulic properties. This would be necessary given that the failure surface often appears partly in the impoundment.

In the present work, only a first approach to address the problem is undertaken in order to demonstrate the high level of accuracy that can be achieved using Model Order Reduction to solve a complex, hydro-mechanically coupled, transient nonlinear problem, and discuss the contribution of such technology to real-time predictive monitoring. In that line, some simplifying modeling assumptions have been adopted. The impoundment is only treated as a load applied to the embankment and the water table upstream is considered stable even after the loading. The dam is considered to be founded on an impervious layer.

The modeled domain corresponds to the original embankment, while the deposited tailing material upstream, as well as the added material on top are modeled as mechanical loads. Initial conditions represent a steady state reached for a known water level imposed upstream, as shown in Fig. 2.2.

Loads p_w and p_e occur due to water and tailings deposit respectively. They are evaluated as,

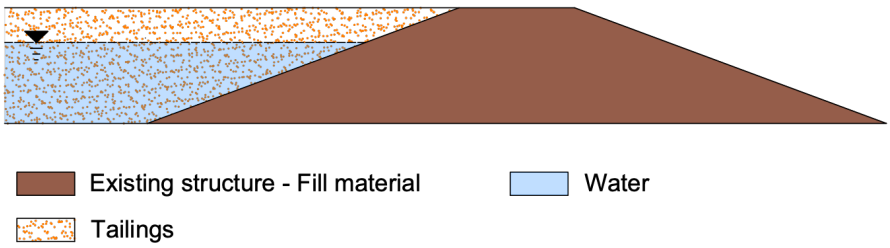
$$p_w = \gamma_w \times (WL - y) \tag{2.46}$$

$$p_{ex} = \begin{cases} K_a\left[\gamma_t(H - WL) + (\gamma_t - \gamma_w)(WL - y)\right] & \text{if } y \leq WL \\ K_a\gamma_t(H - y) & \text{if } WL < y < H \end{cases}, \tag{2.47}$$

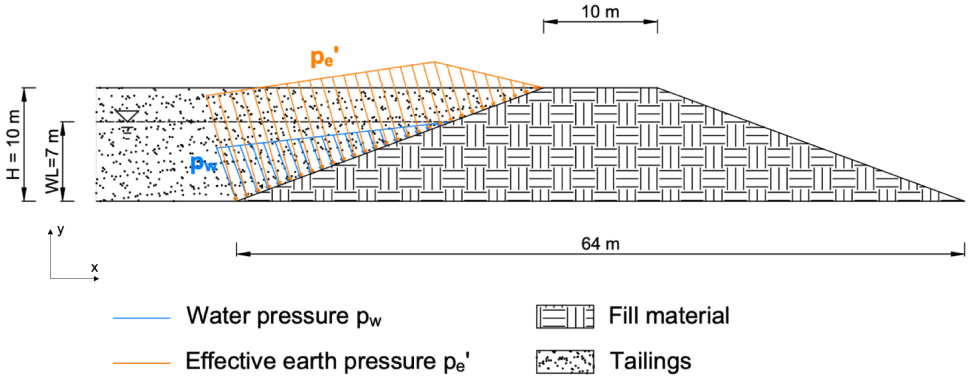
$$p_{ey} = \begin{cases} \left[\gamma_t(H - WL) + (\gamma_t - \gamma_w)(WL - y)\right] & \text{if } y \leq WL \\ \gamma_t(H - y) & \text{if } WL < y < H \end{cases}, \tag{2.48}$$

$$p_e = \sqrt{{p_{ex}}^2 + {p_{ey}}^2}, \tag{2.49}$$

where WL = 7m is the water level upstream of the dam and H = 10m the dam's height. Specific weights γ_w, γ_t correspond to water and tailings material respectively, $K_a = \frac{1-\sin(\phi)}{1+\sin(\phi)}$, is the active earth pressure coefficient, ϕ being the



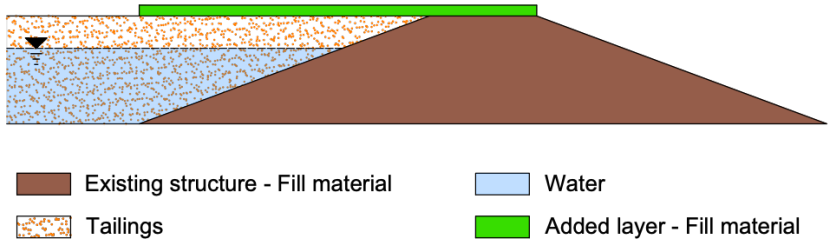
(a) Conceptual sketch



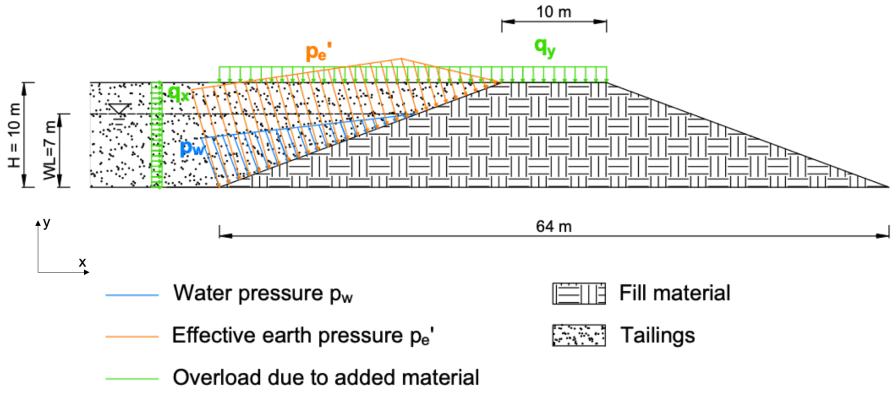
(b) Loads due to water and tailings as modeled

Figure 2.2: Initial conditions.

tailings angle of friction, and x, y denote the horizontal and vertical directions. The of use of K_a for the lateral earth pressure follows the simplifying assumption that the retaining embankment has been allowed to yield sufficiently for the full active earth pressure to be activated. With the modulus of subreaction method [109], modeling the transition from the at-rest to the full active earth pressure implies the introduction of a nonlinearity, since the earth horizontal reaction coefficient depends on the displacement of the domain. K_a was here considered to avoid that complexity, and given that the horizontal displacement reached on the top of the dam when the surcharge load has been applied, is large enough to be consistent with the mobilization of the full active earth pressure based on empirical rules often used in geotechnics



(a) Conceptual sketch



(b) Loads due to water and tailings as modeled

Figure 2.3: Final conditions.

[9]. The fill material is considered to be cohesion-less, as is often the case for tailings dams [92, 120].

Next, a load that corresponds to a level raise by 1m is gradually applied to the top of the structure. The load increases over a duration of 10 days and is then kept constant for the rest of the simulation. This setup is meant to simulate realistic conditions, that is, a small raise, no more than a meter per year, that is preceded and followed by a period of approximately equal time, during which no tailings are deposited, and the fill material is given some time to consolidate. The water table upstream, remains stable throughout the simulation. The final conditions are displayed in Figure 2.3.

The loads attributed to the 1 meter thick layer of added material are

estimated as,

$$q_x = K_a \times q_y, \tag{2.50}$$
$$q_y = \gamma_f \times 1m, \tag{2.51}$$

where γ_f is the added fill material specific weight.

Therefore the conditions that bound the problem are of type Dirichlet, Newman and Robin. In Figure 2.4, 5 boundary parts are defined as $\Gamma_{UD}, \Gamma_{UW}, \Gamma_B, \Gamma_D, \Gamma_T$ that denote Upstream Dry (above water table), Upstream Wet, Bottom, Downstream and Top boundary parts respectively.

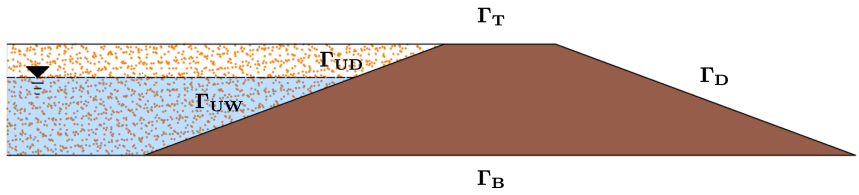


Figure 2.4: Boundary parts: Upstream Dry, Upstream Wet, Bottom, Downstream, Top

The boundary conditions introduced to the Finite Element model are written as

$$\mathbf{u}_x = \mathbf{u}_y = 0 \quad \text{on } \boldsymbol{\Gamma_B}, \tag{2.52}$$
$$\boldsymbol{\sigma} \cdot \mathbf{n} = \begin{bmatrix} q_x + p_{ex} & q_y + p_{ey} \end{bmatrix}^\intercal \quad \text{on } \boldsymbol{\Gamma_{UD}}, \tag{2.53}$$
$$\boldsymbol{\sigma} \cdot \mathbf{n} = \begin{bmatrix} q_x + p_{wx} + p_{ex} & q_y + p_{wy} + p_{ey} \end{bmatrix}^\intercal \quad \text{on } \boldsymbol{\Gamma_{UW}}, \tag{2.54}$$
$$\boldsymbol{\sigma} \cdot \mathbf{n} = \begin{bmatrix} 0 & q_y \end{bmatrix}^\intercal \quad \text{on } \boldsymbol{\Gamma_T}, \tag{2.55}$$
$$p = \gamma_w \times (WL - y) \quad \text{on } \boldsymbol{\Gamma_{UW}}, \tag{2.56}$$
$$\mathbf{q} \cdot \mathbf{n} = 0 \quad \text{on } \boldsymbol{\Gamma_{UD}} \cup \boldsymbol{\Gamma_B} \cup \boldsymbol{\Gamma_T}, \tag{2.57}$$
$$\mathbf{q} \cdot \mathbf{n} = \langle \beta p \rangle \quad \text{on } \boldsymbol{\Gamma_D}. \tag{2.58}$$

The relations among the boundary parts definitions given here, to the ones given in equations (2.8) - (2.9) and (2.11) - (2.13) may be written $\Gamma^u_D = \boldsymbol{\Gamma_B}$, $\Gamma^u_N = \boldsymbol{\Gamma_{UD}} \cup \boldsymbol{\Gamma_{UW}} \cup \boldsymbol{\Gamma_T}$, $\Gamma^p_D = \boldsymbol{\Gamma_{UW}}$, $\Gamma^p_N = \boldsymbol{\Gamma_{UD}} \cup \boldsymbol{\Gamma_B} \cup \boldsymbol{\Gamma_T}$, $\Gamma^p_R = \boldsymbol{\Gamma_D}$.

Parameter	Symbol	Units	Value
Gravitational acceleration	g	m/s	10
Water bulk modulus	K_w	MPa	2.2×10^3
Specific weight of water	γ_w	kN/m^3	10
Embankment fill soil material			
Particle density	ρ_s	kg/m^3	2.7×10^3
Young's Modulus	E	MPa	40
Poisson's ratio	ν	-	0.3
Porosity	η	-	0.38
Embankment fill material - Van Genuchten Model [115]			
Saturated VWC	Θ_s	-	0.38
Residual VWC	Θ_r	-	0.038
Parameter ($\approx$ inverse of air entry suction head)	α	m^{-1}	0.1
Fitting Parameter	m	-	0.184
Saturated hydraulic conductivity	k_s	m/s	$[10^{-9}, 10^{-7}]$
Tailings and added fill material			
Added fill material specific weight	γ_f	kN/m^3	21
Tailings specific weight	γ_t	kN/m^3	21
Tailings friction angle	ϕ	$^\circ$	35

Table 2.1: Values of physical parameters used in the model

Implementation: Full and reduced order solvers

A FEM code for the problem stated above was developed in the FEniCS open-source platform. Following, the Reduced Basis method was used to create a low-order solver for the parametrized problem.

The values of the parameters used in the model are given in Tables 2.1 and 2.2. The values were chosen such that they fall into ranges that are usually observed in tailings dams [14, 92].

Discretization	
Mesh	1382 elements, 774 nodes, unstructured
FE Displacement	P2
FE Pressure	P1
Time integration: θ-scheme	
θ	0.75
Time step Δt	0.1 days

Table 2.2: Numerical parameters used in the model in the 2D scheme

The parameter chosen to be examined is the material saturated hydraulic conductivity k_s. As mentioned above, hydraulic properties of the materials that exist in tailings dams feature high uncertainty and may vary in time. The parametric domain is taken such that the extreme values are realistic in the framework of tailings dams. The saturated hydraulic conductivity of the fill material takes values in $[10^{-9}, 10^{-7}]$(m/s).

Solving the loading problem with FE

In Figures 2.5a-2.5c the pore water pressure fields acquired by the FE model, solving for $k_s = 10^{-9}$m/s, for 3 different time instances are shown. Figure 2.5a corresponds to the initial state before the loading is applied.As mentioned above, the loading period lasts 10 days, during which the overload is gradually increased, simulating a layer of fill material that is gradually deposited on top of the structure. Figure 2.5b represents the pore pressure state after 10 days of loading. During the loading time, changes in the displacement field reflect settlement in the dam due to overload.

The hydro-mechanical coupling induces an increase of pore pressure, as the material is compressed, and the pore space reduced. At first, there is no deformation in the saturated zone, and the load is carried by the pore water. Hence, overpressure is built due to the weight of the added layer of material, making the water table rise. The process of consolidation starts immediately after loading. As water is expelled from the pores, there is a reduction in soil volume and a transfer of pressure from water to soil particles. After 10 days, the load ceases to increase and the process of consolidation continues as the water table falls. The simulation stops when a steady state is reached. The

pressure field is considered to be in steady conditions when the L_2 norm of the difference between the fields corresponding to two subsequent time steps, is smaller than 10^{-2}. For $k_s = 10^{-9}$m/s this occurs at t = 60, 6 days and the final solution fields can be visualized in Figure 2.5c.

Setting up the ROM: Offline stage

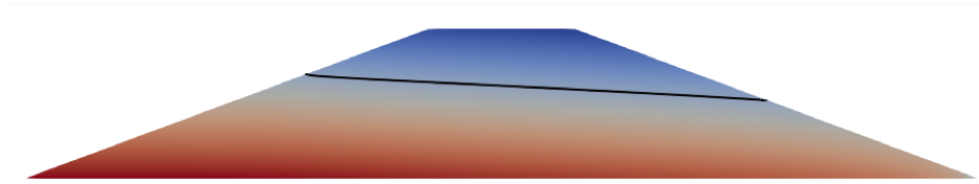
As explained in section 2.3, the Reduced Basis is constructed by means of sampling the high-fidelity solution manifold, that is, in this case, the set of solutions obtained by the full-order Finite Element solver.

Specifically, the full order problem was solved for $k_s = [1, 3, 5, 7, 10, 30, 50, 70, 100] \times 10^{-9}$, and the solutions were stored in two separate snapshot matrices as in equations (2.44) and (2.45). The snapshots were selected such that the values are close to evenly spaced in the logarithmic scale of the parametric domain. Each of the snapshots have a duration of 10 days, during which the load is applied, plus the time that is needed for steady state conditions to be reached. The time required for steady state conditions to be reached after loading depends on the hydraulic conductivity of the material. The built-up pressure requires longer time to dissipate in a less permeable material. Thus the snapshots have different durations.

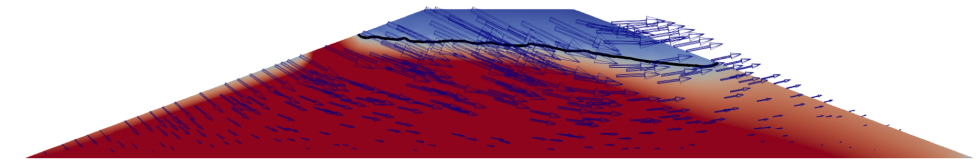
Singular value decomposition was applied to the two matrices resulting in the left singular vectors, that were truncated to yield the Reduced Bases. The truncation criterion is based on the singular values. In Figure 2.6 the singular values that correspond to each of the left singular vectors for the two fields are plotted. The y-axis is in logarithmic scale and it is normalized with respect to the first -and largest- singular value. That is, the values of the y-axis are calculated as,

$$\log_{10}(\frac{\sigma_i}{\sigma_1}), \tag{2.59}$$

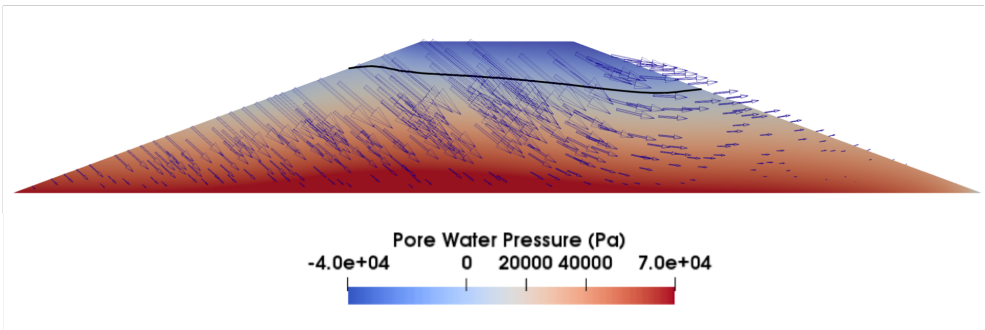
where σ_i is a plotted singular value and σ_1 is the largest singular value, and corresponds to the first vector of each left matrix. The values drop rapidly in both cases. The first vectors contribute significantly to the description of the solution set, and must, therefore, be included into the Reduced Basis, while as the singular value decreases, the corresponding vectors convey less information about the data, that is, the snapshot matrix.



(a) Initial conditions. The Steady state with fixed upstream water table at 7 m. Displacement is considered null.

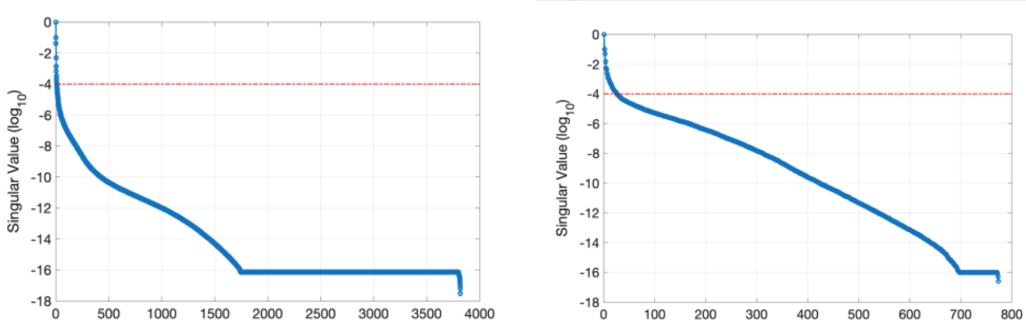


(b) At t = 10 days, after 10 days of loading. The maximal overpressure is reached. At this point, the loading stops and pressure starts dissipating.



(c) At t = 60, 6 days, end of the simulation. The black line indicates the position of the phreatic line.

Figure 2.5: Pore water pressure and displacement field at time instances. The arrows point in the direction of the displacement and the sizes are proportional to the displacement magnitude.



(a) Displacement field

(b) Pressure field

Figure 2.6: Truncation of the matrix of left singular vectors.

It is shown in Figure 2.6, that the singular values that correspond to the displacement field decrease faster than the ones that correspond to the pressure field. This is likely related to the selected parameter k_s, a hydraulic parameter that mostly affects pressure field in a more pronounced manner than the displacement field. The red dashed line denotes the truncation threshold. In this case, vectors that correspond to singular values that are smaller than the first one by 4 orders of magnitude or more, are discarded. Obtaining a ROM with an accuracy higher than that, would be beyond the scope of the ROM, since the accuracy in the problem is limited by that of the sensor measurements.

This truncation threshold yields a basis that is comprised of just 9 vectors for displacement and 25 for pressure, thus 34 is the size of the reduced system of equations. To put this number of reduced unknowns in perspective, the high-fidelity problem dimension, related to mesh resolution and polynomial degree is 6632 degrees of freedom.

The pressure basis is larger than the displacement one, even though resulting from the same tolerance as illustrated in Figure 2.6. This is related to the complexity of the solution manifold of the two fields and may be attributed to the fact that the chosen parameter is a hydraulic property that affects the pressure field more, or with the fact that the nonlinear terms of the model mostly appear in the flow balance equation.

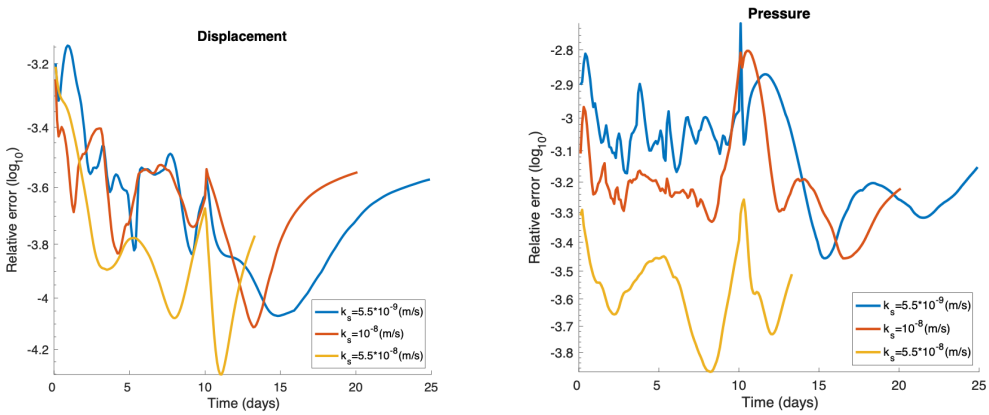
Online stage: Results and comparison

Having populated the transformation matrices $\mathbf{B_u}$ and $\mathbf{B_p}$, the problem may now be solved, for any parameter within the examined range, by assembling the system of equations and projecting it to the reduced space in which the approximation will be sought, as shown in equation (2.41). In Figure 2.7 the relative error of the low order approximation with respect to the high-fidelity solution, is plotted over time. The error is estimated as,

$$e = \left\| \frac{X_{RB} - X_{FEM}}{X_{FEM}} \right\|_2, \tag{2.60}$$

where X_{RB} and X_{FEM} represent the approximation and the high-fidelity solution respectively.

Of the three values examined in Figure 2.7, one is a snapshot value, namely $k_s = 10^{-8}$ m/s, and the other two are values that were not sampled.



(a) Displacement field (b) Pressure field

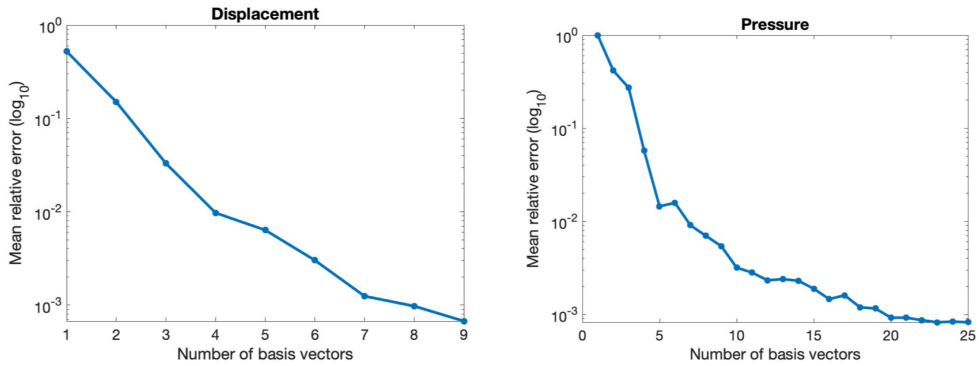
Figure 2.7: Relative error of low-order approximation of the pressure field with respect to full order solution over time, estimated over the entire domain, for 3 different parametric values.

The errors for both fields and for all parametric values remain quite low, despite the small number of base vectors used. In fact, the error is much lower than the typical accuracy of measurement of the instrumentation that corresponds to the quantities evaluated. Note that the error does not seem

to be significantly smaller for the case of the snapshot value $k_s = 10^{-8}$ m/s. This indicates that the Reduced Basis sufficiently describes the solution states that correspond to the entire spectrum of parametric values. Moreover, it is worth mentioning that the reduced order model yields a small error even for the parametric value that corresponds to a snapshot, which is an expected behavior, considering that the snapshot matrix was truncated after the singular value decomposition.

The ROM runs 3 times faster, by average, than the full-order model. That is a significant boost of computational efficiency, especially considering the relatively low mesh resolution.

In Figure 2.8, the mean approximation error obtained for different RB sizes is shown. For each solution field, 3 parametric values were considered, namely $k_s = [5.5 \times 10^{-9},\ 10^{-8},\ 5.5 \times 10^{-8}]$ m/s. The mean approximation error was computed over all the time steps of all 3 parametric values, as in the expression of equation (2.60). For each field, the basis size is increased, while the size of the basis for the other field is kept constant, and equal to the largest value under consideration.



(a) The size of the pressure basis is kept at 25 modes, and the size of the displacement basis is increased from 1 to 9.

(b) The size of the displacement basis is kept at 9 modes, and the size of the pressure basis is increased from 1 to 25.

Figure 2.8: Mean relative approximation error over all the time steps of the solutions obtained for 3 different values in the parametric domain, using increasing number of basis vectors.

The Figure demonstrates the fast improvement of the approximation accuracy, with the first few added vectors, reaching errors in the order of magnitude of 10^{-3} rather fast. After the first few vectors have been included,

the improvement in accuracy becomes slower, as vectors that correspond to smaller singular values are added.

To test the effect of mesh resolution to the computational time reduction, another scheme was run, for the same problem, using a denser mesh. In this scheme the high-fidelity system of equations has 2975 degrees of freedom for pressure and 23164 for displacement. The same number of snapshots were taken and corresponded to the same parametric values. The reduced bases that yield the same level of accuracy as the first scheme have 23 base vectors for the pressure field and 9 for the displacement field. The number of vectors needed is very close to the previous scheme, despite the large difference in the size of the high-fidelity problem. In other words, the size of the reduced scheme is decoupled from the size of the high-fidelity scheme [50]. This is an essential advantage of the Reduced Basis method. It is worth noting that in the case of nonlinear problems, like the present, the reduced scheme cannot be fully decoupled from the FEM scheme. Due to the state-dependence of the FEM operators in the linearized system, they have to be reassembled in every iteration of the linearization scheme. The full order stiffness matrices and force vectors must assembled and then transformed into the reduced order ones. The assembly of these operators is related to the high-fidelity dimension. Therefore the efficiency gains are bounded. This issue is often addressed by the empirical interpolation methods [50, 93], the implementation of which, will be discussed in the following Chapter.

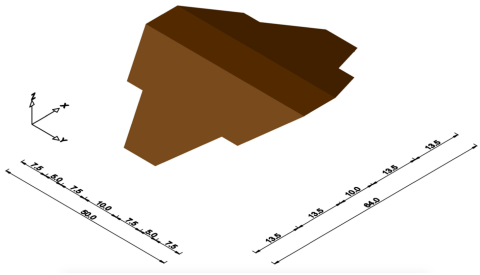
Extension to 3D problems

Literature suggests that 2 dimensional models cannot reflect the complex and variating seepage field [68]. Thus, 3 dimensional models have been proposed for the stability study of tailings dams [67, 127]. In this section, a seepage problem, similar to the 2 dimensional one above, will be solved in a 3D setting and a ROM will be developed and evaluated with respect to its accuracy and computational efficiency against the high-fidelity model.

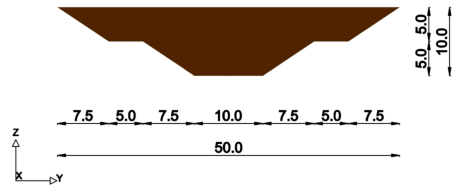
The new geometry approximates an embankment constructed in a narrow, steep sided valley. This is a generic, invented geometry, meant to simulate common conditions in the construction of embankment dams.

In Figure 2.9 the Y axis indicates the third dimension, while axes X and Z correspond to the dimensions that were considered in the plane strain approximation of the previously solved problem. The domain has a variating cross-section along the Y axis. The two-dimensional domain of the previous setting corresponds to the middle section of the dam, ie the part that has its

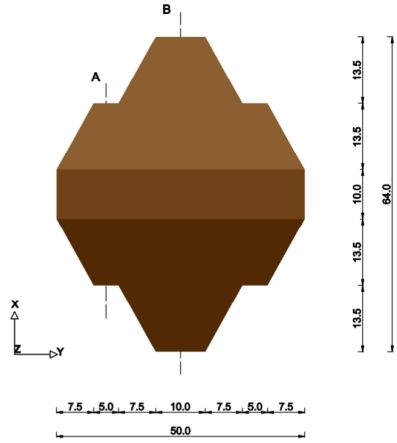
foundation on the valley (Figure 2.9d Bottom). The parts that are founded on the lateral slopes, have smaller cross sections (Figure 2.9d Top). The side slopes have been assumed excavated in two levels, creating two 5 m-high slopes of inclination 1,5:1. The domain is a 50 m-long prism, has 2 axes of symmetry, and an identical cross-section as the 2D domain shown in Figure 2.2.



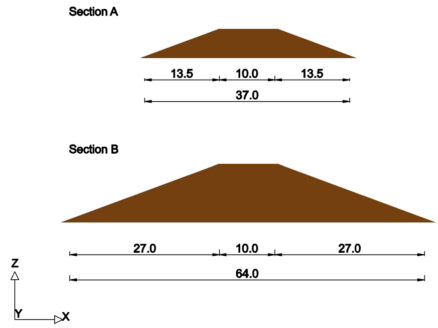
(a) 3D Isometric view



(b) Front view



(c) Top view



(d) Cross sections Top: Section plane A Bottom: Section plane B

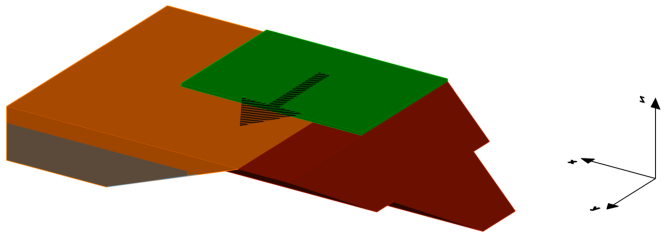
Figure 2.9: Views and cross sections of the 3 dimensional domain. The geometry assumes an embankment that has been constructed in a narrow valley, and has its foundations on two side slopes at the two extremes in the direction of Y axis. The side slopes are not displayed.

The 3D tetrahedral mesh was created in Gmsh open-source mesh generating software [38]. The mesh is made out of 41477 tetrahedral and triangular cells

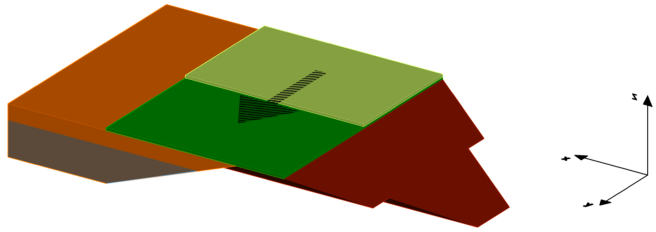
and 9896 nodes. Using the Taylor-Hood (P2-P1) element for the description of the displacement and pressure fields as before, the resulting system is comprised of 202179 and 9896 degrees of freedom for the two fields respectively.

As for the problem setup, it remains similar to the 2D case with some alterations. The initial condition is evaluated by solving the steady state problem for an upstream water level at 7 m. The added material is now placed in two layers of 50 cm each. The first layer is deposited in the first 5 days, gradually, across the Y axis. Then the second 50 cm layer is added on top, during days 5-10, again, gradually across the Y axis. After the period of 10 days, the full load that represents the 1m-thick layer remains constant. The bottom boundary is mechanically constrained. The parameters used have the values listed in Table 3.1. During the level raise phase in a tailings dam life, thin layers of fill material are deposited along the length of the structure. The load increases gradually along the 3rd axis of the dam, i.e. the dimension that is not considered in a 2D plane strain approximation. The problem simulates conditions that may occur in the context of an actual tailings dam and cannot be sufficiently approximated in a 2D plane strain setup.

In Figures 2.10a and 2.10b the loading is illustrated for clarity. Two layers of soil are placed on top of the dam, in order to raise the level of the dam by 1 meter in the upstream manner, as shown in Figure 1.3.



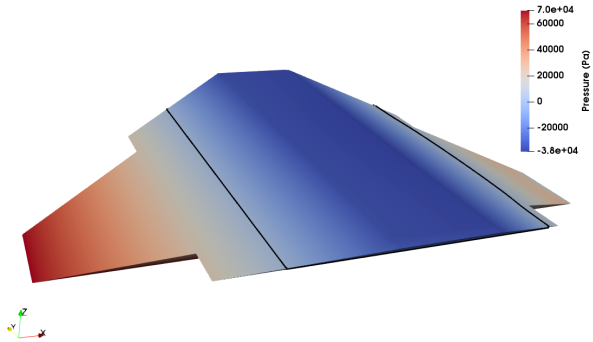
(a) After 2,5 days.



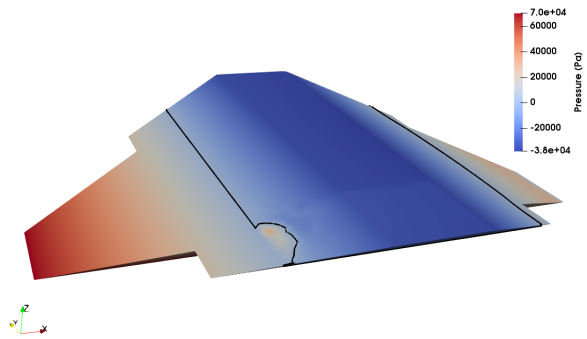
(b) After 7,5 days

Figure 2.10: Deposition of the two 50 cm - thick layers of fill material on the top and upstream side of the structure. The first layer is shown in dark green color and the second in light green color. After 2,5 days of loading half of the first layer has been deposited. After 7,5 days the first layer and half of the second layer have been deposited. Upstream is in the direction of the x axis. Unsaturated tailings are depicted in orange, saturated tailings in gray. The lateral foundation slopes are not depicted.

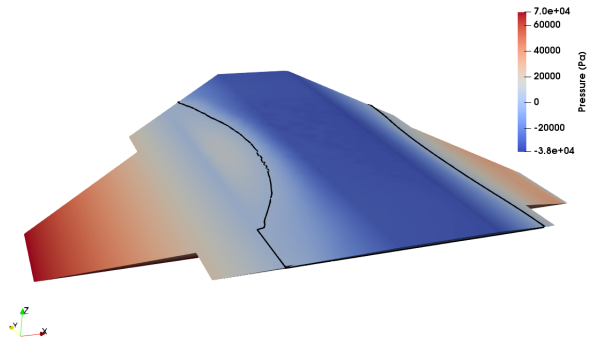
In Figure 2.11, the effect of the gradual load application on the pore pressure field is illustrated. As the load is applied from one extreme to the other in the y direction, the water table in the area rises to be then lowered after pressure is dissipated through consolidation. By the time the loading procedure is finished after 10 days, in some parts of the dam the water table has almost reached its final position after consolidation.



(a) t=0



(b) t = 1 day



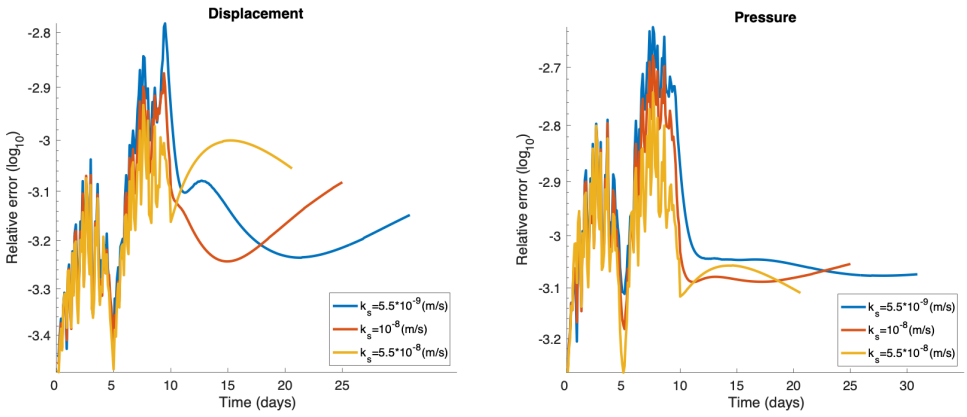
(c) t = 10 days

Figure 2.11: Pore water pressure distribution on different time instances (a) Initial conditions. Steady state condition with an upstream level at 7m. (b) After 1 day of loading. In some part of the domain overpressure has built up. (c) After 10 days of loading the top of the dam and impoundment. The reservoir is considered to be located on the left side of the figures, i.e. upstream is opposite to the x axis direction.

The Reduced order scheme was created following the same steps as in the 2D case. Snapshots of the high-fidelity solution manifold were taken for parametric values $k_s = [1, 3, 5, 7, 10, 30, 50, 70, 100] \times 10^{-9}$. As above, each snapshot has a duration of 10 days of loading, plus the amount of time needed for steady state to be reached in each case. The two solution fields - displacement and pressure- were stored in snapshot matrices $\mathbf{M_u}$ and $\mathbf{M}_p$, arranged as before, i.e. snapshots comprised of all time steps stored in a serial manner in the columns of the matrices. Singular value decomposition was applied to each matrix separately.

Much like in the 2D case, the matrix of left singular vectors was truncated with a criterion related to the singular values, admitting only vectors that correspond to singular values up to 4 orders of magnitude smaller than the largest one. The truncation yielded reduced bases comprised of 80 base vectors for displacement and 174 for pressure. Therefore the system of equations to be solved now is of dimensionality 214, instead of the full order problem with 212075 degrees of freedom.

The same parametric values as before were examined in order to estimate the accuracy and computational efficiency gain of the RB approximation. In Figure 2.12 the L_2 norms of the relative error -estimated as the difference between full order and reduced order solutions- have been plotted for the two fields.



(a) Displacement field (b) Pressure field

Figure 2.12: Relative error of low-order approximation of the pressure field with respect to full order solution over time, estimated over the entire 3D domain, for 3 different parametric values.

In comparison to the 2D scheme, it seems that the ROM yields slightly lower accuracy, which remains however in the same order of magnitude. The accuracy can easily be improved by choosing a smaller truncation tolerance for the reduced basis. In Figure 2.12, it seems that for the first time steps, up to day 10, the error is noisy/presents some oscillations. This part of the graph corresponds to the timesteps where the load is increased. In the following time steps the load remains fixed, and the error between the low-order and the full-order solutions seems to fluctuate in a smoother way, or stabilize around a value.

For various parametric values that were tested, the ROM was found to run 8 - 15 times faster than the full-order model. This is a significant increase of gain in computational efficiency compared to the 2D case, which indicates that the time gain scales along with the size of the problem at hand.

2.5 Conclusions and discussion

In this Chapter, a low order model for the hydro-mechanically coupled problem has been developed, using the Reduced Basis technique. The POD-based method has been successfully applied. A first attempt to treat the parametrized system of PDEs and confirming a high level of accuracy has been demonstrated.

The computational time needed is reduced in this simplified problem, while the accuracy of the solution is not compromised, in comparison to the full-order Finite Element solution. The method is promising, considering that the time reduction increases with increasing problem size.

It is worth noting that in the framework of tailings dams the usual practice is to model the retaining structure as well as the stored material, due to the fact that failure may originate in the impoundment. Sufficiently describing the pressure and displacement field -or other fields of interest- in such a large inhomogeneous domain can result in large systems, in which case the advantages of a reduced order model would be most prominent.

Chapter 3

Hyper-reduction with Discrete Empirical Interpolation

This chapter is based on:
C. Nasika, P. Díez, P. Gerard, T.J. Massart, S. Zlotnik, Discrete Empirical Interpolation for hyper-reduction of hydro-mechanical problems in groundwater flow through soil
In preparation

3.1 Introduction

In a transient nonlinear problem, a low-order system of equations must be solved multiple times within each time step, as the linearization scheme converges to a solution. This implies the repetitive assembly of FE operators and force/residual vectors that result from nonlinear terms of the PDE. The cost of every FE iteration is related to two expensive processes: the assembly of said operators and the solving procedure. POD-based ROMs decrease the cost associated to the latter, but do not affect the cost of the former. The cost of assembling the operators scales with the dimension of the full-order problem (N_h), and by consequence the efficiency gains that can be achieved with ROMs are limited. This drawback is often referred to as the lifting bottleneck, reflecting the fact that the evaluation of the nonlinear terms implied a lift back to the original dimension of the full-order problem and then a projection to the reduced order [57].

To remedy this issue several techniques have been developed, attempting to decouple the ROM size from the full-order size. Rather common methods among them include Missing Point Estimation [7], that extracts certain system equations that correspond to specially selected grid points where the nonlinearity is computed, and the Trajectory Piecewise-Linear method (TPWL) [98, 99] which represents a nonlinear system as a piecewise-linear one, and approximating it as a weighted sum of a small set linearized models at selected points along a state trajectory.

In this work, the Discrete Empirical Interpolation Method (DEIM) is used. DEIM was originally developed in [21] to treat nonlinear problems in a POD context. DEIM is somewhat related to Missing Point Estimation in the sense that both methods avoid the costly evaluation of inner products that correspond to nonlinear terms, by selecting a few grid points, and merely evaluating the nonlinearities on these points. However, the DEIM method involves the construction of a POD basis for each nonlinear term, in addition to the POD basis related to the state variables. That is, in DEIM a system approximation is defined, in addition to the solution space approximation. The approximation is constructed by projection onto the POD basis and interpolation based on the selected grid points.

The central idea of DEIM is similar to that of the Reduced Basis (RB) method. Assembling and solving the system at very reduced computational cost is feasible if the RO operators are independent of the FE order N_h. This can be done by obtaining an affine approximation of the nonlinear or non-affinely parametrized FE operators in the offline stage, therefore constructing a linear

system approximation. Similarly to the construction of the solution space approximation, the system approximation is based on the recovery of N_d arrays that are stored during the offline stage, and can be linearly combined in the online stage in order to quickly obtain an approximation of the matrix or vector operator. This set of arrays is a POD basis for the space that is spanned by a set of snapshots of the approximated function, and it will referred to as a DEIM basis in the following.

The size of the DEIM basis depends on the complexity of the problem, that is, it depends on the type of underlying nonlinearity, and the range of different states, or behaviors, that the system undergoes. The DEIM basis must be appropriate for accurately approximating any of the possible states of the system. If that range is too wide, the resulting size of the approximation is large, and the efficiency gains obtained by the method remain limited.

One of the developed methods that aims to tackle this issue is the Localized Discrete Empirical Interpolation (LDEIM), a variation of DEIM that was introduced in [88]. This approach aims to the construction of even more efficient ROMs, by minimizing the size of the DEIM basis that is used for system approximation. LDEIM uses machine learning-based methods in order to identify not one, but several local subspaces where the approximation of a nonlinear function can be sought for. Each of the resulting subspaces correspond to a particular region of system behavior. Two approaches to LDEIM were proposed in [88], one which associates the different local DEIM bases to parametric sub-domains, and one that associates subspaces to sets of characteristic system states. The second one is used in this work. The resulting subspaces may be smaller in size compared to the global approximation space that is obtained with conventional DEIM. Therefore, the size of the resulting system approximation problem that is solved online in every iteration is smaller and the assembly of the approximated operator can be performed faster.

In geomechanics, the use of POD-based reduction has been mostly motivated by subsurface flow applications. The authors of [119] appear to be the first to implement POD-based solution space reduction, without the integration of a hyper-reduction technique, to geomechanical problems. The most commonly used hyper-reduction method is TPWL in the context of pure two-phase flow (oil/water or oil/gas) analysis [18, 48, 46, 47] for reservoir simulation. In [56] the coupled hydro-mechanical problem with two-phase oil/water flow is treated in a POD-TPWL approach, again in the context of subsurface flows. In all cases, the motivating applications are related to production optimization, uncertainty quantification, history matching and

many-query problems in general.

The authors of [29] developed a method that is similar to a POD-DEIM approach in the context of Finite Volume method for the two-phase (water,oil) pure fluid flow through a porous media. Similarly the authors of [41] used a POD-DEIM approach for Finite Volumes in the uncoupled flow problem, and additionally performed a comparison of POD-based and TPWL methods. The LDEIM method was examined in a similar context of two-phase fluid flow analysis in [39]. A POD-DEIM approach, has not been applied so far to the coupled flow/geomechanics problem to the knowledge of the author of the present.

The novel contribution presented in this Chapter lies in the implementation of a POD-DEIM and POD-LDEIM ROMs to the fully coupled hydro-mechanical problem is examined in the context of FE analysis for embankment dams. The adopted physical model assumes linear elasticity for the porous media and describes the behavior of partially saturated soils. Nonlinearities are introduced to the problem due to the dependence of certain hydraulic and mechanical properties on pressure for the case of unsaturated soil. The resulting system of PDEs is transient and nonlinear.

This Chapter is structured as follows. In Section 3.2 the applied Model Order Reduction methodologies, DEIM and LDEIM, are detailed and some precisions are reported regarding their implementation to the treated problem. In Section 3.3, ROMs are created to solve a hydro-mechanical problem featuring the mechanical loading of a 3D embankment dam. The problem is described, and results pertaining to the computational efficiency and the accuracy of the ROMs solution are reported, and compared to the FE model. Finally, in Section 3.4 the results of this Chapter are summarized, some important conclusions regarding the examined methods are discussed, and comments are made regarding the limitations of the method.

3.2 Methodology

Hyper-reduction of the right-hand side vector with the Discrete Empirical Interpolation Method

The key objective of DEIM is the efficient evaluation of the reduced vector $\mathbf{B}^{\intercal}\mathbf{f}(\mathbf{x}, \boldsymbol{\mu})$ found on the right-hand side of the reduced problem in equation (2.34). Without the use of a hyper-reduction technique, the evaluation of right-hand side vector $\mathbf{f}(\mathbf{x}, \boldsymbol{\mu})$ necessarily implies access to the full order problem and projection to the reduced space after assembly.

The DEIM method follows a similar strategy to that of the Reduced Basis method. The idea is to identify a low-order space where a nonlinear function may be approximated with acceptable accuracy. In the present case the nonlinear function is the right-hand side vector. N_d denotes the dimension of that low-order space. This space is different from the solution approximation space with basis vectors $\mathbf{B}$, but it is determined in the same manner, that is, in an offline stage where POD is applied to snapshots of the nonlinear function.

Similarly to the RB method previously described, a snapshot matrix containing actualizations of the full-order vector $\mathbf{f}$ is populated as,

$$\mathbf{M}_d = \left[\mathbf{f}^1 ... \mathbf{f}^{N_s}\right], \tag{3.1}$$

where N_s is the number of snapshots taken, as in, the number of full-order vectors evaluated and saved in the snapshot matrix. In the context of a time-dependent problem, $\mathbf{f}$ refers to matrices, in the columns of which correspond to snapshots for each time step.

In the context of an RB/DEIM approach where the objective is to construct both a solution approximation and an approximation of the system, the snapshot matrix $\mathbf{M}_d$ can be populated in the offline stage, either simultaneously or sequentially to the collection of solution snapshots. In this work, both offline processes are performed simultaneously with the solution of the full FE problem for different parametric values.

SVD is applied to the snapshot matrix to obtain orthonormal singular vectors for the space that is spanned by the snapshots. The matrix decomposition reads,

$$\mathbf{M}_d = \hat{\mathbf{V}}\mathbf{S}\mathbf{W}^\intercal, \tag{3.2}$$

where $\hat{\mathbf{V}} \in \mathbb{R}^{N_h \times N_s}$ a matrix whose columns are the singular basis vectors and $\mathbf{S}$ contains the singular values that correspond to each vector. Similarly to the procedure described in Section 2.3, the first N_d columns of $\hat{\mathbf{V}}$ are selected based on the relative magnitude of their respective singular value as,

$$\mathbf{V} = \text{first } N_d \text{ columns of } \hat{\mathbf{V}}, \tag{3.3}$$

where the reduced dimension N_d is determined given that the subsequent singular values are negligible,

$$\log_{10}\left(\frac{\sigma_1}{\sigma_{N_d}}\right) \leq \text{tol} \tag{3.4}$$

in which σ_1 is the first and largest singular values found in $\mathbf{S}$ and σ_{N_d} the singular value corresponding to the last selected vector. tol is a tolerance

selected based on the level of accuracy required in the system approximation. Based on the experience obtained in this work, the DEIM tolerance has to be equal to, or smaller than the one selected in the solution approximation.

The right-hand side vector approximation in the N_d sized linear space is defined,

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\mu}) \approx \mathbf{V}\mathbf{c}(\mathbf{x}, \boldsymbol{\mu}) = \mathbf{v}_1 c_1 + \mathbf{v}_2 c_2 + \ldots + \mathbf{v}_{N_d} c_{N_d}, \tag{3.5}$$

where $\mathbf{c}(\mathbf{x}, \boldsymbol{\mu}) \in \mathbb{R}^{N_d}$ is a vector of coefficients used for the linear combination of basis vectors contained in $\mathbf{V}$. In the following, the dependence of $\mathbf{f}$ and $\mathbf{c}$ on $(\mathbf{x}, \boldsymbol{\mu})$ is suppressed for clarity.

In the online stage, at each iteration of the solution procedure applied to the transient nonlinear problem, the evaluation of the combination coefficients $\mathbf{c}(\mathbf{x}, \boldsymbol{\mu}) \in \mathbb{R}^{N_d}$ is required in order to fully determine the approximation (3.5). To achieve that, an N_d sized problem must be solved in every iteration. DEIM proposes to identify the coefficients by imposing exact equality of the approximation and the full-order function on N_d points. In other words, DEIM allows the construction of an approximation of $\mathbf{f}$, given the precise value of N_d entries of the full-order operator $\mathbf{f}$. That is, N_d components of $\mathbf{f}$ that allow computing the N_d values $c_1, c_2, \ldots, c_{N_d}$. The assembly of the vector at only N_d points is expected to be faster than full assembly.

What is left is to determine those N_d entries of $\mathbf{f}$ which, for the case of a right-hand side vector, correspond to degrees of freedom of the full-order system. This is another task that is completed in the offline stage. In a DEIM context, in all iterations the evaluated entries are the same. The indices corresponding to these entries are written $\wp = [\wp_1, ..., \wp_{N_d}]^\intercal \in \mathbb{R}^{N_d}$.

$\mathbf{P} \in \mathbb{R}^{N_h \times N_d}$ is a selection matrix, the columns of which are selected columns of N_d-sized identity matrix. If the i-th column of $\mathbf{V}$ corresponds to selected index $\wp_i$, then the i-th column of $\mathbf{P}$ is the $\wp_i$-th column of the identity matrix. In other words,

$$\mathbf{P} = [\mathbf{e}_{\wp_1}, ..., \mathbf{e}_{\wp_{N_d}}] \in \mathbb{R}^{N_h \times N_d}, \tag{3.6}$$

where $\mathbf{e}_{\wp_i} = [0, ..., 0, 1, 0, ..., 0]^\intercal \in \mathbb{R}^{N_h}$ is the $\wp_i$-th column of the identity matrix. In other words,

$$\mathbf{P}^\intercal \mathbf{f} = \mathbf{P}^\intercal \mathbf{V}\mathbf{c}. \tag{3.7}$$

Therefore, the coefficient vector that must be identified at each iteration of the online stage becomes,

$$\mathbf{c}(\mathbf{x}, \boldsymbol{\mu}) = (\mathbf{P}^\intercal \mathbf{V})^{-1} \mathbf{P}^\intercal \mathbf{f}, \tag{3.8}$$

and introducing this expression to equation (3.5) we obtain,

$$\mathbf{f} \approx \mathbf{V}\mathbf{c} = \mathbf{V}(\mathbf{P}^\intercal \mathbf{V})^{-1}\mathbf{P}^\intercal \mathbf{f}, \tag{3.9}$$

where the product $\mathbf{V}(\mathbf{P}^\intercal \mathbf{V})^{-1}$ can be pre-computed and stored in the offline stage. In the online stage, only the product $\mathbf{P}^\intercal \mathbf{f}$ must be evaluated, that is, the $\mathrm{N_d}$ selected components of $\mathbf{f}$. The operator $\mathbf{P}^\intercal$ sets zeros to all the entries of $\mathbf{f}$ except for the $\mathrm{N_d}$ selected ones, so that, in order to compute the product, it is enough to evaluate $\mathbf{f}$ in the selected entries.

To identify the selected indices and the matrix $\mathbf{P}$ a routine that requires as input the vectors of $\mathbf{V}$ has been proposed by the authors of [21]. The steps are described in Algorithm 3.1.

Algorithm 3.1 DEIM algorithm

Input: $\{\mathbf{v}_\ell\}_{\ell=1}^{\mathrm{N_d}} \in \mathbb{R}^{\mathrm{N_h}}$ linearly independent
Output: $\wp = [\wp_1, ..., \wp_{\mathrm{N_d}}]^\intercal \in \mathbb{R}^{\mathrm{N_d}}$
1: $\wp_1 = \max\{|\mathbf{v}_1|\}$
2: $\mathbf{P} = [\mathbf{e}_{\wp_1}]$, $\wp = [\wp_1]$
3: **for** $\ell = 2$ to $\mathrm{N_d}$ **do**
4: $\quad \tilde{\mathbf{V}}$ = first $\ell - 1$ columns of $\mathbf{V}$
4: $\quad$ Solve $(\mathbf{P}^\intercal \tilde{\mathbf{V}})\mathbf{c} = \mathbf{P}^\intercal \mathbf{v}_\ell$
5: $\quad \mathbf{r} = \mathbf{v}_\ell - \tilde{\mathbf{V}}\mathbf{c}$
6: $\quad \wp_\ell = \max\{|\mathbf{r}|\}$
7: $\quad \mathbf{P} \leftarrow [\mathbf{P} \quad \mathbf{e}_{\wp_\ell}]$, $\wp \leftarrow \begin{bmatrix} \wp \\ \wp_\ell \end{bmatrix}$
8: **end for**

Algorithm 3.1 constructs a set of indices based on the provided POD basis $\mathbf{V}$. The procedure assumes that the input basis vectors $\{\mathbf{v}_\ell\}_{\ell=1}^{\mathrm{N_d}}$ are in decreasing order of their corresponding singular value. The first index is selected based on the first basis vector. It is the index that corresponds to the largest in magnitude entry of $\mathbf{v}_1$. This selection is justified by considering the approximation error in the first direction of the POD basis. If one must pick only one component in which to sample, or, precisely evaluate the function, then it is only logical to pick the largest in magnitude, so that the resulting approximation error is as small as possible [76].

Then, sequentially, the algorithm selects for each vector $\mathbf{v}_\ell$ the entry where the residual between $\mathbf{v}_\ell$ and its approximation using the previously selected

interpolants is largest (Line 5 of Algorithm 3.1). In other words, the algorithm greedily seeks to reduce the approximation error by selecting for precise evaluation the point that is worse approximated by interpolating the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_{\ell-1}\}$ at interpolation points $\{\wp_1, \ldots, \wp_{l-1}\}$ [21].

It is worth noting that the DEIM interpolation is uniquely determined by the POD basis $\mathbf{V}$, as the selected interpolation indices depend on the basis.

Based on the described steps so far, the $\mathbf{V}(\mathbf{P}^\intercal\mathbf{V})^{-1}$ is fully defined and stored. The system of equations to be solved in the online phase, as defined in (2.34), after introducing the DEIM approximation of the right-hand side vector, becomes,

$$\mathbf{B}^\intercal\mathbf{A}(\mathbf{x},\boldsymbol{\mu})\mathbf{B}\boldsymbol{\alpha}(\boldsymbol{\mu}) = \mathbf{B}^\intercal\mathbf{V}(\mathbf{P}^\intercal\mathbf{V})^{-1}\mathbf{P}^\intercal\mathbf{f}(\mathbf{x},\boldsymbol{\mu}), \tag{3.10}$$

where the only term of the right-hand side that must be evaluated online is $\mathbf{f} \in \mathbb{R}^{N_h}$ with N_d non-zero entries.

DEIM for matrix operators (MDEIM)

The nonlinear matrix Finite Element operator $\mathbf{A}(\mathbf{x},\boldsymbol{\mu})$ in equation (2.33) can be approximated using a combination of projection to a low-order space and interpolation, in a similar way as shown for the vector function $\mathbf{f}(\mathbf{x},\boldsymbol{\mu})$.

A convenient way to implement the method for matrix operators is to express the matrices as vectors, by vertically stacking their columns one on top of the other. In the following, the vectorized matrix operator is written, $\mathbf{a}(\mathbf{x},\boldsymbol{\mu}) = \text{vec}(\mathbf{A}(\mathbf{x},\boldsymbol{\mu})) \in \mathbb{R}^{N_h^2}$ and the dependence on $(\mathbf{x},\boldsymbol{\mu})$) is suppressed for clarity. The approximation of the operator then reads,

$$\mathbf{a} \approx \mathbf{V}_A(\mathbf{P}_A^\intercal\mathbf{V}_A)^{-1}\mathbf{P}_A^\intercal\mathbf{a}, \tag{3.11}$$

where $\mathbf{V}_A \in \mathbb{R}^{N_h^2 \times N_d}$ contains the POD basis adopted for the approximation of the operator. Its vectors are vectorized matrix elements to be linearly combined to approximate the matrix operator in an N_d sized space. To identify the basis $\mathbf{V}_A$ we follow the same steps as described above, after collecting matrix snapshots and storing them in vector format as,

$$\mathbf{M}_A = [\text{vec}(\mathbf{A}_1)...\text{vec}(\mathbf{A}_{N_s})] = [\mathbf{a}_1...\mathbf{a}_{N_s}]. \tag{3.12}$$

The POD basis $\mathbf{V}_A$ is constructed by applying SVD to the snapshot matrix $\mathbf{M}_A$ and selecting the N_d singular vectorized matrices with the largest corresponding singular values.

$\mathbf{P}_A \in \mathbb{R}^{N_h \times N_d}$ is the selection matrix that defines the interpolation points selected by the DEIM algorithm. The vectorized operator $\mathbf{a} \in \mathbb{R}^{N_h^2}$ features the precisely estimated values that correspond to the selected interpolation points, and must be populated in the online phase.

The matrix approximation reads,

$$\mathbf{A} \approx \text{vec}^{-1}\left(\mathbf{V}_A \left(\mathbf{P}_A^\intercal \mathbf{V}_A\right)^{-1}\right) \text{vec}^{-1}\left(\mathbf{P}_A^\intercal \mathbf{a}\right), \tag{3.13}$$

where vec^{-1} denotes the process of reverting the vectorized operators to matrix ones, by horizontally stacking the vertically stacked columns. The product $\text{vec}^{-1}(\mathbf{V}_A(\mathbf{P}_A^\intercal \mathbf{V}_A)^{-1})$ is pre-computed and stored in the offline stage. Finally, by introducing the stiffness matrix approximation to equation (3.10), the problem becomes,

$$\mathbf{B}^\intercal \text{vec}^{-1}\left(\mathbf{V}_A \left(\mathbf{P}_A^\intercal \mathbf{V}_A\right)^{-1}\right) \text{vec}^{-1}\left(\mathbf{P}_A^\intercal \mathbf{a}\left(\mathbf{x}, \boldsymbol{\mu}\right)\right) \mathbf{B}\boldsymbol{\alpha}(\boldsymbol{\mu}) = \mathbf{B}^\intercal \mathbf{V} \left(\mathbf{P}^\intercal \mathbf{V}\right)^{-1} \mathbf{P}^\intercal \mathbf{f}\left(\mathbf{x}, \boldsymbol{\mu}\right), \tag{3.14}$$

where $\mathbf{B}$ is the POD basis constructed for the solution state approximation.

DEIM for the reduction for hydro-mechanical groundwater flow problems

In the following, the vectorization of matrices by stacking columns is implied, as explained above, but is not always explicitly indicated for clarity of notation.

To apply the DEIM method to all the nonlinear terms in the coupled system at hand, each component in equation (2.41) must be approximated separately. The system becomes,

$$\begin{bmatrix} \mathbf{B_u}^\intercal \hat{\mathbf{K}} \mathbf{B_u} & \mathbf{B_u}^\intercal \mathbf{V}_Q (\mathbf{P}_Q^\intercal \mathbf{V}_Q)^{-1} \mathbf{P}_Q^\intercal \hat{\mathbf{Q}}(p) \mathbf{B_p} \\ \mathbf{B_p}^\intercal \mathbf{V}_C (\mathbf{P}_C^\intercal \mathbf{V}_C)^{-1} \mathbf{P}_C^\intercal \hat{\mathbf{C}}(p) \mathbf{B_u} & \mathbf{B_p}^\intercal \mathbf{V}_H (\mathbf{P}_H^\intercal \mathbf{V}_H)^{-1} \mathbf{P}_H^\intercal \hat{\mathbf{H}}(p, k_s) \mathbf{B_p} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_\mathbf{u} \\ \boldsymbol{\alpha}_p \end{bmatrix}^{i+1} = \begin{bmatrix} \mathbf{B_u}^\intercal \mathbf{V_u} (\mathbf{P_u}^\intercal \mathbf{V_u})^{-1} \mathbf{P_u}^\intercal \hat{\mathbf{f}}_\mathbf{u}(p) \\ \mathbf{B_p}^\intercal \mathbf{V}_p (\mathbf{P}_p^\intercal \mathbf{V}_p)^{-1} \hat{\mathbf{f}}_p(p, k_s) \end{bmatrix}, \tag{3.15}$$

where the matrices $\mathbf{V}_Q, \mathbf{V}_C, \mathbf{V}_H$ contain POD bases for the 3 components of the global matrix that contain nonlinearities, and $\mathbf{P}_Q, \mathbf{P}_C, \mathbf{P}_H$ are the respective selection operators that determine the selected entries to be evaluated online for the construction of each submatrix approximation. In equation (3.15), the notation for matrices $\hat{\mathbf{Q}}, \hat{\mathbf{C}}, \hat{\mathbf{H}}$ does not explicitly indicate that in the online

computations, vectorized versions containing the exact values of only the selected entries are to be computed, but it is implied.

To elaborate on the implementation of DEIM for matrices (MDEIM) the focus is brought to the approximation for one of the submatrices, $\hat{\mathbf{Q}}(\mathrm{p}) \in \mathbb{R}^{\mathrm{N_u} \times \mathrm{N_p}}$, where $\mathrm{N_u}$ is the number of degrees of freedom of the high-fidelity (FE) problem that are related to the description of the displacement, and $\mathrm{N_p}$ the ones that are related to pressure. Then the POD basis $\mathbf{V}_\mathrm{Q} \in \mathbb{R}^{(\mathrm{N_u} \cdot \mathrm{N_p}) \times \mathrm{N_{dQ}}}$ has $\mathrm{N_{dQ}}$ columns and $\mathrm{N_u} \cdot \mathrm{N_p}$ rows, as it is a basis for the vectorized operator. The product $(\mathbf{P}_\mathrm{Q}^\intercal \mathbf{V}_\mathrm{Q})^{-1} \in \mathbb{R}^{\mathrm{N_{dQ}} \times \mathrm{N_{dQ}}}$ is a square matrix of the size of the DEIM approximation. Therefore, the quantity

$$\mathbf{B_u}^\intercal \mathbf{V}_\mathrm{Q} (\mathbf{P}_\mathrm{Q}^\intercal \mathbf{V}_\mathrm{Q})^{-1},$$

cannot be computed and stored at this stage where the DEIM interpolants are vectorized, due to dimension mismatch. That is, in equation (3.15) the operators are in matrix un-vectorized form as in,

$$\mathrm{vec}^{-1}(\mathbf{V}_\mathrm{Q} (\mathbf{P}_\mathrm{Q}^\intercal \mathbf{V}_\mathrm{Q})^{-1}) \in \mathbb{R}^{\mathrm{N_u} \times \mathrm{N_p} \times \mathrm{N_{dQ}}}.$$

This is not explicitly indicated in equation (3.15), and will be implied in the following, to avoid overwhelming notation.

Then, the dimensions of the following product satisfy,

$$\mathbf{B_u}^\intercal \mathbf{V}_\mathrm{Q} (\mathbf{P}_\mathrm{Q}^\intercal \mathbf{V}_\mathrm{Q})^{-1} \in \mathbb{R}^{\mathrm{N_{r_u}} \times \mathrm{N_p} \times \mathrm{N_{dQ}}},$$

where $\mathrm{N_{r_u}}$ is the size of the Reduced Basis for the displacement field. This product may be stored in the form of a 3-dimensional matrix, that has a depth of the DEIM approximation size $\mathrm{N_{dQ}}$. In that sense, and with the goal of eventually evaluating the term,

$$\underbrace{\mathbf{B_u}^\intercal \mathbf{V}_\mathrm{Q} (\mathbf{P}_\mathrm{Q}^\intercal \mathbf{V}_\mathrm{Q})^{-1}}_{\mathrm{N_{r_u}} \times \mathrm{N_p} \times \mathrm{N_{dQ}}} \underbrace{\mathbf{P}_\mathrm{Q}^\intercal \hat{\mathbf{Q}}(\mathrm{p})}_{\mathrm{N_{dQ}} \times 1} \underbrace{\mathbf{B_p}}_{1 \times \mathrm{N_p} \times \mathrm{N_{r_p}}}$$

one may evaluate each component of the 3D matrix in the sense of $\mathrm{N_{dQ}}$ separately. Then, for every component of the term in the $\mathrm{N_{dQ}}$ sense, the product $\mathbf{P}_\mathrm{Q}^\intercal \hat{\mathbf{Q}}(\mathrm{p})$ is a scalar and can be permuted with the matrix $\mathbf{B_p} \in \mathbb{R}^{\mathrm{N_p} \times \mathrm{N_{r_p}}}$. So finally, for each direction i of the $\mathrm{N_{dQ}}$ directions of the MDEIM approximation space, the product,

$$\underbrace{\mathbf{B_u}^\intercal \left(\mathbf{V}_\mathrm{Q} (\mathbf{P}_\mathrm{Q}^\intercal \mathbf{V}_\mathrm{Q})^{-1} \right)_i}_{\mathrm{N_{r_u}} \times \mathrm{N_p}} \underbrace{\mathbf{B_p}}_{\mathrm{N_p} \times \mathrm{N_{r_p}}} = \underbrace{\mathbf{B_u}^\intercal \mathbf{V}_\mathrm{Q} (\mathbf{P}_\mathrm{Q}^\intercal \mathbf{V}_\mathrm{Q})^{-1} \mathbf{B_p}}_{\mathrm{N_{r_u}} \times \mathrm{N_{R_p}}} \tag{3.16}$$

can be precomputed and stored. Thus, N_{dQ} matrices of size $N_{r_u} \times N_{R_p}$ must be stored, as opposed to N_{dQ} matrices of size $N_{r_u} \times N_p$ and N_{dQ} matrices of size $N_p \times N_{r_p}$, resulting in a less memory intensive algorithm, and a more efficient one, since the tasks performed online for the reconstruction of the approximation do not depend on the original problem size N_{r_p}.

Notes on the implementation of vector and matrix DEIM

Algorithm 3.1 results in a selection of N_d entries of the nonlinear function $\mathbf{f}(\mathbf{x}, \mu)$ or $\mathbf{a}(\mathbf{x}, \mu)$ that must be precisely evaluated in the online stage, that is, the functions $\mathbf{f}_d$ and $\mathbf{a}_d$ that are computed online must be equal to $\mathbf{f}$ or $\mathbf{a}$ on the selected entries. To precisely evaluate these entries, some of the elements of the FE must be accessed. The less the elements that need be accessed online, the higher the computational gains of the method. This implies certain interventions to the FE assembly routines that ensure that the loop over all finite elements is restricted to the few necessary ones, or in other words, to the *reduced mesh*. In Figure 3.1 the concept of the reduced mesh is illustrated for the case of a matrix operator. Depending on the selected matrix entry, one or two nodes may have non-zero contributions to the value of the function there, and all the adjacent elements of these nodes must be included in the loop over the reduced mesh.

Since a selected entry might correspond to multiple elements in the reduced mesh, the size of the mesh increases faster than the number of selected entries needed to approximate a nonlinear function. This has a computational cost that may counteract the benefits of the DEIM and MDEIM. To remedy this issue an unassembled variant of DEIM (UDEIM) has been developed in [6], that aims to approximate the unassembled nonlinear quantities. The main advantage of this method is that it results in more sparse reduced meshes, as for every selected entry, only one element is related. However, this method is more memory-intensive than the conventional DEIM as it requires the storage of each element contribution to every node separately. In the offline stage, where multiple snapshots must be stored and SVD must be applied to large snapshot matrices, this can prove to be a significant issue, especially for large problems. Moreover it requires modifications to the assembly routine of the FE code. In this case, in order to reap the benefits of the FEniCs platform and its fast, automatic assembly routine, the conventional DEIM is preferred.

One of the major challenges in implementing DEIM is the fact that the offline stage of the method can be particularly memory intensive. For the

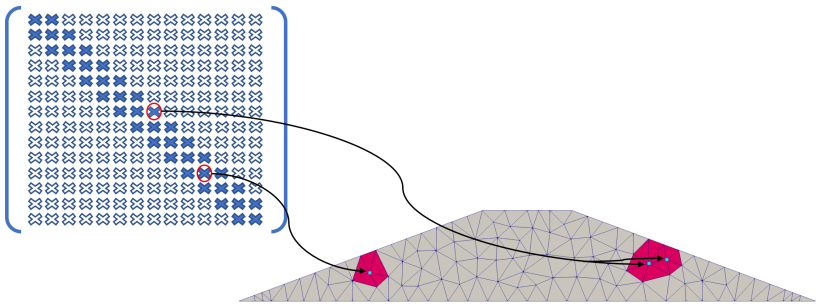


Figure 3.1: Illustration of the concept of the reduced mesh in a FE-MDEIM context. On the left, a sparse matrix FE. Two selected entries of the function by the DEIM algorithm are in red circles. To precisely evaluate these entries, all the adjacent elements of the involved nodes that have a non-zero contribution to the selected matrix entry must be accessed. For each dof, several elements, illustrated on the right with pink color must be accessed.

approximation of a matrix operator $\hat{\mathbf{Q}} \in \mathbb{R}^{N_{\mathbf{u}} \times N_p}$, the method requires the evaluation and storage of snapshot matrices sized $(N_{\mathbf{u}} \cdot N_p) \times N_s$. If sparse matrix structures are used, the size of the stored matrices is the number of non-zero elements, N_{nz}. However, depending on the software that is used, performing some of the required operations on sparse matrices can be very inefficient. The implementation of SVD, DEIM algorithm, or other tasks that require changing the structure of the involved matrices are very demanding when performed on sparse structures. However, as suggested in [78], auxiliary dense matrices that contain only the non-zero elements of the original snapshot matrices can be used. This is done by eliminating all the zero rows of the snapshot matrix. Since the columns of the matrix contain vectorized snapshots of the same operator, the sparsity structure of each row is bound to be the same. Therefore, all the tasks are now performed on N_{nz} sized dense matrices.

State-based Localized Discrete Empirical Interpolation (LDEIM) for a generic nonlinear function f

An important limitation of DEIM is that for some nonlinear systems, particularly when the system exhibits a wide range of behaviors, a large number of DEIM basis vectors is needed for an accurate approximation of the function.

This limits the efficiency of the resulting ROM. In the Localized Discrete Interpolation Method a nonlinear function approximation is constructed by creating not one, but several reduced spaces. Each sub-space is appropriate for the approximation of a certain region of states. Hence, the subspace dimension becomes smaller than the global DEIM approximation space, which translates to a smaller reduced mesh and higher ROM efficiency.

To accomplish that, LDEIM utilizes machine learning techniques both in the offline and the online space. In the offline space, clustering methods are used to identify these state sub-regions among all possible system states. Essentially, similar snapshots are grouped together based on unsupervised learning. In the online stage classification algorithms are used to identify which subspace is most suitable for the approximation of the current system state as the computation proceeds.

The method can handle a large number of large-sized snapshots and a large number of local approximation spaces, thanks to the use of a DEIM-based feature extraction that results to a low-dimensional representation of the snapshots that is used for the clustering and the classification processes.

It is worth noting that the idea of localized DEIM and localized POD has been explored in various settings. Two versions of LDEIM are presented in [88], a parameter-based and a state-based one. In the parameter-based LDEIM the local spaces are defined based on sub-regions of parametric values and the online selection of an appropriate sub-space is also based on the parameters. Moreover, groups of snapshots can be created based on time for a transient problem [83]. In the problem of groundwater-flow through porous media however, the studied parameters result in similar solutions and system states, making the parameters an unfit criterion for clustering and classification. Furthermore, the authors found that for this problem, a time-based division of the snapshots does not result in smaller approximation spaces. Hence, the state-based version of LDEIM is selected and presented here.

The implementation of the state-based localized DEIM method as proposed in [88] implies the pre-construction of a global DEIM scheme as described in Section 3.2. In the offline stage, a snapshot matrix $\mathbf{M}_d$ of fully assembled functions must be populated as in equation (3.1)) and the global N_d-sized DEIM interpolants $\mathbf{V}$ and $\mathbf{P}$ must be computed.

Offline stage: snapshot clustering and sub-spaces identification

Once the global DEIM scheme is constructed, the sub-spaces that correspond to sub-regions of system states must be identified by clustering the snapshots

of $\mathbf{M}_d$ into groups based on their similarity. The measure of similarity is in this case based on the Euclidean distance of a DEIM-based representation of the snapshots, as opposed to a Euclidean distance of the full snapshots. To create the low-order representation of the snapshots, the interpolation points of the global-DEIM scheme are selected as,

$$\mathbf{M}_{ld} = \left[\mathbf{P}^\intercal \mathbf{f}^1 ... \mathbf{P}^\intercal \mathbf{f}^{N_s}\right] \in \mathbb{R}^{N_d \times N_s}. \tag{3.17}$$

This selection of entries is called feature extraction in the context of machine learning.

Subsequently, a clustering algorithm is used to group the elements of $\mathbf{M}_{ld}$. In this work, the K-means algorithm is used, and more particularly the python implementation of the scikit-learn package is used [87]. The basic steps in K-means algorithm are summarized in Algorithm 3.2.

The method requires as input the number k of groups of snapshots, or clusters of snapshots to be created. Initially k centroids $\mathcal{F}_j$ are selected from the dataset $\mathbf{M}_{ld}$ such that they are distant from each other. This technique typically yields better results than random initialization [87]. The method iteratively optimizes the clusters by assigning each snapshot to the cluster with the nearest centroid. Then the centroids are updated taking the mean value of all elements of their respective cluster. The difference between the old centroids and the new centroids is computed and the process is repeated until this value is smaller than a threshold.

Algorithm 3.2 K-means algorithm

Input: Number of clusters k, dataset $[\mathbf{P}^\intercal \mathbf{f}_1 ... \mathbf{P}^\intercal \mathbf{f}_{N_s}] \in \mathbb{R}^{N_d \times N_s}$, stopping tolerance tol

Output: Clusters of snapshots $\mathbf{M}_{ld}^j$, centroids $\mathcal{F}_j$

1: Initialize clusters with centroids $\mathcal{F}_j$ distant from each other

2: Assign each element $\mathbf{P}^\intercal \mathbf{f}_z$ to nearest centroid cluster: $\underset{j \in [1,k]}{\text{minimize}} |\mathbf{P}^\intercal \mathbf{f}_z - \mathcal{F}_j|_{\mathcal{L}_2}$

3: Compute new centroids $\mathcal{F}_j$ as mean of all elements of cluster j

4: Repeat steps 2-3 until $|\mathcal{F}_j^{old} - \mathcal{F}_j^{new}| \leq$ tol

The result of the clustering technique is a partition $\mathbf{M}_{ld}^1 \uplus \mathbf{M}_{ld}^2 \uplus \ldots \uplus \mathbf{M}_{ld}^k$ of the snapshot matrix $\mathbf{M}_{ld}$ and the corresponding centroids. The clusters contain low-order approximations of the originally collected snapshots. Snapshot subsets $\mathbf{M}_d^1 \uplus \mathbf{M}_d^2 \uplus \ldots \uplus \mathbf{M}_d^k = \mathbf{M}_d$ are obtained from the associated full-order snapshots.

POD bases and interpolation points are identified for each subset of snapshots $\mathbf{M}_d^i$, implementing SVD and the DEIM algorithm as described in section 3.2. The resulting approximations spaces and selection matrices read $(\mathbf{V}^1, \mathbf{P}^1), \ldots, (\mathbf{V}^k, \mathbf{P}^k)$.

This process must be performed for all vector and matrix operations that need be approximated with DEIM. Eventually, the interpolants that correspond to all subspaces must be stored. Re-examining the approximation of the component $\hat{\mathbf{Q}}$ of the global matrix, equation (3.16) now becomes,

$$\underbrace{\mathbf{B_u}^\intercal \left(\mathbf{V}_Q^j (\mathbf{P}_Q^j{}^\intercal \mathbf{V}_Q^j)^{-1}\right)_j}_{N_{r_u} \times N_p} \underbrace{\mathbf{B_p}}_{N_p \times N_{r_p}} = \underbrace{\mathbf{B_u}^\intercal \mathbf{V}_Q^j (\mathbf{P}_Q^i{}^\intercal \mathbf{V}_Q^j)^{-1} \mathbf{B_p}}_{N_{r_u} \times N_{r_p}}, \quad j \in [1, k], \tag{3.18}$$

where $\mathbf{V}_Q^j$ and $\mathbf{P}_Q^j$ correspond to the j-th local DEIM approximation space. That is, k matrices of size $N_{r_u} \times N_{r_p} \times N_{dQ}$ must be pre-stored.

Online stage: identification of appropriate approximation space

To efficiently assemble nonlinear operators in the online phase, the appropriate cluster must be identified in every time-step or iteration. To identify the cluster, the lastly evaluated function must be classified and assigned to one of the k clusters. This can be done using a classification algorithm. In this case the *Nearest Neighbor Search* was used, more particularly the implementation found in the scikit-learn python package [87]. This is a simple algorithm that assigns an element to a cluster based on its proximity to the pre-computed centroids $\mathcal{F}_1, \ldots, \mathcal{F}_k$, in a Euclidean context. It is very similar to step 2 of the K-means Algorithm 3.2.

Note that the computed centroids are in the form of low-order DEIM-based representations. Therefore for the classification a similar representation of the new element must be evaluated, based on the global DEIM approximation. In this work, the classification process occurs once in every time step and not within the time steps. This is because it has been observed that repeating the classification process after each iteration of the Picard scheme results in the same cluster selection, that is, the cluster does not change within a time step, but rather it typically changes from time step to time step.

Therefore, to assemble the function $\mathbf{f}^{i+1}$, where i + 1 denotes the current time-step, one must classify the element $\mathbf{P}^\intercal \mathbf{f}^i$, where $\mathbf{f}^i$ is the lastly evaluated function. The function actualization $\mathbf{f}^i$ is approximated using the previously selected DEIM sub-space, that may or may not be different than the one selected in the current time step i + 1.

3.3 Application

In this section a ROM is created based on the methodologies that are discussed in Section 3.2. A parametric problem is setup and solved using FEM and ROM, for the purpose of comparing and evaluating the ROM performance in terms of efficiency and accuracy.

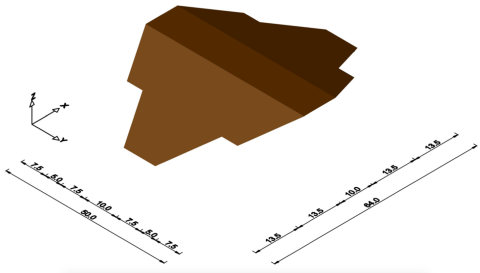
Problem setup

The examined problem is relevant in the context of tailings dams [77]. This is a type of earthfill embankment dam designed for the permanent storage of mining waste. Those dams are typically constructed near mining facilities, and it is usual practice that their crest level is raised by adding layers of fill-material on top of the existing embankment, in order to increase the reservoir capacity, to accommodate increased waste production. Hence, the problem that is studied here is one in which a distributed load is applied on the top boundary of a 3-dimensional domain that represents a dam. The added load simulates the deposition of a fill material layer in the process of raising the dam level. Literature suggests that 2 dimensional models cannot reflect the complex and variating seepage field [68]. Thus, 3 dimensional models have been proposed for the stability study of tailings dams [67, 127].

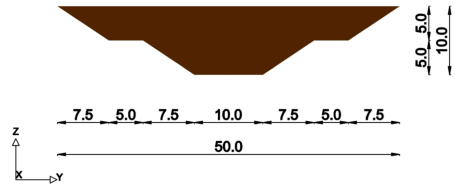
This illustrative problem has been examined in a previous publication by the authors of [77]. All the details of the problem setup mentioned in the present work are identical to the ones used in the previous publication.

The geometry shown in Figure 3.2 approximates an embankment constructed in a narrow, steep sided valley. It is a generic, invented geometry, meant to simulate common conditions in the construction of embankment dams.

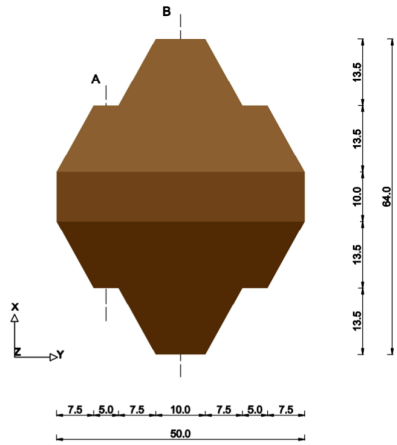
The domain is a 50 m-long prism, with 2 axes of symmetry and a variating cross-section along the Y axis. The parts that are founded on the lateral slopes have smaller cross sections (Figure 3.2d Top). The side slopes have been assumed excavated in two levels, creating two 5 m-high slopes of inclination 1,5:1.



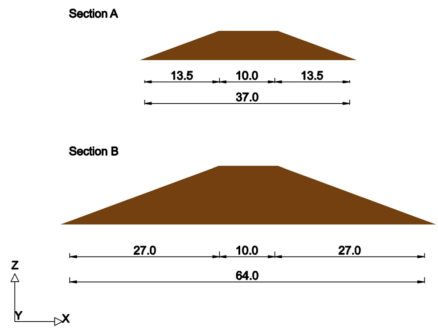
(a) 3D Isometric view



(b) Front view



(c) Top view

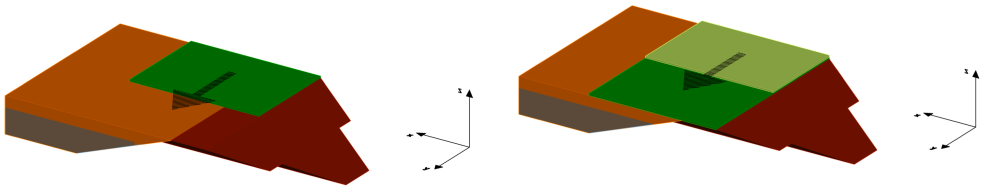


(d) Cross sections Top: Section plane A Bottom: Section plane B

Figure 3.2: Views and cross sections of the 3 dimensional domain. The geometry assumes an embankment that has been constructed in a narrow valley, and has its foundations on two side slopes at the two extremes in the direction of Y axis. The side slopes are not displayed.

A 3D tetrahedral mesh was created in Gmsh open-source mesh generating software [38]. The mesh is made out of 41477 tetrahedral and triangular cells and 9896 nodes. Using the Taylor-Hood (P2-P1) element for the description of the displacement and pressure fields, the resulting system consists of 202179 and 9896 degrees of freedom for the two fields respectively.

The initial condition is evaluated by solving the steady state problem for an upstream water level at 7 m. In the upstream side, the deposition of tailings material, with specific weight γ_t reaches the level of the dam's crest. The material is saturated up to the level of 7m. The added material, of specific



(a) After 2,5 days (b) After 7,5 days

Figure 3.3: Deposition of the two 50 cm - thick layers of fill material on the top and upstream side of the structure. The first layer is shown in dark green color and the second in light green color. After 2,5 days of loading half of the first layer has been deposited. After 7,5 days the first layer and half of the second layer have been deposited. Upstream is in the direction of the x axis. Unsaturated tailings are depicted in orange, saturated tailings in gray. The lateral foundation slopes are not depicted.

weight γ_f, is placed in two layers of 50 cm each. The first layer is deposited in the first 5 days, gradually, across the Y axis. Then the second 50 cm layer is added on top, during days 5-10, again, gradually across the Y axis. The bottom boundary is mechanically constrained and hydraulically impermeable. A fixed water pressure head of 7 meters is imposed in the upstream boundary, which remains fixed throughout the simulation. Moreover, a mechanical load is imposed as a Neumann condition in the upstream boundary (Equation (2.9)), resulting from the lateral earth pressure due to the tailings material in the impoundment. A seepage condition (Equation (2.13)) is imposed on the downstream boundary. During the level raise phase in a tailings dam life, thin layers of fill material are deposited along the length of the structure. The load increases gradually along the Y axis. In Figures 3.3a and 3.3b this loading sequence is illustrated for clarity. Two layers of soil are placed on top of the dam, in order to raise the level of the dam by 1 meter.

The values of the physical properties and numerical parameters used in the model are given in Tables 3.1 and 3.2. The values were chosen such that they fall into ranges that are usually observed in tailings dams [14, 92].

The parameters that are considered to vary in this work are both hydraulic

Parameter	Symbol	Units	Value
Gravitational acceleration	g	m/s	10
Water bulk modulus	K_w	MPa	2.2×10^3
Specific weight of water	γ_w	kN/m^3	10
Embankment fill soil material			
Particle density	ρ_s	kg/m^3	2.7×10^3
Young's Modulus	E	MPa	[50, 150]
Poisson's ratio	ν	-	0.3
Porosity	η	-	0.38
Saturated VWC	Θ_s	-	0.38
Residual VWC	Θ_r	-	0.038
Parameter ($\approx$ inverse of air entry suction head)	α	m^{-1}	[0.01, 1]
Fitting Parameter	m	-	0.184
Saturated hydraulic conductivity	k_s	m/s	$[10^{-9}, 10^{-7}]$
Tailings and added fill material			
Added fill material specific weight	γ_f	kN/m^3	21
Tailings specific weight	γ_t	kN/m^3	21
Tailings friction angle	ϕ	$^\circ$	35

Table 3.1: Values of physical soil properties used in the model. Value ranges of properties that are considered as parameters in the following reduced order models are given in [ ]

Discretization	
Mesh	41477 elements, 9896 nodes, unstructured
FE Displacement	P2
FE Pressure	P1
Time integration: θ-scheme	
θ	0.75
Time step Δt	0.25 days

Table 3.2: Numerical parameters used in the model

soil properties. Namely, the saturated hydraulic conductivity k_s, the parameter α, which is related to the air entry value of the soil, and the Young's modulus E are considered as variating parameters. These 3 properties are chosen due to the high level of uncertainty they often feature in geotechnical applications, that renders them appropriate objects of examination using data assimilation methods. As explained in Section 4.1 Model Order Reduction is particularly advantageous in the context of data assimilation or other many-query problems, it therefore seems fitting to study uncertain parameters using a ROM.

The parametric domain is taken such that the extreme values of ranges are realistic in the framework of earthfill dams [14] .

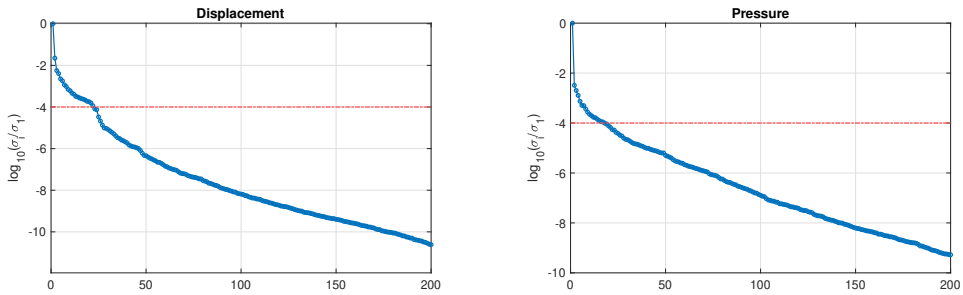
In the following, 3 setups will be examined. The first parametric problem that is presented, considers only one parameter, namely k_s, the second considers all 3 parameters, k_s, E, α and the final setup considers one parameter but a denser FE mesh. For every setup, both a global DEIM and a local DEIM approach are used. The goal of the study is to examine how the computational efficiency achieved with a ROM scales with increasing number of parameters and an increasing size of the original problem.

ROMs created using different reduction methods are referenced with the abbreviations RB, RB-DEIM, RB-LDEIM. The first one refers to a ROM with low-order solution space approximation using the Reduced Basis method and no hyper-reduction. The other two refer to ROMs with hyper-reduction using the global and localized versions of DEIM respectively.

ROM for a problem with 1 parameter: saturated hydraulic conductivity

In this section, only the saturated hydraulic conductivity is considered as a parameter. The other 2 mentioned properties remain fixed at E = 80 MPa and $\alpha = 0.1\ \mathrm{m}^{-1}$.

The solution space reduction is performed as explained in Section 2.3 using the Reduced Basis method. In this case, to create the Reduced Basis, snapshots were taken for $k_s = [10^{-9}, 5 \times 10^{-9}, 10^{-8}, 5 \times 10^{-8}, 10^{-7}]$ m/s. These 5 values are enough to create an accurate basis. For this and all the following cases, the truncation tolerance for the solution space basis is -4. Based on this tolerance, the most relevant orthonormal vectors are selected to form a low-order space basis, as indicated in equation (3.4). In Figure 3.4, the fast decay of the singular values that correspond to the vectors the result from applying SVD to the snapshot matrices for both fields is illustrated.



(a) Displacement (b) Pressure

Figure 3.4: Snapshot truncation based on singular values. The y axis is a log plot of the normalized singular values that correspond to the first singular vectors. The red line represents the truncation tolerance -4.

The author's experience using DEIM has led to the conclusion that for this type of problem, more snapshots are required for the identification of a low-order space for system approximation than for solution space approximation. Therefore, more snapshots were taken after the Reduced Basis was identified, this time using a ROM. Snapshots of the fully assembled submatrices $\hat{\mathbf{Q}}$, $\hat{\mathbf{C}}$, $\hat{\mathbf{H}}$ and the right-hand side vector $[\hat{\mathbf{f}}_{\mathbf{u}}, \hat{\mathbf{f}}_{\mathrm{p}}]^{\intercal}$ were taken for 10 parametric values. The assembled operators that correspond to every iteration of every time-step were saved in snapshot matrices.

In Figure 3.5 the singular values that correspond to each approximated function are plotted. The final number of othronormal vectors in the POD basis for each function depends on the selected truncation tolerance and the rate at which the singular values decrease. Figure 3.5 shows how the singular values that correspond to coupling matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{C}}$ decrease much faster than those that correspond to matrix $\hat{\mathbf{H}}$ and the right-hand side vector $[\hat{\mathbf{f}}_{\mathbf{u}}, \hat{\mathbf{f}}_{\mathrm{p}}]^{\intercal}$. The experience of the author shows that a tolerance as large as -4 yields sufficient accuracy for the coupling matrices, and very few vectors are selected to approximate them. Specifically, for the tested cases that are presented in the following, the POD bases for matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{C}}$ were of size as small as 2 vectors.

For the approximation of matrix $\hat{\mathbf{H}}$ and the right-hand side vector, the singular values drop slower. This may be related to the fact that these functions result from parametrically dependent variational forms. The number of vectors needed in a POD basis generally depends on how complex the manifold of all possible values of the underlying function is. It appears that $\hat{\mathbf{H}}$ and the vector

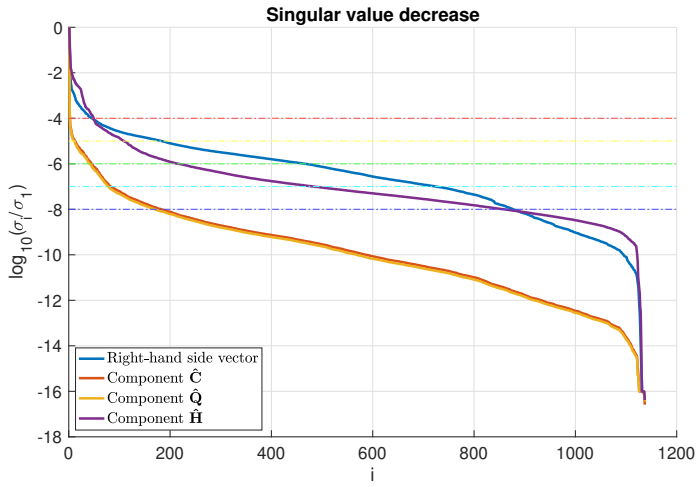


Figure 3.5: Singular values for the functions that are approximated using DEIM. Truncation tolerances that are used for the different functions and range from -8 to -4 are illustrated with dashed lines.

$[\hat{\mathbf{f}}_{\mathbf{u}}, \hat{\mathbf{f}}_{\mathrm{p}}]^{\intercal}$ can take a wide range of values in the examined parametric problem. Therefore, the size of the reduced mesh used to solve the problem, depends on the truncation tolerance applied for these two terms.

In Figure 3.6 the numbers of elements that are selected by the DEIM procedure (Algorithm 3.1) are shown. This example corresponds to the case where the first singular vectors of the snapshot matrix that corresponds to the $\hat{\mathbf{H}}$ component and the right-hand side vector are retained, based on a tolerance of -8. The tolerance used for the other matrices, $\hat{\mathbf{Q}}$ and $\hat{\mathbf{C}}$, is larger, namely -4.

In the following, ROMs that result from different tolerances are presented. All refer to a variation of the tolerance for component $\hat{\mathbf{H}}$ and right-hand side vector $[\hat{\mathbf{f}}_{\mathbf{u}}, \hat{\mathbf{f}}_{\mathrm{p}}]^{\intercal}$.

It is worth noting that the elements selected as a result of the implementation of Algorithm 3.1 to the POD bases that correspond to all components are similar. They are mostly located in the unsaturated part of the domain, and close to the phreatic line. Given the fact that the nonlinearities examined are activated only in the unsaturated part, this result is reasonable and in accordance to expectations.

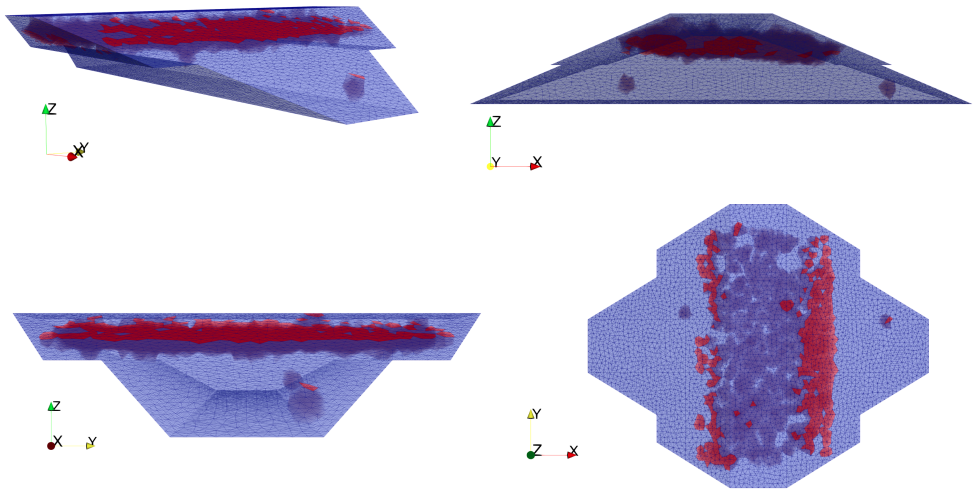


Figure 3.6: Illustration of the reduced mesh that is yielded using a truncation tolerance of -8 for singular values that correspond to the $\hat{\mathbf{H}}$ global matrix component. The elements that must be accessed online are highlighted with red color.

In Table 3.3, several different models are compared in terms of efficiency for the solution of the problem introduced in Section 3.3. The problem is solved for a parametric value that was not sampled in the offline stage, namely $k_s = 5.5 \times 10^{-8}$ m/s.

In the first column the truncation tolerance for the DEIM bases for terms $\hat{\mathbf{H}}$ and $\mathbf{f}$ is listed. In the second column the number of finite elements that are visited online for the assembly of matrix and vector operators are listed. The total number of elements in the full mesh is listed in the last two rows. The problem has 54 unknowns that result from the Reduced Basis method, except for the case where the problem is solved with plain FE. Of these 54 unknowns, 30 are related to the description of the pressure field, and 24 to the displacement field.

The total iterations needed for the problem to be solved (note that there are 40 time steps, with a fixed time step over 10 days), are usually less when the problem is solved with the Reduced Basis method. When DEIM is used, in some cases in which a larger truncation tolerance is used, more iterations are needed for convergence. Similar results have been reported in [6] for a

	Truncation tolerance	Elements for assembly	Problem unknowns (H+M)	Total iterations	Total duration (s)	Time gain T_{RO}/T_{FO}	Assembly duration (s)	Solving duration (s)
Global DEIM +RB	-8	10269	30+24	114	418.28	1/20.27	1.67	0.0003
	-7	10115	30+24	114	419.03	1/20.23	1.48	0.0003
	-6	8417	30+24	280	887.31	1/9.56	1.30	0.0003
	-5	4725	30+24	456	Loss of convergence			
Local DEIM 5 clusters +RB	-8	7797	30+24	126	405.53	1/20,91	1.25	0.0003
	-7	7696	30+24	121	450.30	1/18.83	1.07	0,0003
	-6	6329	30+24	128	365.63	1/23.19	1.07	0,0003
	-5	4717	30+24	859	Loss of convergence			
Local DEIM 10 clusters +RB	-8	8078	30+24	151	757.58	1/11.19	1.56	0.0003
	-7	6525	30+24	174	801.87	1/10.57	1.35	0.0003
	-6	5971	30+24	117	306.15	1/27.70	0.92	0.0004
	-5	4480	30+24		Loss of convergence			
RB		41477	30+24	112	1411.54	1/6.01	10.86	0.0003
FEM		41477	212075	383	8479.53	1	3.988	9.6730

Table 3.3: Results for solving a hydro-mechanical parametric problem with one parameter, k_s, using different Model Order Reduction schemes (described in Section 3.2). A FE model is compared to a ROM where merely solution state reduction has been performed with the Reduced Basis method (written RB), ROMs using DEIM or Localized DEIM with different truncation tolerances for functions $\hat{\mathbf{H}}$ and $\mathbf{f}$.

Newton-Raphson scheme. The fact that the RB-DEIM model may require more iterations can be an issue for the resulting efficiency of the model, as it counteracts the gains that are related to fast assembly. This is possibly related to the accuracy of the approximated functions. Nevertheless, it seems that when an appropriate tolerance is selected for hyper-reduction, the number of additional iterations required is small enough and the final computational gain is still significant.

The column titled "Time gain T_{RO}/T_{FO}" shows the ratio between the durations of each ROM and the FEM - subscript RO stands for Reduced Order and FO for Full Order.

The column titled "Assembly duration" features the time it takes for the

reduced operators to be assembled, which in the case of the RB model, includes the full assembly of the matrices and vectors and their projection to the reduced space, and in the case of RB-DEIM models, it includes assembly in the reduced mesh and projection to the system approximation spaces, and the solution approximation space. This column, and the next one, titled "Solving duration" refers to per iteration times and the values are calculated as the average of all needed iterations.

A review of the results shows that for this case with one parameter, both the RB-DEIM and the RB-LDEIM models yield significant efficiency gains. The fastest case is achieved using a tolerance of -6 to construct a model with LDEIM.

The error in most cases remains within acceptable levels, in the order of magnitude of 10^{-4}, as is shown in Figure 3.7. This is not the case for schemes with a too large truncation tolerance for hyper-reduction (all schemes with -5, global DEIM scheme with -6 tolerance). There, the approximation error is too large and eventually the convergence is lost. The error at a given timestep is estimated as,

$$e = \left\| \frac{X_{RB} - X_{FEM}}{X_{FEM}} \right\|_2, \tag{3.19}$$

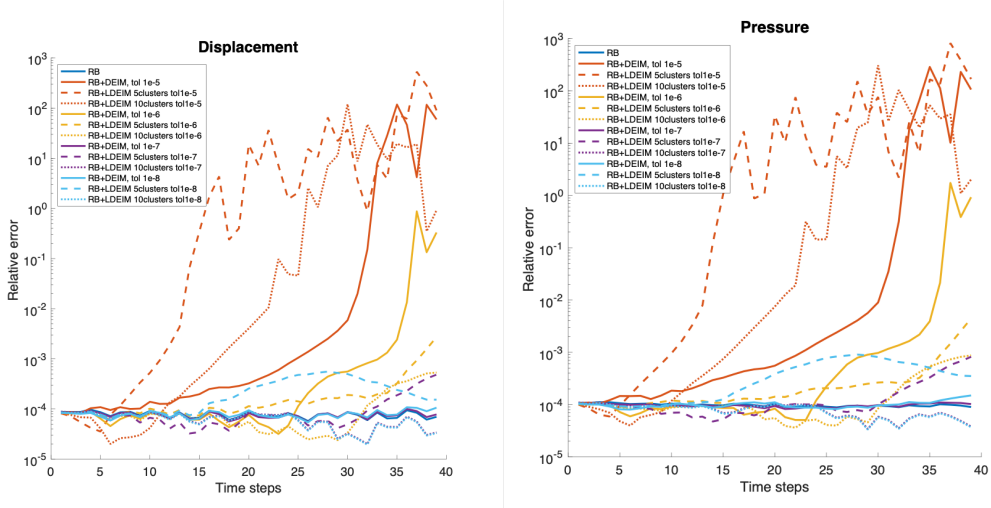
where X_{RB} and X_{FEM} represent the approximation and the high-fidelity solution respectively.

ROM for a problem with 3 parameters

In this Section a similar parametric problem is solved, this time considering 3 parameters, the saturated hydraulic conductivity k_s, the coefficient α, which is related to the air entry value of the soil and appears in Equation (2.15), and the soil's Young's modulus E. The examined value ranges for these parameters are given in Table 3.1.

The offline stage, collecting snapshots of solution vectors to create the solution space approximation is run all over again, this time sampling values in the 3-dimensional parametric space. Specifically, to build the POD basis for solution space approximation, 3 values were sampled for each parameter, resulting in a total of 27 parametric value combinations, selected such that they are evenly spaced along the log scale of each parametric domain.

For the construction of the system approximation spaces (hyper-reduction) more values were sampled. Specifically 5 values for each parameter were selected in a similar way, resulting in 125 sampled values. In this case, the



(a) Displacement field (b) Pressure field

Figure 3.7: Relative error ROM-FEM for pressure and displacement fields.

calculations were performed using the RB model that had been previously constructed. It is worth noting, that the offline stage for this case that examines a multi-dimensional parametric space, is significantly longer than in the previous case. However, as can be sheen in Table 3.4, the resulting ROMs do not drastically differ in size from the ones created in Section 3.3.

The RB method results in a POD basis with 75 vectors, which is, as expected, larger than the 54-vector basis used in Section 3.3, but remains in the same order of magnitude. The same holds true for the reduced mesh used in each scheme that is presented in Table 3.4. The resulting number of visited elements is in some cases larger but not in a different order of magnitude than in the 1-parameter case. This indicates that even though more parameters are examined, the solution manifold and the system state manifold do not feature much larger variation, allowing for the costs of ROMs to remain low.

Reviewing Table 3.4, it seems that LDEIM has a more pronounced comparative advantage compared to the global DEIM in this case, where 3 parameters are considered. In some cases, using a localized POD basis for system approximation results in faster convergence. This observation is in accordance to previous studies that report more accurate and robust results using localized

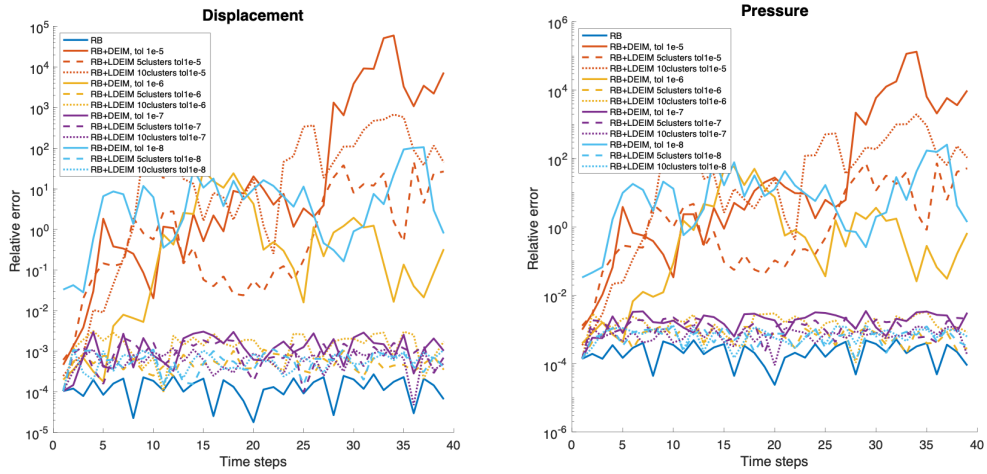
	Truncation tolerance	Elements for assembly	Problem unknowns (H+M)	Total iterations	Total duration (s)	Time gain T_{RO}/T_{FO}	Assembly duration (s)	Solving duration (s)
Global DEIM +RB	-8	12379	51+24	250	917.28	1/9.24	1.67	0.0004
	-7	11966	51+24	244	896.88	1/9.45	1.60	0.0004
	-6	10474	51+24	Loss of convergence				
	-5	6616	51+24	Loss of convergence				
Local DEIM 5 clusters +RB	-8	10216	51+24	132	554.57	1/15.29	1.58	0.0004
	-7	9672	51+24	121	372.70	1/22.75	1.61	0.0004
	-6	9418	51+24	121	385.42	1/22.00	1.08	0.0004
	-5	7884	51+24	Loss of convergence				
Local DEIM 10 clusters +RB	-8	10823	51+24	152	657.29	1/12.90	1.31	0.0004
	-7	10741	51+24	117	367.03	1/23.10	1.33	0.0004
	-6	9219	51+24	123	389.19	1/21.79	1.36	0.0004
	-5	8047	51+24	Loss of convergence				
RB		41477	51+24	102	1385.40	1/6.12	11.24	0.0004
FEM		41477	212075	383	8479.53	1	3.988	9.6730

Table 3.4: Results for solving a hydro-mechanical problem with different Model Order Reduction schemes (described in Section 3.2). In this problem 3 parameters are considered k_s, α and E. This table contains data about the the problem solution for values $k_s = 7e-8$ m/s, $\alpha = 0.1$ m^{-1} and E = 80 MPa. A FE model is compared to a ROM where merely solution state reduction has been performed with the Reduced Basis method (written RB), ROMs using DEIM or Localized DEIM with different truncation tolerances for functions $\hat{\mathbf{H}}$ and $\mathbf{f}$.

bases reduced order schemes, instead of global ones, for systems with large parameter variation [5, 39].

Scaling with full-order model size

In this Section the problem is solved with a much denser FEM mesh, in order to examine how the efficiency gain scales with the size of the full-order problem. One parameter, k_s is considered, and aside from the mesh, every other aspect



(a) Displacement field

(b) Pressure field

Figure 3.8: Relative error ROM-FEM for pressure and displacement fields.

in the setup of the problem remains the same. The most notable conclusion to be drawn from Table 3.5, relates to the gain in computational efficiency achieved in this scheme. A comparison of the time gain columns in Table 3.5 and Table 3.3, shows clearly that DEIM results to a significantly higher speedup factor when applied to FE problems with a larger size.

Moreover, as shown in Figure 3.9, the gain in computational efficiency does not entail a loss of accuracy with respect to the RB model. Again, the method yields accurate results provided a small enough truncation is used for the creation of POD bases in the hyper-reduction process.

Finally, it seems once again that raising the number of clusters does not result to notably more efficient models.

3.4 Conclusions and discussion

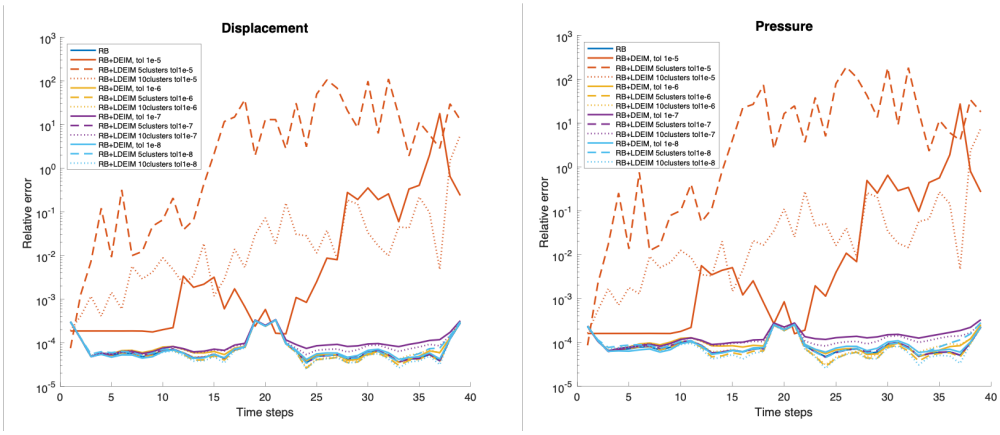
In this Chapter, Model Order Reduction was implemented for the coupled system of hydro-mechanical equations that governs the water flow through partially saturated soil. Combinations of the Reduced Basis method with the DEIM and LDEIM methods were implemented, and their benefits were

	Truncation tolerance	Elements for assembly	Problem unknowns (H+M)	Total iterations	Total duration (s)	Time gain T_{RO}/T_{FO}	Assembly duration (s)	Solving duration (s)
Global DEIM +RB	-8	12419	30+25	108	498.95	1/66.88	2.21	0.0003
	-7	12213	30+25	111	494.59	1/67.47	1.99	0.0003
	-6	8579	30+25	163	576.53	1/57.88	1,67	0,0003
	-5	3667	30+25	Loss of convergence				
Local DEIM 5 clusters +RB	-8	9629	30+25	120	604.07	1/55.24	2.21	0.0004
	-7	9477	30+25	126	572.58	1/58.28	2.02	0.0004
	-6	6885	30+25	105	486.81	1/68.55	1.61	0.0004
	-5	3904	30+25	Loss of convergence				
Local DEIM 10 clusters +RB	-8	7985	30+25	118	539.41	1/61.86	1.81	0.0004
	-7	7646	30+25	99	464.00	1/71.92	1.69	0.0004
	-6	6398	30+25	110	466.17	1/71.58	1.56	0.0004
	-5	3904	30+25	Loss of convergence				
RB		117088	30+25	106	3863,55	1/8.64	31,87	0,0002
FEM		117088	568850	526	33369,55	1	10,70	53,1154

Table 3.5: Results for solving a hydro-mechanical parametric problem with one parameter, k_s, using different Model Order Reduction schemes (described in Section 3.2). A FE model is compared to a ROM where merely solution state reduction has been performed with the Reduced Basis method (written RB), ROMs using DEIM or Localized DEIM with different truncation tolerances for functions $\hat{\mathbf{H}}$ and $\mathbf{f}$.

examined and compared. A 3D parametric illustrative problem was solved with the ROMs that results form implementing RB, DEIM and LDEIM, in order to compare their performances. The problem consists of mechanically loading the top boundary of an embankment dam. Three variations of this problem was considered: one considering a single parameter; another with 3 parameters and a third were the single parameter problem was solved with a denser FE mesh.

It was found that POD-based ROMs particularly when combined with hyper-reduction via DEIM and LDEIM result in a significant increase of the computational efficiency in solving the examined problem. Specifically, it was shown that for the single-parameter case the solution was up to 27 times faster



(a) Displacement field (b) Pressure field

Figure 3.9: Relative error ROM-FEM for pressure and displacement fields.

than when solved with FEM. For the original problem with 3 parameters it was up to 23 times faster, while for the single parameter case with a denser mesh it was up to 72 times faster. These results imply that the method yields a speedup that scales with increasing size of the full-order problem. However, it is worth noting here the coarsest FE mesh that yields convergent results should be the finest FE mesh considered, as there is no need for further refining. This sets an upper bound to the tendency for increasing gains with increasing size of the full-order model. In this work, a finer mesh than the coarsest acceptable has been used in Section 3.3, in order to demonstrate the effect.

It was shown that the size of POD-basis chosen for nonlinear function approximation affects the accuracy and the efficiency of the resulting model. An approximation space of lower order may result in a smaller number of visited elements by DEIM, and consequently to faster assembly. However, selecting a smaller basis may also lead to slower convergence and thereby, to a slower solver overall.

Finally it was shown that LDEIM yields in all cases the fastest solution among the examined ROMs, and is particularly advantageous with respect to DEIM for cases in which a higher parametric space is considered. Increasing the number of clusters from 5 to 10 does not particularly speedup the solution.

All things considered, the combination of RB and DEIM is an appropriate method for problems of hydro-mechanically coupled flow through porous media. The choice of truncation tolerance and, in the case of localized DEIM,

the choice of number of clusters, is not straightforward. Moreover, there is no specific criterion to select these numbers a priori. The experience that was acquired through the present work is that the truncation tolerance should be chosen smaller than what is used in the solution approximation POD basis.

It is worth noting here, that the performance of ROMs in terms of efficiency, are highly dependent on the specific problem under investigation. Depending on the specific setup, the method may result to faster or slower ROMs. This remains to be explored in future implementations of the method for other types of hydro-mechanical coupled problems.

Chapter 4

Parametric inference via inverse problem solving

This chapter is based on:
C. Nasika, P. Díez, P. Gerard, T.J. Massart, S. Zlotnik, Hydro-mechanical soil parameters identification in earthfill dams using Reduced Order Models
In preparation

4.1 Introduction

This Chapter focuses on the simultaneous identification of the Young's modulus E of the soil, as well as two hydraulic soil parameters, namely the saturated hydraulic conductivity k_s, and the parameter α, that characterizes the soil retention. The method is here applied to earthfill dams, but may be extended to any type of soil structure. Several applications of inversion techniques in earth dams using field data can be found in the literature. In those cases the authors use measurements of hydraulic pore pressures as well as outflow rate data from weirs or flowmeters installed in the structures, in order to remedy issues with the uniqueness and reliability of the inverse modeling results. The authors of [22, 132] and [52] used pore pressure and water discharge field measurements taken during a reservoir impounding process, to back-calculate the permeability of fractured rocks in existing dams foundations. The authors of [112] used the same method to back-calculate the permeability of different parts (foundation, core, filters, shell) of an existing earthfill dam based on actual pressure measurements.

There are different ways to formulate and solve an inverse problem. The method that will be examined in this paper involves the minimization of an objective function measuring the distance between observed and numerically computed data. Other methods include the reformulation of the forward problem to one that seeks for the unknown parameters [42, 102]. This method was developed in the early stages of studying inverse problems in geomechanics, and implies several simplifying physical assumptions that are not valid in most real problems. Another commonly used method is the class of probabilistic methods of combining prior knowledge and field measurements to estimate parameters using Bayesian techniques [89, 84, 118, 124, 125].

As mentioned in Chapter 1, the issue of ill-posedness, inherent in the inverse problem for hydraulic parameter identification, that may result in non-unique and/or unstable solutions has been studied extensively [110, 59, 113, 43, 71, 134]. Specifically, authors have investigated the uniqueness of the inverse problem solution, when hydraulic soil properties, like k_s and α must be identified simultaneously. The studies were initially based on data from laboratory transient flow experiments aiming to characterize the relations between pressures, conductivity and water content. The parameter inference was initially based on cumulative outflow data, from different types of experimental data. In these studies, authors identified a non-uniqueness problem, which they attributed to the insensitivity of the objective function to the parameters. Attempts to resolve the problem involved using additional pressure head

or water content data. Eventually, the author of [43] argued that in some experimental setups, the use of outflow data did not improve the sensitivity of the objective function, and that a pore water pressure-based objective function was best. The authors of [71] argue that the uniqueness and the stability of the solution to the inverse problem in groundwater flow depends on many factors, like the number of available measurements, sensor positioning, availability of variation measurements both in pressure and in suction.

Authors who attempted to identify hydraulic soil properties using field monitoring sensor measurements have resorted to combining water pressure and discharge rate data in order to overcome the problem of non-uniqueness that they identified [22, 52, 132, 112]. In the present work the issue of ill-posedness is examined and a novel regularized objective function, exclusively based on pore water data, is proposed to improve the characteristics of the problem.

There are various algorithms designed for optimization -in this case, minimization. The two fundamental categories are gradient-based and gradient-free methods. The former requires the computation of the objective function derivatives, rendering its implementation an arduous procedure. These algorithms usually lead to fast convergence but are often 'trapped' in local minima and their outcome depends on the initial guess. Gradient-free methods, that will be used here, do not require differentiability of the objective function and therefore can be used to optimize objective functions that are discrete, discontinuous, or noisy. Among the most common algorithms of this kind are Genetic Algorithms, like Differential Evolution Genetic Algorithm (DEGA) [107] that is used in this work. When properly implemented, these algorithms perform a search of the entire parametric domain, thus providing better chances at finding the global minimum, providing one exists.

The major disadvantage of such algorithms is their high computational cost, since their strategy is based on repetitive objective function evaluations, and their slower convergence with respect to gradient-based solvers. A similar problem is faced when using bayesian techniques for parameter identification. When the objective function evaluation requires a Finite Element Method (FEM) computation, the cost may become prohibitive. It is therefore common practice to use surrogate models that can yield repetitive evaluations with a much lower computational cost. A common method in groundwater flow problems is the use of artificial neural networks (ANN), trained to approximate the relation between a permeability field and the pore water pressure distribution of the domain for a specific problem [35, 112, 132].

Here, a different approach is explored, employing a Reduced Order Model

(ROM) [94, 77, 62] instead of the computationally expensive FE model, to speed up the iterative calculation of pressure and deformation fields in the embankment, that is required in an optimization scheme. A ROM of a fully-coupled, hydro-mechanical (HM) model is used to describe the water flow through a partially saturated soil. The developed ROM will be shown to significantly enhance the computational efficiency of the method, while yielding accurate results, and leading to parametric predictions that are equivalent to those obtained using a FE model.

The novelty of the work that is presented in this Chapter lies in the use of ROM for back analysis of both mechanical and hydraulic soil parameters based on field pressure data from embankment dams. The method is validated with examples based on synthetic and real data from piezometers installed in the dam. The illustrative examples are based on the Glen Shira Lower Dam, which is part of a pumping storage scheme in Northern Scotland [91]. Another contribution of the work refers to the study of the effect of a regularized objective function to overcome the issue of non-uniqueness in the inverse groundwater flow problem.

To this end, the present Chapter is structured as follows. Section 4.2 contains the description of the Reduced Order Modeling methodology as well as an introduction to the employed back-analysis strategy via objective function minimization. In Section 4.3, the method is applied to a rapid drawdown problem using synthetically generated data for the inverse problem. In Section 4.4, the hydraulic properties of the Glen Shira lower dam are back-analyzed based on the real pressure head data that were observed on site during a rapid drawdown. Finally, in Section 4.5 the conclusions of the study are listed, and some comments are made on some remaining limitations of the methodology, outlining future works required to address them.

4.2 Methodology

Constitutive relations and governing equations

The governing and constitutive relations adopted in the developments of this Chapter are detailed in Chapter 2.

In this Chapter, the adopted relation between the hydraulic conductivity the soil-water system and the pore water pressure is the cubic law mentioned in Equation 2.18. This relation is is selected in accordance to [91].

A Reduced Order Model for coupled hydro-mechanical problems

In the context of parameter identification based on field sensor data, the unknown parameter must be identified by solving an inverse problem. In this case, if the forward problem is one where the unknown is the pore water pressure field, the inverse problem implies that some information about the pore pressure is known, and the unknown is now a set of one or more uncertain parameters in the model. In order to solve an inverse problem using a gradient-free optimization strategy, the parametric space must be explored by solving the forward problem for different parametric values. If this procedure is based on a full order approach (e.g. FE), the computational cost may be prohibitive. Hence, in this Chapter, a ROM will be employed to efficiently solve the inverse problem.

The Reduced Basis method is used to obtain a ROM for the coupled hydro-mechanical system of equations, as detailed in Section 2.3 of Chapter 2.

Inverse problem solving for parameter identification

Data/Sensor measurements

The problem is based on sensor observations of a measurable quantity related to the monitored structure or system. In this case, the available sensor data are measurements of water pressure obtained over a period of time on different parts of the body of the dam. If N sensors are installed and pressure measurements are obtained at N_t time instances, then the dataset at hand is written,

$$\begin{bmatrix} \bar{\mathbf{p}}_1^\intercal \\ \bar{\mathbf{p}}_2^\intercal \\ \vdots \\ \bar{\mathbf{p}}_N^\intercal \end{bmatrix} \in \mathbb{R}^{N \times N_t}, \tag{4.1}$$

where the overbar and the index indicate that each row corresponds to all the time history of pressure in one given sensor $i \in \mathbb{R}^N$ and reads, $\bar{\mathbf{p}}_i^\intercal = [\bar{p}_i^1, \cdots, \bar{p}_i^{N_t}] \in \mathbb{R}^{N_t}$. We point out for clarity that the previously mentioned $\mathbf{p} \in \mathbb{R}^{N_h^p}$ denotes the solution vector of nodal pressures, while $\bar{\mathbf{p}}_i \in \mathbb{R}^{N_t}$ denotes a vector of pressures in time.

Similarly, the pore water pressure as computed numerically by the adopted model at the point s(i) of the numerical domain Ω which corresponds to sensor

i, at N_t time instances reads,

$$\hat{\mathbf{p}}_{s(i)}^{\intercal} = [\hat{p}_{s(i)}^{1}, \cdots, \hat{p}_{s(i)}^{N_t}] \in \mathbb{R}^{N_t}. \tag{4.2}$$

Objective function

As mentioned in Section 4.1, in the past literature, authors who examined the hydraulic inverse problem, have resorted to objective functions that combine pressure head data and flow rate data to obtain a unique and stable solution of the inverse problem [52, 22, 132, 112]. Other authors have concluded that an objective function based on flow rate and pore water pressures does not provide better parameter sensitivity than one based on pore water pressure alone [43]. In this work, we argue that the simultaneous identification of parameters k_s and α using pore water pressure data alone is possible, and we propose a regularized version of the objective function, which takes into account the physical meaning of the parameters under consideration, and seems to reduce the effect of multiple local minima that exist in the non-regularized objective function.

The choice made in this work, to adopt an objective function that is based solely on water pressure data, is related to the fact that only pressure head data is available to the author of the present, for the case of a actual embankment dam (Section 4.4). Besides, pore pressure data is commonly available in the context of earth-fill dam monitoring.

Both parameters k_s and α are related to the conductivity of the 2-phase medium, that is, they are related to the "ease" with which a particular fluid, in this case water, is allowed to move through the pores of the particular soil at hand. In a transient problem, these parameters are related to the pace at which the phenomenon evolves, meaning, the pace at which a disturbance in the upstream boundary, such as a change in the upstream water level, reaches a sensor that is located at a certain distance downstream of the boundary. With that information in mind, a regularized objective function based on pore water pressure data is proposed, which reads,

$$F_{obj} = \sum_{i=1}^{N} \left(\sum_{j=1}^{N_t} \left(\bar{p}_i^j - \hat{p}_{s(i)}^j \right)^2 + w \sum_{j=1}^{N_t} \left(\dot{\bar{p}}_i^j - \dot{\hat{p}}_{s(i)}^j \right)^2 \right), \tag{4.3}$$

where $\dot{\bar{p}}_i^j$ and $\dot{\hat{p}}_{s(i)}^j$ are the time derivative of the measured pore water pressure at sensor i and time instance j, and the time derivative of the numerically computed pressure at point s(i) and time instance j respectively. The weighting

coefficient w accounts for the differences in magnitude between the two terms. The time derivative of the measured water pressure at sampling time j is approximated as,

$$\dot{\bar{p}}_j = \frac{\bar{p}_i^j - \bar{p}_i^{j-1}}{\Delta t}, \tag{4.4}$$

where $\Delta t = t_j - t_{j-1}$, is the time increment between two subsequent sampling times. The time derivative of the numerically computed pressure is similarly evaluated.

The minimization strategy

The minimization algorithm used in the present work is the Differential Evolution Genetic Algorithm (DEGA), and particularly the SciPy Python package implementation of the algorithm [107]. Differential Evolution is a global search technique that does not require a derivative evaluation of the error function. The optimized function therefore does not need to be differentiable and it can be evaluated in a black-box procedure, in this case a FEM or ROM. However, this method often requires a large number of objective function evaluations, and as a result, in this case, a large number of model queries.

DEGA is a population-based evolutionary computation, which iteratively seeks to optimize a problem by maintaining a population of candidate solutions, while creating new candidate solutions by combining existing ones and keeping whichever solution has the best fitness score. The best fitness score is in this case expressed as the lowest objective function value. The new candidate solutions, similar to other evolutionary algorithms, are created by mutation and crossover. The algorithm parameters, like the population size, the mutation and crossover constant must be adjusted to the problem at hand, in order to ensure that the search covers the entire parametric domain, is not trapped in local minima and converges as fast as possible. DEGA, similarly to other global search algorithms, does not guarantee convergence towards the global minimum, but can generate good quality solutions when implemented properly. DEGA has been used for back-analysis in geomechanical problems ans is proven to be an appropriate method for this type of problem [117, 131].

4.3 Inverse problem based on synthetic sensor data for a rapid drawdown case

Setup of the problem: Rapid drawdown in the Glen Shira dam

To validate the parameter identification method described in the previous sections, an application that is based on an existing water dam is examined. The inverse problem is solved first using synthetic data, and then using real pore pressure data from piezometers that were installed in the actual dam. Both applications, with the synthetic data and the real data, are related to a rapid drawdown problem. The case has been extensively studied in [91], in which the authors compare the performance of 4 different models in the description of the piezometric measurements. The models assume physical hypotheses of increasing complexity. A pure flow uncoupled model, a coupled elastic model considering instantaneous drawdown of maximum intensity, followed by pore water pressure dissipation, a coupled elastic model with gradual drawdown, and a coupled non-elastic analysis considering a gradual drawdown were all considered. The term gradual drawdown here indicates that the change in water level is applied as a time-dependent hydraulic boundary condition in the upstream slope. The authors found that the data are best described using a coupled hydro-mechanical analysis and considering a gradual drop of the upstream water level. The approach that considers an elastic constitutive model captures well the observed pore water pressures, since it seems that no plastification occurred during the rapid drawdown.

The Glen Shira Lower Dam is part of a pumping storage scheme in Northern Scotland. The maximum cross section of the dam and the position of the sensors that were installed in the dam and recorded the pore pressure measurements used in Section 4.4 are presented in Figure 4.1. The embankment is made of a compacted moraine soil and is reinforced with a rockfill shell that covers the upstream slope. The mechanical and hydraulic parameters used in this work are the same as the ones in [91] and are listed in Table 4.1. Other details regarding the construction of the dam and the selection of material properties are mentioned in [91].

The total water level drawdown that was applied to the Glen Shira Dam over 4 days was 9.1 m. It was imposed over 4 stages of rapid water lowering, followed by short stages of constant upstream water level. The changing water level is shown in Figures 4.3a, 4.3b. The initial condition is obtained by instantaneously filling up the reservoir up to the initial upstream water

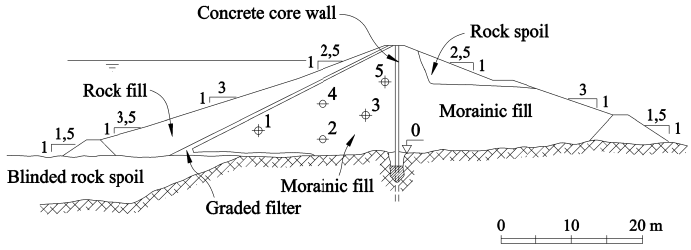


Figure 4.1: Maximum cross-section of Shira Dam. The positions of piezometers 1 to 5 are indicated. Re-sketched after [91]

level, keeping it stable, and running the transient simulation until right before steady state conditions have been reached and the computed pore pressures in the positions of the 5 sensors are as close as possible to the initial registered sensor measurements.

In this work, only the upstream half of the domain has been modeled, since the concrete core wall that is shown in Figure 4.1 is practically impervious and very stiff, such that the displacement and pressure field in the downstream part can be neglected in the context of the inverse problem. The graded filter is neglected, as it does not influence the pressure-displacement states of the dam. Instead, the parts of the mesh that correspond to the graded filter are assigned the same rockfill material that is considered in the upstream shell. This assumption is also adopted in [91]. The mesh was created in Gmsh open-source mesh generating software [38] and is shown in Figure 4.2. Two subdomains are considered, one representing the rockfill shell and one the compacted Morainic fill.

The displacement and the water flow are set to zero at the bottom, and at the right-side boundary, where the concrete wall is assumed. The material bellow the dam is considered impermeable. The upstream water is modeled as a Dirichlet pressure boundary condition applied on a time dependent part of the upstream boundary. In the mechanical equilibrium equation a Neumann condition is introduced accounting for the weight of the water. It is likewise applied on a time dependent part of the upstream boundary domain. Finally, a seepage Robin-type condition (Equation (2.13)) is applied in the upstream boundary.

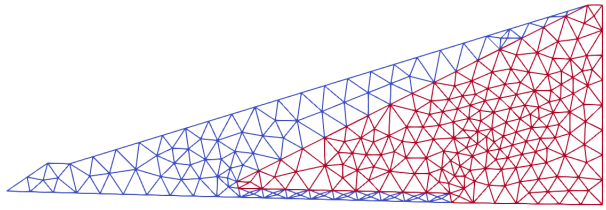


Figure 4.2: Finite Element mesh of the upstream part of the 2D cross section of Shira Dam. The blue part of the domain is assigned rockfill material and the red part, Morainic fill material.

Problem parametrization

The saturated hydraulic conductivity k_s, parameter α, which is related to the air entry suction of the soil, and the Young's modulus E are thus chosen for identification. The hydraulic soil properties k_s and α appear in Equations (2.14) and (2.18). Understanding the hydraulic properties of an unsaturated soil is an essential requirement for the prediction of pore pressure in a transient groundwater problem. The use of optimization techniques for the identification of hydraulic properties is common in the literature, particularly in the context of agricultural engineering and mainly with respect to experimental data [110, 134, 126, 59, 113, 28, 55, 43, 96]. The choice of k_s as a parameter to be identified is rather straightforward, as the uncertainty on hydraulic conductivity in geotechnical problems and its dependence on the compaction state of the material is well documented [30, 122, 75].

Often in geotechnics the assumption is made that the water retention curve is unique for a soil, though it has been shown that it depends on the mechanical state of the soil [80] [79]. The differences are mostly reflected on the parameter α that is related to the air entry suction of the soil. This parameter has been shown to feature variability that may stem from physical processes that are not explicitly described in the model [80, 79, 32, 73]. It is therefore expected to feature high uncertainty in the context of an earthfill dam, depending on the initial soil compaction during the construction phase, as well as the loading, and wetting/drying history during the functioning phase.

Finally, regarding the selection of parameter E, the notion that the mechanical coupling affects the rate of pressure diffusion in a rapid drawdown problem, that has been elaborated in [91], motivated the exploration of inverse

Property	Symbol	Units	Morraine soil	Rockfill
Young's Modulus	E	MPa	$[30, 300]$	100
Posson ratio	ν	-	0.3	0.3
Particle density	ρ_s	kg/m^3	2.7×10^3	2.7×10^3
Saturated VWC	Θ_s	-	0.282	0.4
Residual VWC	Θ_r	-	0	0
Parameter	α	m^{-1}	$[0.01, 1]$	1
Fitting Parameter	m	-	0.2	0.4
Saturated hydraulic conductivity	k_s	m/s	$[10^{-9}, 10^{-7}]$	10^{-4}

Table 4.1: Mechanical and hydraulic parameters used for the analysis of the Shira Dam. For properties treated as parameters in the inverse analysis, ranges considered are given in [ ]

identification of a mechanical property based on pressure sensor data. Besides, E is a parameter worth exploring in the context of earthfill dams. Even though its influence in a seepage problem such as the one treated, is expected to be small, it is shown that in certain problems in geomechanics, significant uncertainties may stem from the mechanical properties [129, 95, 84].

The ranges in which parameters k_s, α and E take values are selected such that they represent realistic orders of magnitude for the respective properties for a relatively impervious soil material used in the body of an earthfill dam. The values resulted form a review of various references studying earthfill dams [36, 91, 97, 27, 121, 131].

Synthetic data generation

Studying the synthetic problem allows the assessment of the minimization strategy, as well as the minimized objective function, without the complexity introduced by uncertainties that exist in a real-life problem. In geotechnical problems uncertainty may be related to various sources: inherent variability of the soil mass due to unpredictable geomechanical processes that continuously

modify the soil in situ [84]; measurement errors; insufficient site investigation; poor capacity of the selected physical models to describe real processes; and transformation uncertainty, often introduced when laboratory measurements are transformed to the sought-after parameter using empirical models [89].

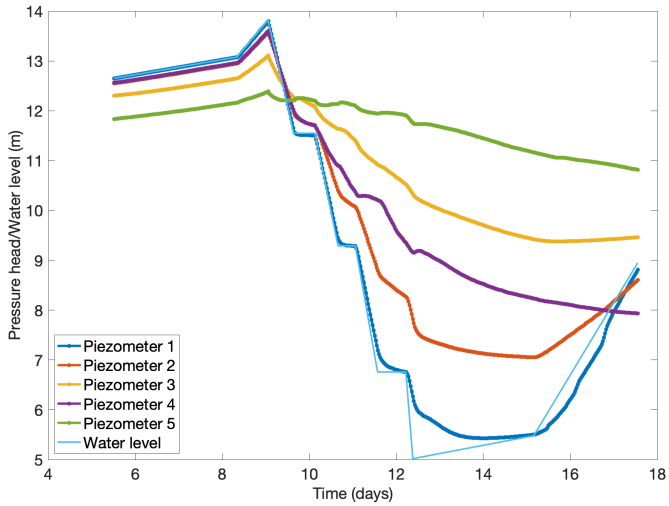
In this section, an application based on synthetic data is first developed. The goals of this section are to show that the ROM is an appropriate surrogate for FEM for inverse problem solving (Section 4.3), and that the proposed objective function is a good choice for tackling the problem (Section 4.3). The pore pressure data is computed with the ROM that is described in Section 2.3, or using the full order FE model, without any numerical noise introduced to the data. In Figure 4.3a the synthetically computed pressure heads in the 5 virtual sensors are shown, as well as the upstream water level as it fluctuates during the water drawdown event. The data that is plotted, is computed with the ROM described in Section 2.3. However, in some cases in the following, in order to examine the ROM's accuracy, the objective function evaluation is performed with ROM, but the synthetic sensor data are computed with the FE model. These are not shown here, but since the error between the results of the two models is very small, as will be shown in Section 4.3, the difference between that data is not clearly visible in a plot. The data is computed using the parametric values mentioned in Table 4.1, and $k_s = 1.6 \times 10^{-8}$m/s, and $\alpha = 0.2\text{m}^{-1}$, which are the values used in the coupled analysis run for the Shira dam in [91].

In Figure 4.3b the water pressure head data as registered in the actual sensors of the Shira dam are shown. This data will be used in Section 4.4.

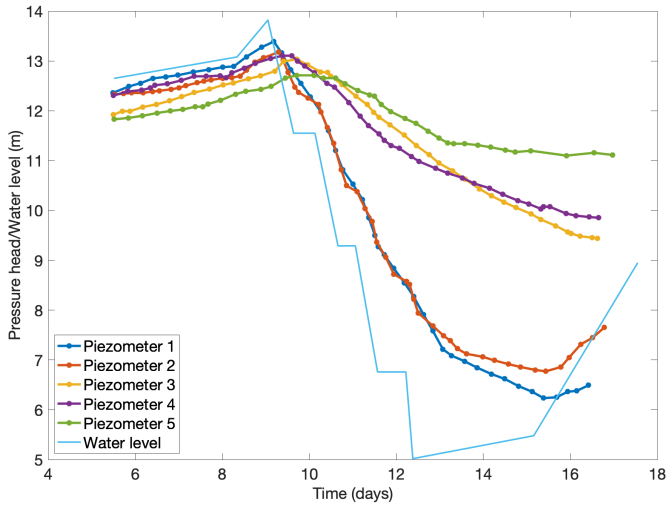
Implementation

The mesh that is shown in Figure 4.2 yields a problem with 286 degrees of freedom for the description of the pressure field and 2160 for the description of the displacement field. As explained in Section 2.3, setting up a ROM that seeks an approximate solution of a system of equations, requires a pre-process stage, or an offline stage, where snapshots that correspond to various parametric values in the feasible space, are collected and processed. In this case the transient pressure and displacement fields were evaluated for 125 sets of k_s, α and E parameters, that correspond to the combinations of 5 k_s values, equidistantly spaced in a 10-logarithmic scale, 5 similarly selected values of α and 5 for E.

As explained in Section 2.3, SVD is applied to the snapshot matrices in



(a)



(b)

Figure 4.3: Water level during the drawdown event, and sensor data used in the parameter identification problems of this work. 4.3a. Synthetic data computed by a ROM, during a rapid water drawdown event, on points that correspond to the 5 Piezometers shown in Figure 4.1. 4.3b. Real data registered in Shira Dam during a rapid water drawdown event, in 5 the Piezometers, and upstream water level during the event. The y-axis represents pressure head for the 5 curves labeled as Piezometer 1,2,3,4,5 and water level for the last curve. The reference level is specified in Figure 4.1.

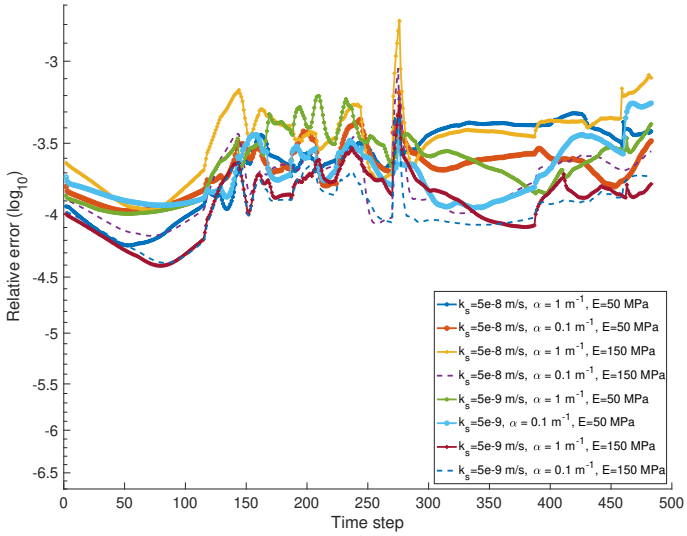
order to obtain orthonormalized basis vectors for the space that is spanned by the snapshots and their corresponding singular values σ_i. The few vectors that correspond to the largest singular values are kept, according to,

$$\log_{10}(\frac{\sigma_i}{\sigma_1}) > \text{tol}, \tag{4.5}$$

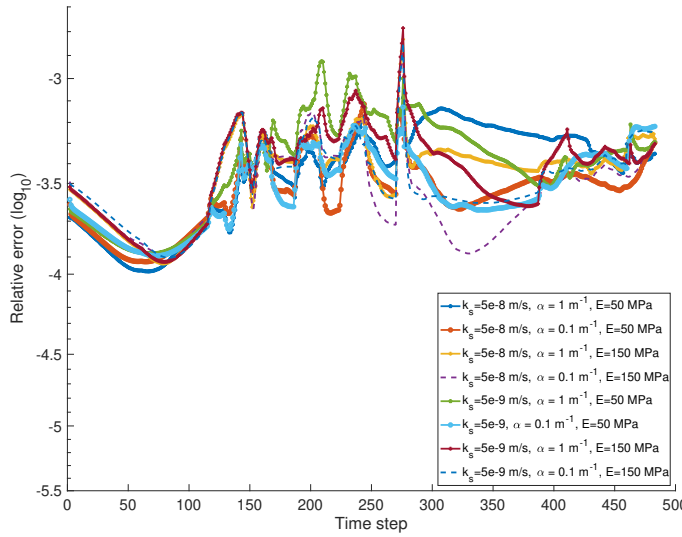
where σ_1 is the largest singular value and tol is a tolerance decided based on the level of accuracy that is required in the investigated problem. In this case, tol $= -4$ was selected, since a relative error between the approximate solution of ROM and the high-fidelity FEM solution smaller than 10^{-4} is not necessary when dealing with quantities such as water pressure and displacement. This process lead to two Reduced Bases of 40 and 59 basis vectors for displacement and water pressure respectively. To examine the accuracy of the created ROM, some simulations were conducted for parametric values that were not sampled in the offline stage (i.e. that were not in the snapshots). The relative error in the displacement and pressure fields, as seen in Figure 4.4, is in the order of magnitude of 10^{-3}, which is considered an acceptable level of accuracy for this application. The error is estimated as in Equation (3.19).

It appears that an offline stage with only 125 problem solutions yields a Reduced Basis in which an adequately accurate approximation can be sought for, for any value in the entire multidimensional parametric domain. This cost is very small compared to the multiple hundreds, or even thousands of objective function evaluations - where each evaluation translates into a transient problem solution - that can be required for parameter identification with back analysis.

Many simulations of this problem have been conducted, in the context of the validation of the ROM accuracy, inverse problem solving, and visualization of the objective function for different parametric values. Overall, the ROM is 3-4 times faster than the FE model. This is a significant efficiency boost even though the problem at hand is 2D and remains rather small since the FEM discretization used yields a rather small number of unknowns even for the full-order case. In engineering practice, such small problems are not common. As shown by the authors in [77], the efficiency gains of a ROM for the hydro-mechanical problem become even more prominent for problems with more unknowns of the full-order problem.



(a)



(b)

Figure 4.4: 4.4a Relative error ROM-FEM in the displacement field for various parameter values. 4.4b Relative error ROM-FEM in the pressure field for various parameter values

Study of the objective function and its sensitivity to the parameters

In order to study the performance of the regularized objective function some 3D surface plots of the function in the parametric search space are now shown. In the following, the two terms of the objective function (4.3) are referenced as terms A and B, and their contributions to the objective function are plotted separately as well as combined in order to visualize the effect of the regularization. Both terms therefore read,

$$A = \sum_{i=1}^{N} \sum_{j=1}^{N_t} \left(\bar{p}_i^j - \hat{p}_{s(i)}^j \right)^2, \tag{4.6}$$

$$B = w \sum_{i=1}^{N} \sum_{j=1}^{N_t} \left(\dot{\bar{p}}_i^j - \dot{\hat{p}}_{s(i)}^j \right)^2, \tag{4.7}$$

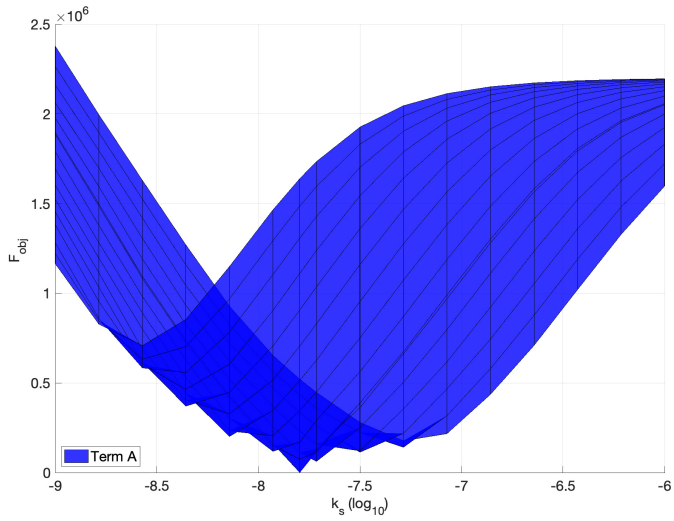
$$F_{obj} = A + B. \tag{4.8}$$

It is difficult to visualize the objective function with respect to all three parameters under consideration at once. In order to study the sensitivity of the objective function two problems were studied separately -the simultaneous variation of the two hydraulic parameters k_s and α while E is kept fixed, and the variation of k_s and E while α is kept fixed. In this Section, the weighting factor is set to $w = 50$.

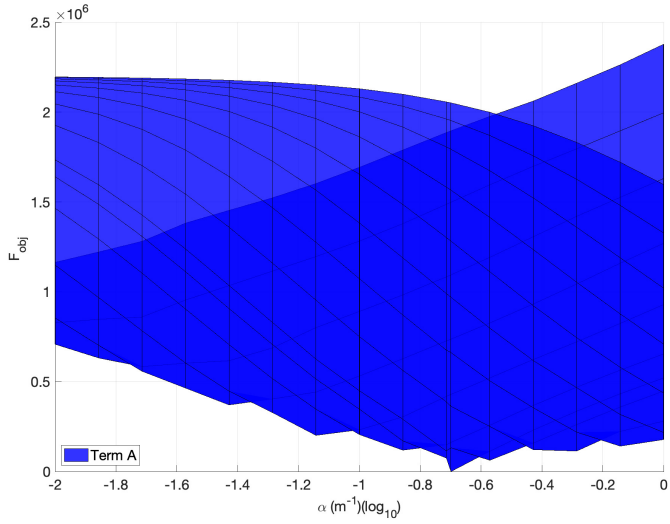
Parameters k_s and α

In the following, parameter E is fixed to 100 MPa which is the value that was used in [91]. Parameters k_s and α take values within the ranges mentioned in Table 4.1.

In Figure 4.5, the shape of an exclusively pressure-based objective function can be visualized. It is clear that even though this problem has a unique solution that minimizes the squared differences between observed and calculated data, there are many local minima in the function. These minima do not appear in Figure 4.6, where the contribution of the B term has been plotted. In Figure 4.7, where the regularized objective function is plotted, the local minima are noticeable, though they are much weaker than in Figure 4.5. In Figure 4.8, both contributions and their sum are plotted in the same graph for comparison. The surface representing the objective function, that is, the summation of both

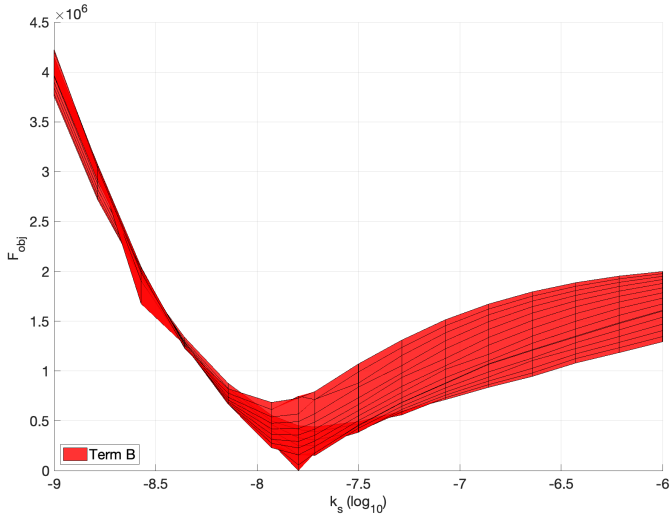


(a) View: Term A vs. k_s

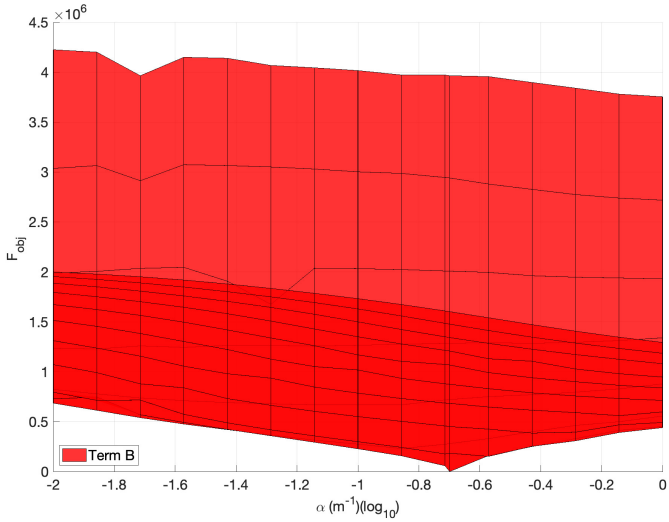


(b) View: Term A vs. α

Figure 4.5: Response surface of term A for different values of k_s and α



(a) View: Term B vs. k_s



(b) View: Term B vs. α

Figure 4.6: Response surface of term B for different values of k_s and α

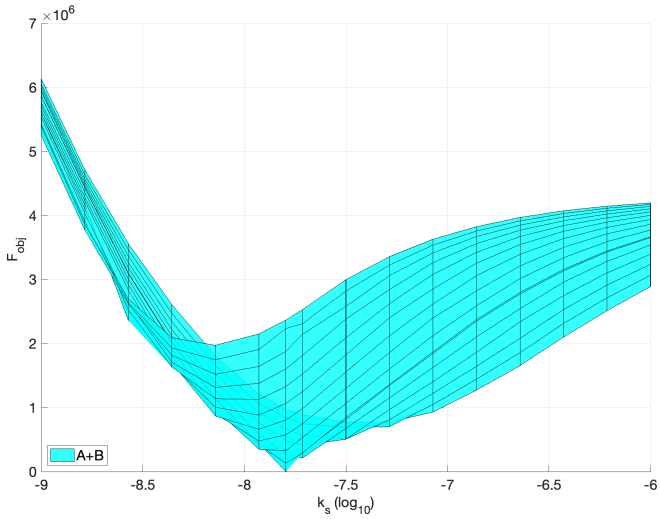
terms A and B, displays, as expected a sharper drop around the minimum value compared to the function accounting only for pressure differences.

These surface plots indicate that both terms A, B and their summation F_{obj}, are more sensitive to variations in k_s and less so to variations in α. This can be deduced by observing the steep slope of the function around the optimum k_s and the flat "valley" around the true value of α. Another observation relates to the fact that the objective function seems to have a steeper increase toward the smaller conductivity values than in the direction of larger k_s values. Those results are in accordance with [59].

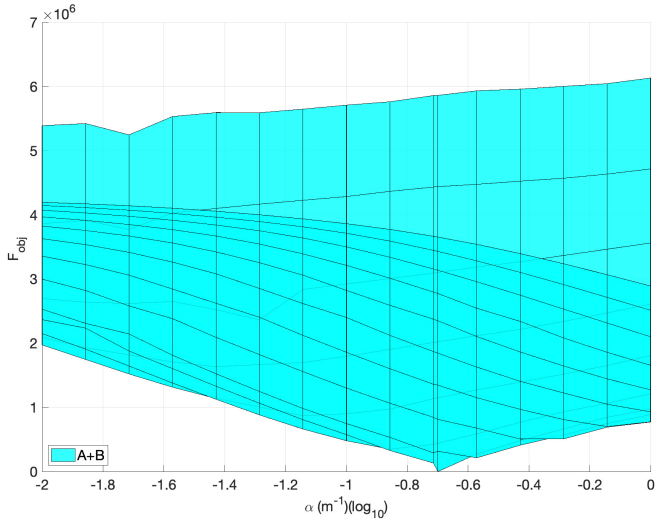
The existence of local optima in an objective function can lead to a convergence to a false solution, or delay the convergence to the global minimum, even if a global search algorithm is used, such as a evolutionary algorithm. By adjusting the weighting factor w, the influence of the regularizing term can be tuned, depending on the chosen optimization strategy. For example, when using an evolutionary algorithm, such as the widely used Genetic Algorithms and the DEGA which is used in this work, an objective function with less local minima, allows the use of a smaller mutation rate, which can speed up the convergence process. In this work, the factor w was selected such that the values of the two terms of the objective function are of the same order of magnitude. This value differs depending on the specific problem under scrutiny, the sensing frequency, and the number of sensors used.

In order to validate the adequacy of the ROM in terms of accuracy when it is used in an inverse problem some response surface plots are shown, that illustrate the magnitude of the error introduced by the ROM with respect to an FE model. This is critical in the problem that is under investigation, in which the evaluation of an objective function formulated as in (4.8) implies the computation of a norm of differences over time, and thus, the time integration of the pressure solutions that feature some error with respect to FEM. In Figure 4.9 the response surfaces of the objective function (4.8) for the cases where both the model response and the data were computed with ROM, the model response and the data were computed with a FE model, and the data was generated with a FE model while the model response was computed using a ROM.

The three surfaces are practically identical, with discrepancies that are smaller than the order of magnitude of the objective function values. This is an indication that the accuracy obtained with a ROM is adequate for a parameter



(a) View: F_{obj} vs. k_s



(b) View: F_{obj} vs. α

Figure 4.7: Response surface of the regularized objective function F_{obj} for different values of k_s and α

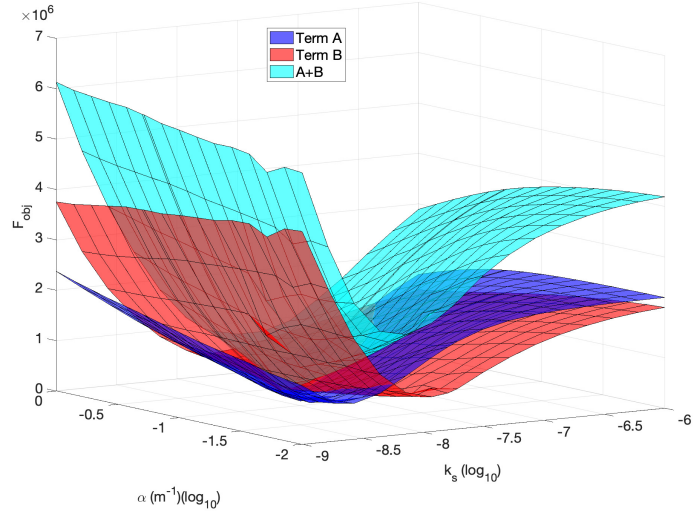


Figure 4.8: Values of the contributions A and B and the total F_{obj} for different values of the parameters k_s and α

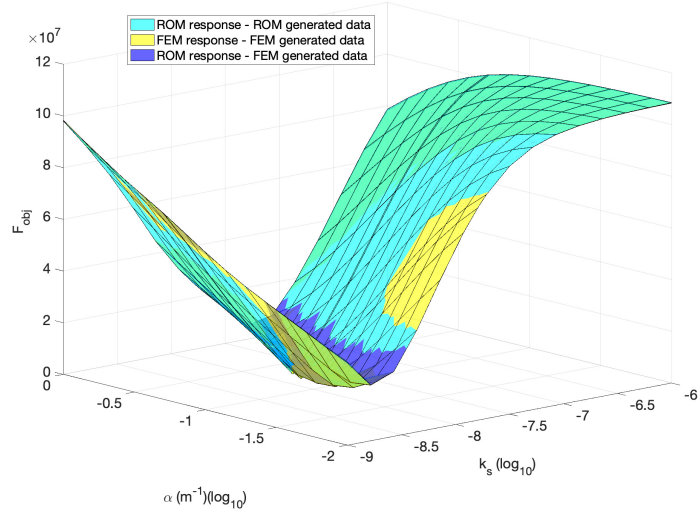


Figure 4.9: Response surface of the objective function (4.8) for synthetic data generated with ROM or FEM and model responses computed with ROM or FEM

identification problem.

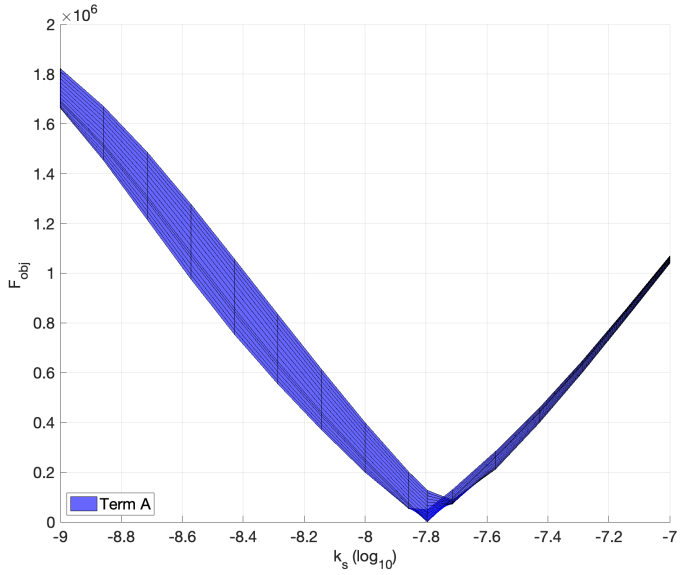
Parameters k_s and E

In the following, parameter α is fixed to 0.2 m^{-1}, which is the value that was used in [91]. Parameters k_s and E take values within the ranges mentioned in Table 4.1.

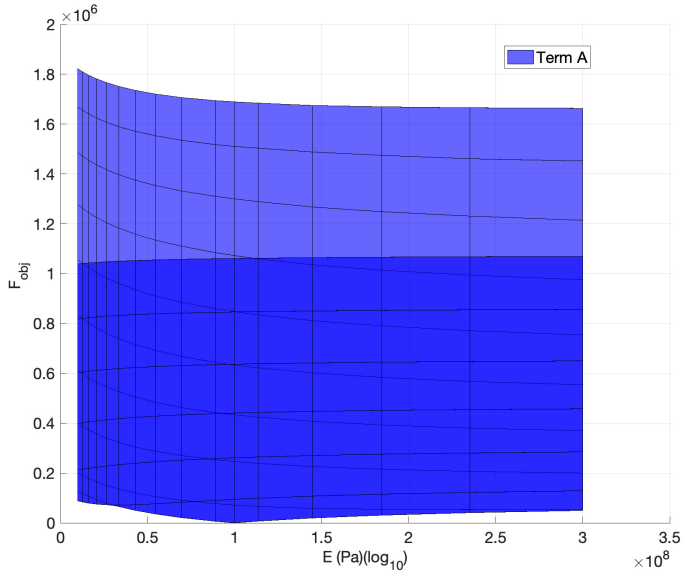
It is not uncommon in geotechnics to consider the mechanical and the hydraulic problems separately, and ignore the coupling effect, though the availability of commercial numerical tools that consider the HM coupling tends to eliminate this practice in recent years. In the case of rapid drawdown, using an uncoupled pure flow analysis is equivalent to the assumption of an incompressible soil skeleton. As argued in [91], this is not a realistic assumption for the Shira Dam. Changes in the stress state of the dam, due to the mechanical un-loading of the upstream boundary as the water level drops, are expected to affect the consolidation process, that is, the rate at which the pore pressure diffuses. Hence, the soil stiffness is expected to have an effect on the pressure drop recorded by piezometers 1-5, or, in other words, the inference of the Young's Modulus E, based on the pressure sensor data, may be possible, provided that the soil is not stiff enough to be considered incompressible.

To test this effect, the response surface of the objective function for varying parameters k_s and E is visualized in Figures 4.10 - 4.12. For this study, the synthetic data was computed using the ROM described in Section 2.3 for values $k_s = 1.6 \times 10^{-8}$ m/s and E = 100 MPa.

The sensitivity of the objective function is once again significantly larger for k_s than for the other parameter E. The pressure-based term of the function is rather flat around the target value (Figure 4.10). The regularizing pressure derivative term seems to be steeper around the target value (Figure 4.11) which implies higher sensitivity. The contribution of Term B leads to a steeper F_{obj} (Figure 4.12b) and this effect can be magnified by adjusting the weighting factor in Equation (4.8), so as to enhance the influence of Term B. Overall, this problem does not seem to suffer from multiple local minima.

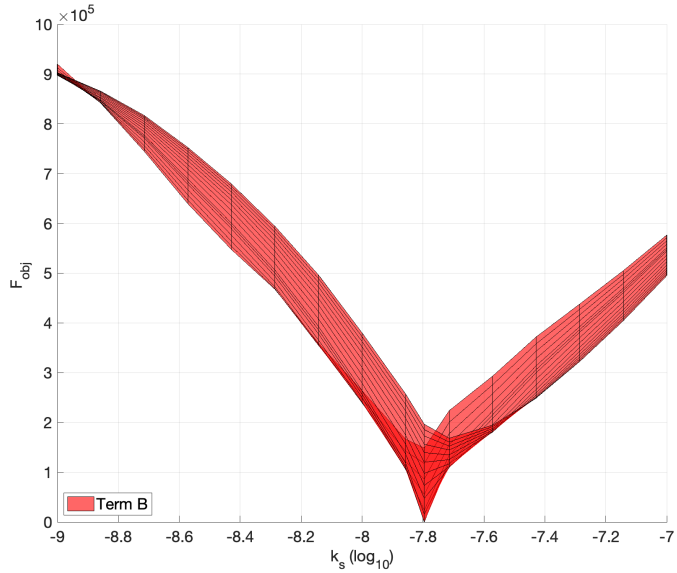


(a) View: Term A vs. k_s

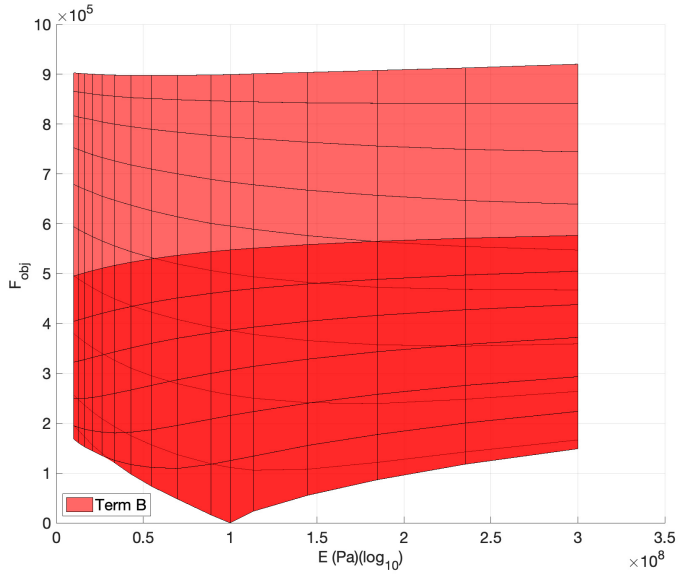


(b) View: Term A vs. α

Figure 4.10: Response surface of term A for different values of k_s and E

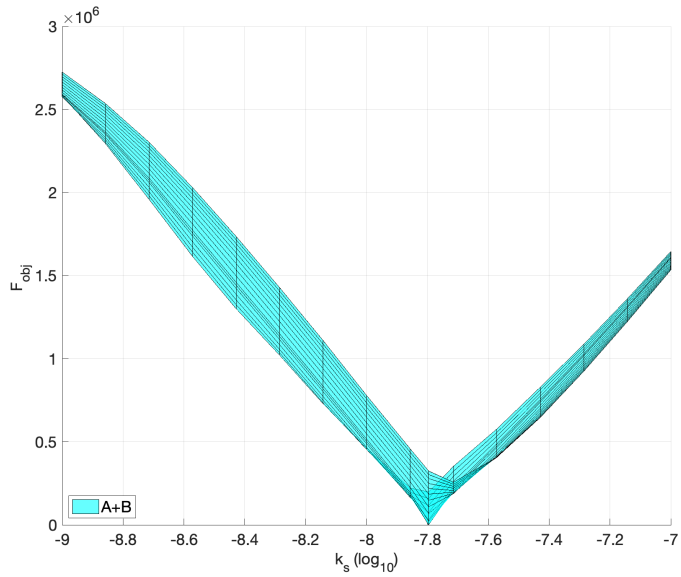


(a) View: Term B vs. k_s

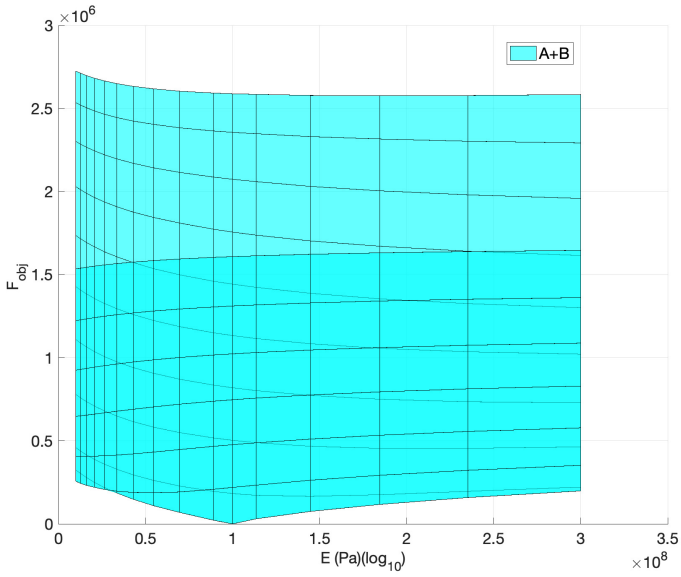


(b) View: Term B vs. E

Figure 4.11: Response surface of term B for different values of k_s and E



(a) View: F_{obj} vs. k_s



(b) View: F_{obj} vs. E

Figure 4.12: Response surface of the regularized objective function F_{obj} for different values of k_s and E

	k_s (m/s)	α (m^{-1})	E (MPa)	Minimization time (days)	Number of evaluations	Final F_{obj}
Inv. k_s, α	1.6×10^{-8}	0.2	100	1.33	1722	0.10
Inv. k_s, E	1.6×10^{-8}	0.2	100	1.22	1593	0.043
Inv. k_s, α, E	1.6×10^{-8}	0.2	100	2.84	2572	0.49

Table 4.2: Identified parametric values based on synthetic sensor data for Shira Dam generated by the full-order FE model. The model response was estimated using ROM

Parameters k_s and α and E

Finally, DEGA was used to infer the target values $k_s = 1.6 \times 10^{-8}$ m/s, $\alpha - 0.2\,m^{-1}$ and E = 100 MPa. For the cases in which two of the three parameters are sought for, the third is fixed to the corresponding target value. The model response is calculated with ROM and the synthetic data is created with FEM. In Table 4.2, the results of the objective function minimization procedure are presented. In the last column the final value of the minimized objective function is given when the process is stopped.

The minimization procedure was stopped after 50 generations. In all 3 cases, the target values were obtained. The results do not correspond to the precise parametric values that were used for the generation of the data, but rounding off the resulting values $\log10(k_s)$, $\log10(\alpha)$, E to the second decimal place, the values presented in Table 4.2 are obtained, that are virtually identical to the target values.

4.4 Inverse problem based on actual sensor data

In Section 4.3 the inverse identification problem was examined without considering the effect of measurement noise and ignoring the modeling error, that is, the error that is attributed to physical processes that affect the real-life problem, but are not considered in the physical model used. In this Section, a problem based on real data is solved to verify the feasibility of

soil hydraulic and mechanical characterization by the regularized objective function optimization based on pressure sensor data.

Sensor measurements from drawdown in the Glen Shira dam

Five porous stone piezometer disks, previously calibrated against mercury columns, were located in the places shown in Figure 4.1. The authors conclude in their paper [91] that the possibility of instrumental error is of minor order and can be neglected. The observations are shown in Figure 4.3b.

Results

In this Section the inverse problem of simultaneous identification of parameters k_s, α and E, and the problems of identification of parameter pairs k_s-α, and k_s-E that were studied in Section 4.3 are solved, and the results are listed and commented. The inversions are based on the actual sensor pressure data that were recorded in Shira Dam during a drawdown incident.

The problem is solved multiple times to verify the capacity of the method to provide a stable solution. The procedure is repeated 3 times for each combination of parameters, using the regularized objective function F_{obj} = A + B that accounts for squared differences between computed and evaluated pressures and pressure derivatives, using a weighting factor $w = 1$. For all inversions, the ROM is used to evaluate the model response.

The outcomes of all inversions are reported in Tables 4.3 - 4.5. The number of objective function evaluations and the total time required for the optimization algorithm to converge are also reported. Finally, in the last column of Tables 4.3 - 4.5, the final value of the minimized objective function is listed. The final value of F_{obj} gives an insight on the quality of the achieved fit, as a smaller final objective function value indicates a better fit.

For the inversions listed in Table 4.4, the Young's modulus E is fixed to the value of 100 MPa. For the inversions listed in Table 4.5, the Young's modulus α is fixed to the value of 0.2 m^{-1} .

The results seem to indicate that the three real data-based problems have unique solutions. The optimization method does not converge to identical values all 3 times that a problem is solved, but provides sets of realistic solutions that adequately characterize the soil, and feature small discrepancies among them. Moreover, the final fit obtained by the 4 repetitions of each

	k_s (m/s)	α (m^{-1})	E (MPa)	Minimization time (days)	Number of evaluations	Final F_{obj}
Inv. with ROM	2.01×10^{-8}	0.294	33.30	0.89	971	39.19
Inv. with ROM	1.99×10^{-8}	0.313	38.22	0.78	891	39.22
Inv. with ROM	2.00×10^{-8}	0.315	30.71	0.97	1150	39.22
Inv. with FEM	1.99×10^{-8}	0.245	31.60	1.74	1002	39.19

Table 4.3: Identified parametric values for all 3 parameters k_s, α and E based on actual sensor data recorded during a drawdown in Shira Dam. Results are obtained implementing the inversion methodology 4 times; 3 times using ROM to compute the model response and 1 time using FEM.

	k_s (m/s)	α (m^{-1})	Minimization time (days)	Number of evaluations	Final F_{obj}
Inv. with ROM	6.95×10^{-9}	0.083	0.37	366	40.10
Inv. with ROM	6.79×10^{-9}	0.081	0.40	396	40.09
Inv. with ROM	6.69×10^{-9}	0.078	0.23	243	40.10
Inversion with FEM	2.88×10^{-9}	0.033	1.47	390	42.72

Table 4.4: Identified parametric values for parameters k_s and α based on actual sensor data recorded during a drawdown in Shira Dam. Results are obtained implementing the inversion methodology 4 times; 3 times using ROM to compute the model response and 1 time using FEM.

problem, expressed by the value of the minimized objective function, is stable for each problem.

Among the 3 different problems, the inferred parameters feature notable discrepancies but small differences in terms of how well they fit the data. The obtained values are also different to the ones that are used in [91],

	k_s (m/s)	E (MPa)	Minimization time (days)	Number of evaluations	Final F_{obj}
Inv. with ROM	1.70×10^{-8}	38.27	0.56	528	39.37
Inv. with ROM	1.57×10^{-8}	39.91	0.67	636	39.33
Inv. with ROM	1.57×10^{-8}	41.38	0.40	378	39.33
Inv. with FEM	1.57×10^{-8}	43.28	0.55	354	39.37

Table 4.5: Identified parametric values for parameters k_s and E based on actual sensor data recorded during a drawdown in Shira Dam. Results are obtained implementing the inversion methodology 4 times; 3 times using ROM to compute the model response and 1 time using FEM.

$k_s = 1.6 * 10^{-8}$ m/s, $\alpha = 0.2$ m^{-1} and E = 100 MPa. It is worth noting here that the values used in [91] were determined based on a combination of information given by [86] and the use of empirical methods for characterizing the water-retention curve of a soil. Regarding k_s, the value used in [91] is based on values that were given in [86], and were obtained by laboratory experiments on compacted specimens. Two values were mentioned by [86], 1.6×10^{-8} m/s when the specimen is compacted at optimum water content and 1.6×10^{-7} m/s when compacted wet of optimum. The first value was selected for the numerical analysis in [91], as the dry density reached in the field (19.8 kN/m^3) is higher than the optimum laboratory compaction (19.3 kN/m^3), which leads to a reduction in permeability.

The evaluation of parameter α was based on a simplified empirical procedure that makes use of the grain size distribution, that were published in [86], to derive a water retention curve for the Moraine soil. Parameters α and m are then inferred by fitting the van Genuchten expression (2.14) to the derived water retention curve. Finally the Young's modulus E value, was selected by the authors of [91] as a realistic value for this type of soil. It is therefore not unjustified to expect some uncertainty in the selected values, that was, however, not relevant to the scope of [91].

The pressure heads as computed using the ROM and the identified parametric values are plotted in Figures 4.13a-4.11e. The solutions that correspond

	k_s (m/s)	α (m^{-1})	Minimization time (days)	Number of evaluations	Final F_{obj}
Inv. with FEM	4.89×10^{-10}	0.010	0.51	705	40.48

Table 4.6: Identified parametric values for hydraulic parameters based on actual sensor data recorded during a drawdown in Shira Dam. Results are obtained using a pure hydraulic analysis and an FE model to compute the model response.

to each problem are plotted with a different color. The observed sensor pressure heads are also shown in the plots. It seems that the pressure head values predicted at each sensor are similar for any set of solutions among the ones obtained by the inversion.

It seems that in the case where all 3 parameters are solved for, a better fit is obtained, that is, the final value of the objective function is smaller for these inversions. This indicates that even in a problem like rapid water level drawdown, where the changes are mostly expected to affect the pressure field, the value of the mechanical parameter E is reflected on the water pressure measurements. Even in problems that are often treated as purely hydraulic, the effect of the mechanical coupling in soil should not be neglected.This idea is also was argued in [91], where the coupled analysis is shown to better describe the pressure head measurements, even though a rather stiff soil is assumed.

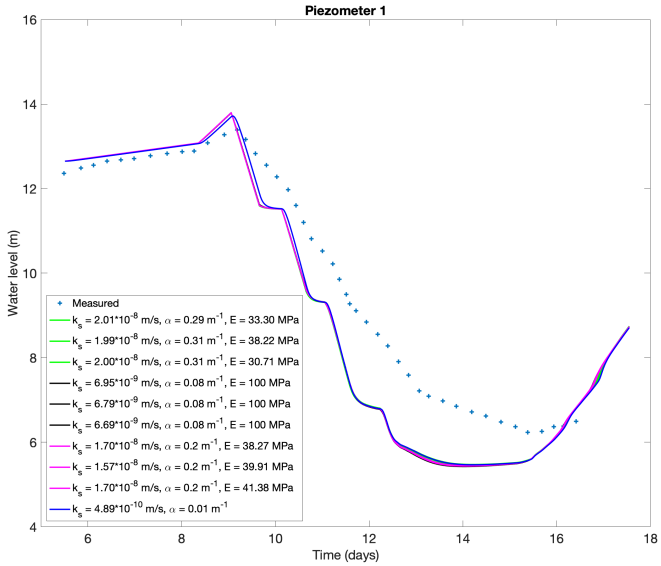
To explore this issue further, the hydraulic parameters k_s and α were identified using pure hydraulic flow analysis. The motivation for this study is to determine whether a good fit and realistic parametric values can be obtained with a mere hydraulic analysis. The search was performed initially in the parametric domains that were used before, that is, $k_s \in [10^{-9}, 10^{-7}]$ m/s, $\alpha \in [0.01, 1]\, m^{-1}$. With this search range, the obtained value for k_s was 10^{-9}, so the smallest acceptable value. Eventually the search range for k_s was extended to $[10^{-10}, 10^{-7}]$ m/s and the minimization was run again. The result is given in Table 4.6 and shown in Figures 4.13a-4.11e. It turns out that the resulting fit is similar to the results of previous problems using coupled analysis. However the obtained parameters are orders of magnitude apart from the ones reported

by [86] and [91]. Particularly, regarding k_s, the obtained value by hydraulic inverse modeling is more than one order of magnitude smaller than the smallest hydraulic conductivity value measured experimentally.

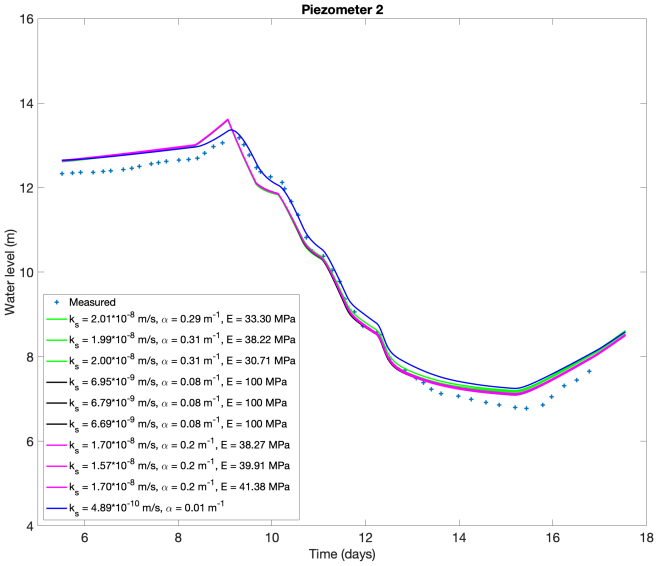
Despite the fact that the method was shown to be effective in a synthetic problem in Section 4.3, the results obtained for the real-data problem, even though realistic in terms of soil characterization, lead to pressure head solutions that are different to the pressure head sensor observations to a notable degree. There is a discrepancy between the computed and the measured data that does not seem to decrease significantly for any combination of parameters, especially for sensors 1 and 4.

This discrepancy may be related to uncertainties in other parametric values, that are not being identified, or it could be related to the insufficiency of the model in describing all relevant physical phenomena to the problem. For example, in this model, single values are given to the properties of the Moraine soil and the rockfill material. Soil is a highly heterogeneous material, and neglecting this heterogeneity could lead to false results. Moreover, the elastic model employed for the constitutive modeling of soil may be unrealistic. As explained in [71], the exiestence of a unique and stable solution to an inverse problem of soil characterization does not guarantee the accuracy of the obtained parametric values.

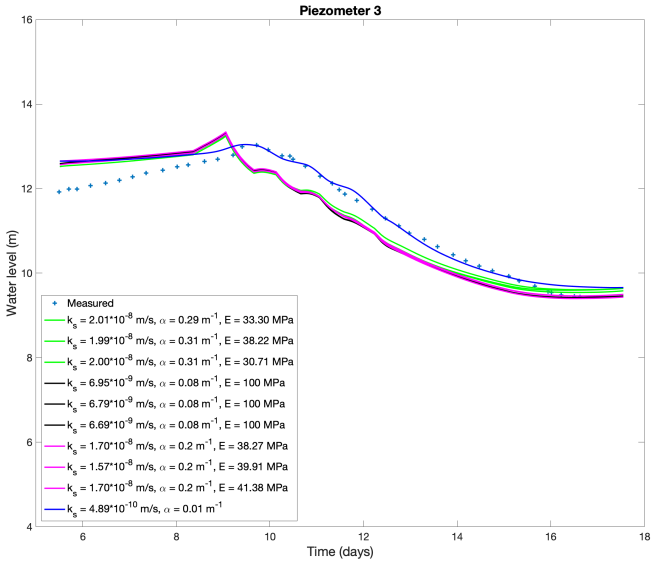
Regardless of the proximity of the identified values to the "true" values of the soil properties, the use of ROM in coupled hydro-mechanical problems is shown to be a valuable tool for exploring groundwater flow problems with back-analysis. The use of ROM significantly reduces the duration of the optimization process, even for a relatively small high-fidelity problem, while yielding accurate results and has a relatively small pre-processing offline cost. Particularly, the ROM seems to yield a more pronounced speedup of the precess in the case where parameters k_s and α are sought for, while E is kept fixed to 100 MPa. This might be related to slower convergence of the linearization iterative scheme (Picard scheme) for higher values of the Young's modulus combined with small values for k_s and α when the FE problem is solved. In the other two problems, where E is inferred, less solutions are computed for high E values, as these yield a large objective function value, that is, a low fitness. By consequence, as the optimization algorithm progresses, the individuals that correspond to high E values comprise a smaller part of the population. The efficiency gains are expected to be even higher when applied



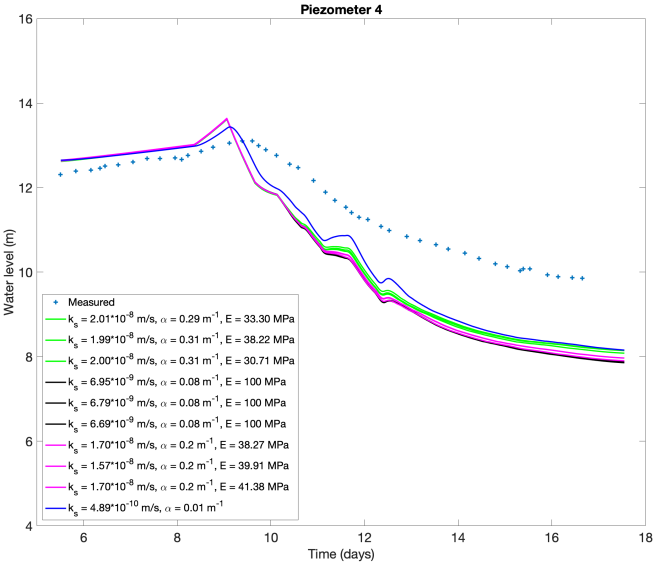
(a) Pressure heads measured and computed on Piezometer 1



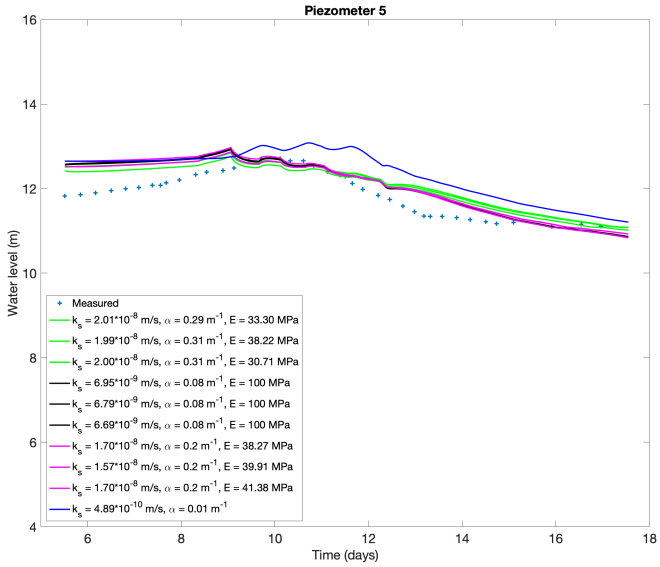
(b) Pressure heads measured and computed on Piezometer 2



(c) Pressure heads measured and computed on Piezometer 3



(d) Pressure heads measured and computed on Piezometer 4



(e) Pressure heads measured and computed on Piezometer 5

Figure 4.11: Pressure heads measured by the piezometers in the Glen Shira dam, and numerically evaluated using the parametric values obtained by the inverse process. Green color: Simultaneous identification of all 3 parameters. Black color: Simultaneous identification of k_s and α. Magenta color: Simultaneous identification of k_s and α. Blue color: Simultaneous identification of k_s and α based on pure hydraulic analysis.

to larger problems [77].

4.5 Conclusions and discussion

In this study the Reduced Basis method was used to obtain approximate solutions to the parametrized coupled hydro-mechanical problem of groundwater flow through soil, with enhanced computational efficiency. The obtained ROM was used in inverse problem solving for mechanical and hydraulic parameter identification by objective function minimization. It is shown that the ROM yields an adequate level of accuracy for use in inverse problem solving and that it speeds up the process significantly. The uniqueness of the problems solution is also examined, especially regarding the simultaneous identification

of hydraulic parameters k_s and α, as the literature indicates that the problem may suffer from ill-posedness. The study of an example based on synthetically generated data, for a problem of rapid drawdown there is in fact a unique solution in the simultaneous identification of both parameters k_s and α. An objective function based on pore water pressure features multiple local minima but one single global minimum. A regularization is proposed using the norm of the time derivatives of the pressure. This method seems to reduce the effect of local minima in the synthetic problem. When the objective function is tested for the problem of k_s and E inference, the regularizing term seems to increase the objective functions sensitivity with respect to the parameters. It is worth noting though, that the weighting factor must be tuned to each solved problem, so that the two terms of the objective function are of comparable magnitude.

Identification of parameters k_s, α and E is also attempted based on actual pressure head sensor data recorded during a rapid drawdown event in Shira dam. The identification yields realistic values for the parameters that seem to result to pressure head predictions that adequately describe the recorded data. The discrepancy between the data and the prediction is thought to be related to the adopted physical model rather than the inversion method.

It is worth mentioning here that in this Chapter, no hyper-reduction methods were used. The DEIM and LDEIM methods described in Chapter 3 were not employed in this case, because as the author of this work noted, they do not yield increased efficiency but do result to reduced accuracy. That is, in this case, where a rapid drawdown problem is studied, the pore pressure states of the dam vary significantly and rapidly, as the phreatic line fluctuates. Therefore, the FE operators take highly variable values. A POD basis containing information that corresponds to this large range of pressure states is rather large, and consequently, results to a ROM that requires visiting a large number of elements in order to be assembled online. In this case however, as seen in Figure 4.2 the full FE mesh is not particularly dense and results to a relatively small number of degrees of freedom. Therefore, employing the DEIM and/or LDEIM methods result to a model with smaller accuracy, that is not significantly faster to solve than a ROM based merely on the Reduced Basis method. This is a limitation of the hyper-reduction methods, that is however strongly related to the particular problem that is treated.

There is a very wide range of topics related to pressure-based parameter identification in groundwater flow through unsaturated soil that may be studied in the future. Regarding Model Order Reduction, other methods, besides the ones studied in this thesis, may be explored to tackle the nonlinear

problem.

Regarding the issue of inverse problem solving, the effect of number and location of the sensors on the quality of the solution may be examined, as well as the quantification of the requirements for a problem to be well-defined, in terms of number and duration of observations, and the type and magnitude of the variation in pressure and suction registered.

Chapter 5

Conclusions and discussion

5.1 Summary of conclusions

This thesis has dealt with developing Model Order Reduction strategies for parametric hydro-mechanical problems in poroelastic materials with partial saturation. The objective here is to create highly efficient ROMs, that can serve as surrogate models to the expensive FE models, and provide accurate solutions fast enough such that many-query applications like parameter identification through data assimilation are enabled. Moreover, it is also within the scope of the thesis to test the developed ROMs in applications of inverse problem solving. The research questions and objectives that are stated in Chapter 1 have been sufficiently answered, and the findings of this work have been demonstrated in numerical examples. The main results and conclusions drawn from this work are summarized below.

The adopted strategy for Model Order Reduction employs POD-based methods for solution space reduction and system space reduction (or hyper-reduction). The Reduced Basis method was used for the former, solving a parametric problem featuring the mechanical loading of an embankment dam in 2D and 3D. At this stage, no hyper-reduction was applied. The problem was solved with ROM up to 15 times faster than with FEM. It was shown that the gain in efficiency using ROM was higher in the 3D case than in the 2D case, that is, the speedup scales with increasing problem size. Meanwhile, the accuracy of the results with respect to FE results was always kept high, with a relative error in the order of magnitude of 10^{-4}.

The POD-based reduction methodology was then further extended to techniques that specifically aim at treating the nonlinear terms in the problem, like DEIM, and its localized version LDEIM. Introducing system space reduction significantly enhanced the computational efficiency achieved with ROMs. DEIM and LDEIM, in combination with RB were implemented and the resulting models were compared. It was shown that the 3D parametric problem of mechanically coupled water flow through soil can be solved approximately 10 to 70 times faster than with FEM. Again, the computational savings were shown to scale with initial problem size. The method was tested on parametric problems with one and three parameters, and its performance remained similar. The global and the localized DEIM were shown to result into similarly performing models, with LDEIM yielding somewhat faster solvers. The created ROMs accuracy remained within acceptable levels in the order of magnitude of 10^{-4}, unless the selected POD bases for nonlinear function approximation were too small, in which case a large error was produced, and the convergence was lost.

The final contribution of this thesis is an approach to parameter identification based on pore pressure sensor data from embankment dams. Inverse problem solving via objective function minimization was explored, for the simultaneous identification of 3 parameters, hydraulic and mechanical soil properties. A rapid upstream water level drawdown problem was examined, and the inverse problem was solved both with synthetically generated sensor data, and with real observed data from a water dam. A ROM was used for model evaluation in the function minimization process. The objective function that was adopted is based on the squared differences between measured and observed water pressures, and includes a regularization term based on the time derivative of the pressures. The objective function without a regularization term features multiple local minima and is rather flat around the global minimum. Adding the regularization term results to a sharper function around the global minimum and reduces the number of local minima. Moreover, it was shown that using a ROM for model evaluation significantly accelerates the minimization process, and does not affect the inverse problem solution. Overall, POD-based surrogates are found to be an appropriate tool for performing parameter identification via objective function optimization.

A synthesis of the conclusions drawn from all three contributions in this study may be summarized in the following points answering the identified research questions

- The Reduced Basis method for model order reduction yields accurate results. The cost of the offline stage is marginal compared to the computational savings that can be achieved in the context of many query problems, due to the fact that only a few snapshots are required to obtain an accurate basis. This holds true even when the parametric domain is of higher dimension.

- Hyper-reduction with DEIM or LDEIM for hydro-mechanical problems with nonlinearities related to the partially saturated state of the soil can significantly enhance the computational efficiency of the model by reducing the cost of assembling FE operators affected by nonlinearity. If large enough POD bases are used for the nonlinear function approximation, the scheme yields results that are as accurate as the ones achieved by assembling the full operators.

- Introducing DEIM or LDEIM to the problem implies a more expensive offline stage. A denser sampling is required to obtain DEIM approximations that result into accurate results. However, depending on the type

of application for which the ROM might be used, the computational savings achieved by the implementation of DEIM in the online stage render it an advantageous strategy despite the expensive offline stage.

- The computational savings obtained by applying RB, DEIM and LDEIM all scale with the size of the treated full-order problem.

- POD-based ROMs are a suitable option for inverse problem solving for parameter identification. Using a ROM as a surrogate to FEM yields accurate results and accelerates the process.

It is worth noting here, that all results that have been reported in this thesis may be highly-problem dependent. Even though the implementation of the presented methods on illustrative examples indicate the order of magnitude of the possible computational savings, those results depend on the complexity of a particular problem and may be more or less favorable when applied to different cases. The same holds true in parameter identification via inverse problem solving, as it has been mentioned in Chapter 4

5.2 Future Developments

The work presented in this thesis, has fulfilled the project objectives and provided an answer to the main research questions posed. In the course of the thesis, various possibilities for future work and extensions of the examined material have become clear. Some of the issues that may be investigated as a continuation of this work are listed below:

- **Physical model:** In this study the soil is considered to behave according to a linear elastic constitutive law. This assumption was adopted in an effort to approach a simpler version of the problem. The intention was to develop a ROM that efficiently treats the nonlinearities that arise due to partial saturation of the soil, ignoring at this stage the nonlinearities that may arise due to a non-elastic constitutive model for the soil. However, this may quickly become an unrealistic assumption for soil in geotechnical applications. There is extensive literature on the mechanical behavior of soils, and non-elastic models have been proposed, that consider many of the complex physical phenomena that govern the behavior of this highly heterogeneous, multiphase material [33, 2, 34, 105]. Studying Model Order Reduction for the coupled hydro-mechanical problem considering

a more involved constitutive model remains an interesting topic to be studied.

- **Hyper-reduction:** As explained in Chapter 1 and 3, various methods have been developed in order to efficiently treat nonlinear problems in the context of Model Order Reduction. In this work DEIM was examined, and found to yield significant computational savings. Studying the advantages of other methods, like TPWL, and comparing the results in terms of accuracy and efficiency may be an interesting field for future investigation.

- **Inverse problem solving** is a wide topic that was only partially treated in this work. The main goal of the work was to identify whether POD-based ROMs can be used as surrogates for this type of problem. However, many aspects of inverse problem solving remain open for investigation. Possible topics to be explored are the conditions required for the inverse problem to be well posed, like the number of spatial and temporal observations required, the type of measurement that is most informative (pressure, suction, displacement, variability in the measured quantity), and the optimal placement of the sensors. Similar tasks have been undertaken by the authors of [71] regarding water flow problems through soil in general, by the authors of [128, 124], for hydro-mechanical inverse problems in a probabilistic context. Exploring the issue in the specific context of deterministic model calibration via optimization in embankment dams, may yield interesting results.

- **Artificial Neural Networks for efficient objective function evaluation:** POD-based Model Order Reduction was found to yield significant computational savings in the context of parameter identification via objective function minimization. However, the ROMs that have been studied in the present have achieved significant speedup, but not a real-time response. In the context of Digital Twins and IoT technologies, solving a many-query problem in real-time is essential, as it enables fast decision-making, informed by quickly obtained what-if scenarios and optimization applications. To achieve an efficient model response, various authors have used Neural Network technologies to replace the forward FE model [27, 90, 112]. The training of an Artificial Neural Network (ANN) however, is in itself a computationally expensive process that requires a large number of forward-model evaluations. It may be worth investigating in the future, the possibility to use a ROM

as a surrogate model in the training of ANNs to reduce the overall computational effort. Besides, it was shown in this work that a rather small offline stage is required to create an accurate ROM, at least as far as the solution space approximation is concerned. Moreover, the full-order function evaluations that are required for the offline stage in the creation of a ROM can serve as training data sets for the ANN. After the completion of the offline stage, more training data sets can be generated using a ROM. It may be worth exploring if a combination of ROM and ANN technologies can yield a computationally optimal scheme for real-time model evaluation.

- **Probabilistic approach to data assimilation:** As discussed in Chapter 4, the problem of parameter identification in water flow through porous material, is often approached in a probabilistic manner in the literature. These methods allow for more flexibility in accounting for the spatial variability in soil properties [95, 84, 118, 85]. It would be interesting to approach parameter identification in the context of embankment dam monitoring using stochastic methods (Maximum likelihood [63], Bayesian updating [53, 84, 128, 125], Kalman filter [123, 118]) and POD-based surrogate models, in order to compare the advantages and disadvantages that the two approaches involve for geotechnical applications.

- **Using ROM in other many-query applications in embankment dam optimization and monitoring:** In this work ROM was investigated for parameter identification, but in the context of Digital Twins, other many-query problems related to the optimization of an embankment dam design and functioning may be pertinent. Specifically for tailings dams, determining maximum allowed level raise rate, or impounding rate within safety-imposed limitations are examples of problems that can be solved with optimization approaches and require high computational efficiency. Moreover, ROM may be employed to efficiently treat geometrical uncertainty [133, 4, 104] in tailings dams. Parametric problems with geometric parameters may be defined to describe the uncertainty in the location of zones of different materials in the impoundment of the dam, or the uncertainty that relates to deformations in parts of the structure where sensor observations are not available. Besides, approaching problems that involve geometric parameters with ROM may enable the solution of design optimization problems, at all stages of the life cycle of

these ever-evolving structures.

Bibliography

[1] M. S. Alnaes, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The FEniCS Project Version 1.5. *Archive of Numerical Software*, 3:9–23, 2015.

[2] E. E. Alonso, A. Gens, and A. Josa. A constitutive model for partially saturated soils. *Géotechnique*, 40(3):405–430, Sept. 1990.

[3] E. E. Alonso, S. Olivella, and N. M. Pinyol. A review of Beliche Dam. *Géotechnique*, 55(4):267–285, 2005.

[4] A. Ammar, A. Huerta, F. Chinesta, E. Cueto, and A. Leygue. Parametric solutions involving geometry: A step towards efficient shape optimization. *Computer Methods in Applied Mechanics and Engineering*, 268:178–193, Jan. 2014.

[5] D. Amsallem, M. J. Zahr, and C. Farhat. Nonlinear model order reduction based on local reduced-order bases. *International Journal for Numerical Methods in Engineering*, 92(10):891–916, Dec. 2012.

[6] H. Antil, M. Heinkenschloss, and D. C. Sorensen. Application of the Discrete Empirical Interpolation Method to Reduced Order Modeling of Nonlinear and Parametric Systems. In A. Quarteroni and G. Rozza, editors, *Reduced Order Methods for Modeling and Computational Reduction*, volume 9, pages 101–136. Springer International Publishing, Cham, 2014.

[7] P. Astrid. Reduction of process simulation models:a proper orthogonal decomposition approach. 2004. Publisher: Technische Universiteit Eindhoven.

[8] A. Aziz, O. Schelén, and U. Bodin. A Study on Industrial IoT for the Mining Industry: Synthesized Architecture and Open Research Directions. *IoT*, 1(2):529–550, Dec. 2020.

[9] F. Azizi. *Applied analyses in geotechnics*. E & FN Spon, London ; New York, 2000.

[10] I. Babuška. Error-bounds for Finite Element Method. *Numerische Mathematik*, 16(4):322–333, 1971.

[11] S. Badia, A. Quaini, and A. Quarteroni. Coupling Biot and Navier–Stokes equations for modelling fluid–poroelastic media interaction. *Journal of Computational Physics*, page 29, 2009.

[12] A. Bao, E. Gildin, A. Narasingam, and J. S. Kwon. Data-Driven Model Reduction for Coupled Flow and Geomechanics Based on DMD Methods. *Fluids*, 4(3):138, July 2019.

[13] L. Barnewold and B. G. Lottermoser. Identification of digital technologies and digitalisation trends in the mining industry. *International Journal of Mining Science and Technology*, 30(6):747–757, Nov. 2020.

[14] R. Bhanbhro. *Mechanical Properties of Tailings: Basic Description of a Tailings Material from Sweden*. PhD thesis, 2014. Publisher: Unpublished.

[15] J. Booker and J. Small. An investigation of the stability of numerical solutions of Biot's equations of consolidation. *International Journal of Solids and Structures*, 11(7-8):907–917, July 1975.

[16] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 8(R2):129–151, 1974.

[17] M. Calvello and R. J. Finno. Selecting parameters to optimize in model calibration by inverse analysis. *Computers and Geotechnics*, 31(5):410–424, July 2004.

[18] M. Cardoso and L. Durlofsky. Linearized reduced-order models for subsurface flow simulation. *Journal of Computational Physics*, 229(3):681–700, Feb. 2010.

[19] M. A. Celia, E. T. Bouloutas, and R. L. Zarba. A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resources Research*, 26(7):1483–1496, July 1990.

[20] R. Charlier, F. Collin, B. Pardoen, J. Talandier, J.-P. Radu, and P. Gerard. An unsaturated hydro-mechanical modelling of two in-situ experiments in Callovo-Oxfordian argillite. *Engineering Geology*, 165:46–63, Oct. 2013.

[21] S. Chaturantabut and D. C. Sorensen. Nonlinear Model Reduction via Discrete Empirical Interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, Jan. 2010.

[22] Y. Chen. Characterization of transient groundwater flow through a high arch dam foundation during reservoir impounding. *Journal of Rock Mechanics and Geotechnical Engineering*, 8(4):462–471, 2016.

[23] F. Chinesta, A. Ammar, and E. Cueto. Recent Advances and New Challenges in the Use of the Proper Generalized Decomposition for Solving Multidimensional Models. *Archives of Computational Methods in Engineering*, 17(4):327–350, Dec. 2010.

[24] F. Chinesta, R. Keunings, and A. Leygue. *The Proper Generalized Decomposition for Advanced Numerical Simulations*. SpringerBriefs in Applied Sciences and Technology. Springer International Publishing, Cham, 2014.

[25] L. Clarkson, D. Williams, and J. Seppälä. Real-time monitoring of tailings dams. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, pages 1–15, Mar. 2020.

[26] M. Davies and T. Martin. Static liquefaction of tailings–fundamentals and case histories. In *Proceedings of Tailings Dams ASDSO/USCOLD*, pages 233–255, Las Vegas 2002, 2002.

[27] J. Deng and C. Lee. Displacement back analysis for a steep slope at the Three Gorges Project site. *International Journal of Rock Mechanics and Mining Sciences*, 38(2):259–268, Feb. 2001.

[28] S. O. Eching and J. W. Hopmans. Optimization of Hydraulic Functions from Transient Outflow and Soil Water Pressure Data. *Soil Science Society of America Journal*, 57(5):1167–1175, Sept. 1993.

[29] M. Esmaeili, M. Ahmadi, and A. Kazemi. A generalized DEIM technique for model order reduction of porous media simulations in reservoir optimizations. *Journal of Computational Physics*, 422:109769, Dec. 2020.

[30] I. Fatt and D. Davis. Reduction in Permeability With Overburden Pressure. *Journal of Petroleum Technology*, 4(12):16–16, Dec. 1952.

[31] E. Florentin and P. Díez. Adaptive reduced basis strategy based on goal oriented error assessment for stochastic problems. *Computer Methods in Applied Mechanics and Engineering*, 225-228:116–127, 2012.

[32] D. G. Fredlund and S. L. Houston. Interpretation of Soil-Water Characteristic Curves when Volume Change Occurs as Soil Suction is Changed. In *Advances in Unsaturated Soils - Proceedings of the 1st Pan-American Conference on Unsaturated Soils, PanAmUNSAT 2013*, Jan. 2013.

[33] D. G. Fredlund and H. Rahardjo. *Soil mechanics for unsaturated soils*. Wiley, New York, 1993.

[34] D. Gallipoli, A. Gens, R. Sharma, and J. Vaunat. An elasto-plastic model for unsaturated soil incorporating the effects of suction and degree of saturation on mechanical behaviour. *Géotechnique*, 53(1):123–135, Feb. 2003.

[35] L. A. Garcia and A. Shigidi. Using neural networks for parameter estimation in ground water. *Journal of Hydrology*, 318(1-4):215–231, Mar. 2006.

[36] A. Gens and E. E. Alonso. Aznalcóllar dam failure. Part 2: Stability conditions and failure mechanism. *Géotechnique*, 56(3):185–201, Apr. 2006.

[37] P. Gerard, A. Leonard, J.-P. Masekanya, R. Charlier, and F. Collin. Study of the soil-atmosphere moisture exchanges through convective drying tests in non-isothermal conditions. *International Journal for Numerical and Analytical Methods in Geomechanics*, 34(12):1297 – 1320, 2009.

[38] C. Geuzaine and J.-F. Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities: THE GMSH PAPER. *International Journal for Numerical Methods in Engineering*, 79(11):1309–1331, Sept. 2009.

[39] M. Ghasemi and E. Gildin. Localized model order reduction in porous media flow simulation. *Journal of Petroleum Science and Engineering*, 145:689–703, Sept. 2016.

[40] M. Ghommem, E. Gildin, and M. Ghasemi. Complexity Reduction of Multiphase Flows in Heterogeneous Porous Media. *SPE Journal*, 21(01):144–151, Feb. 2016.

[41] E. Gildin, M. Ghasemi, A. Romanovskay, and Y. Efendiev. Nonlinear Complexity Reduction for Fast Simulation of Flow in Heterogeneous Porous Media. In *All Days*, pages SPE–163618–MS, The Woodlands, Texas, USA, Feb. 2013. SPE.

[42] G. Gioda and S. Sakurai. Back analysis procedures for the interpretation of field measurements in geomechanics. *International Journal for Numerical and Analytical Methods in Geomechanics*, 11:555–583, 1987.

[43] M. M. Gribb. Parameter Estimation for Determining Hydraulic Properties of a Fine Sand From Transient Flow Measurements. *Water Resources Research*, 32(7):1965–1974, July 1996.

[44] S. Haag and R. Anderl. Digital twin – Proof of concept. *Manufacturing Letters*, 15:64–66, Jan. 2018.

[45] T. Hamade. *Geotechnical Design of Tailings Dams - A Stochastic Analysis Approach*. PhD thesis, McGill University, 2013.

[46] J. He and L. J. Durlofsky. Reduced-Order Modeling for Compositional Simulation by Use of Trajectory Piecewise Linearization. *SPE Journal*, 19(05):858–872, Oct. 2014.

[47] J. He and L. J. Durlofsky. Constraint reduction procedures for reduced-order subsurface flow models based on POD-TPWL: CONSTRAINT REDUCTION FOR POD-TPWL. *International Journal for Numerical Methods in Engineering*, 103(1):1–30, July 2015.

[48] J. He, J. Sætrom, and L. Durlofsky. Enhanced linearized reduced-order models for subsurface flow simulation. *Journal of Computational Physics*, 230(23):8313–8341, Sept. 2011.

[49] A. A. Heshmati R., H. Salehzadeh, and M. Shahidi. Prediction of the Void Ratio Parameter in Mineral Tailings Using Gene Expression Programming. *Advances in Civil Engineering*, 2020:1–12, Oct. 2020.

[50] J. S. Hesthaven, G. Rozza, and B. Stamm. *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*. SpringerBriefs in Mathematics. Springer International Publishing, Cham, 2016.

[51] K. C. Hoang, T.-Y. Kim, and J.-H. Song. Fast and accurate two-field reduced basis approximation for parametrized thermoelasticity problems. *Finite Elements in Analysis and Design*, 141:96–118, 2018.

[52] J.-M. Hong, Y.-F. Chen, M.-M. Liu, and C.-B. Zhou. Inverse modelling of groundwater flow around a large-scale underground cavern system considering the excavation-induced hydraulic conductivity variation. *Computers and Geotechnics*, 81:346–359, Jan. 2017.

[53] Y. Honjo, L. Wen-Tsung, and S. Guha. Inverse analysis of an embankment on soft clay by extended Bayesian method. *International Journal for Numerical and Analytical Methods in Geomechanics*, 18(10):709–734, Oct. 1994.

[54] S. R. Hui, L. Charlebois, and C. Sun. Real-time monitoring for structural health, public safety, and risk management of mine tailings dams. *Canadian Journal of Earth Sciences*, 55(3):221–229, Mar. 2018.

[55] M. Inoue, J. Šimunek, J. W. Hopmans, and V. Clausnitzer. In situ estimation of soil hydraulic functions using a multistep soil-water extraction technique. *Water Resources Research*, 34(5):1035–1050, May 1998.

[56] Z. L. Jin, T. Garipov, O. Volkov, and L. J. Durlofsky. Reduced-Order Modeling of Coupled Flow and Quasistatic Geomechanics. *SPE Journal*, 25(01):326–346, Feb. 2020.

[57] G. Kirsten and V. Simoncini. A matrix-oriented POD-DEIM algorithm applied to semilinear matrix differential equations. *arXiv:2006.13289 [cs, math]*, May 2021. arXiv: 2006.13289.

[58] R. Knutsson, A. Bjelkevik, and S. Knutsson. Slope stability in landform design. In A. Fourie and M. Tibbett, editors, *Proceedings of the 11th International Conference on Mine Closure*, pages 89–98. Australian Centre for Geomechanics, 2016. event-place: Perth.

[59] J. B. Kool and J. C. Parker. Analysis of the inverse problem for transient unsaturated flow. *Water Resources Research*, 24(6):817–830, June 1988.

[60] D. Kossoff, W. E. Dubbin, M. L. Alfredsson, S. J. Edwards, M. G. Macklin, and K. A. Hudson-Edwards. Mine tailings dams: Characteristics, failure, environmental impacts, and remediation. *Applied Geochemistry*, 51:229–245, Dec. 2014.

[61] Y. Larion, S. Zlotnik, T. Massart, and P. Díez. Building a certified reduced basis for coupled thermo-hydro-mechanical systems with goal-oriented error estimation. *Computational Mechanics*, 66, 09 2020.

[62] Y. Larion, S. Zlotnik, T. J. Massart, and P. Díez. Building a certified reduced basis for coupled thermo-hydro-mechanical systems with goal-oriented error estimation. *Computational Mechanics*, 66(3):559–573, Sept. 2020.

[63] A. Ledesma, A. Gens, and E. Alonso. Estimation of parameters in geotechnical backanalysis — I. Maximum likelihood approach. *Computers and Geotechnics*, 18(1):1–27, Jan. 1996.

[64] O. Ledesma, A. Sfriso, and D. Manzanal. Procedure for assessing the liquefaction vulnerability of tailings dams. *Computers and Geotechnics*, 144:104632, Apr. 2022.

[65] Y.-B. Li, Y. Liu, W.-B. Nie, and X.-Y. Ma. Inverse Modeling of Soil Hydraulic Parameters Based on a Hybrid of Vector-Evaluated Genetic Algorithm and Particle Swarm Optimization. *Water*, 10(1):84, Jan. 2018.

[66] F. List and F. A. Radu. A study on iterative methods for solving Richards' equation. *Computational Geosciences*, 20(2):341–353, Apr. 2016.

[67] M.-l. Lu and L. Cui. Three-dimensional seepage analysis for complex topographical tailings dam. *Yantu Lixue(Rock and Soil Mechanics)*, 27(7):1176–1180, 2006.

[68] Z. Lyu, J. Chai, Z. Xu, Y. Qin, and J. Cao. A Comprehensive Review on Reasons for Tailings Dam Failures Based on Case History. *Advances in Civil Engineering*, 2019:1–18, June 2019.

[69] Y. Maday and E. M. Ronquist. The Reduced Basis Element Method: Application to a Thermal Fin Problem. *SIAM Journal on Scientific Computing*, 26(1):240–258, Jan. 2004.

[70] Y. Maday and E. M. Rønquist. A reduced-basis element method. *Comptes Rendus Mathematique*, 335(2):195–200, Jan. 2002.

[71] D. Mao, T.-C. J. Yeh, L. Wan, K.-C. Hsu, C.-H. Lee, and J.-C. Wen. Necessary conditions for inverse modeling of flow through variably saturated porous media. *Advances in Water Resources*, 52:50–61, Feb. 2013.

[72] T. E. Martin and E. C. McRoberts. Some considerations in the stability analysis of upstream tailings dams. page 18, 1999.

[73] C. J. Miller, N. Yesiller, K. Yaldo, and S. Merayyan. Impact of Soil Type and Compaction Conditions on Soil Water Characteristic. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(9):733–742, Sept. 2002.

[74] K. L. Morton. The Use of Accurate Pore Pressure Monitoring for Risk Reduction in Tailings Dams. *Mine Water and the Environment*, 40(1):42–49, Mar. 2021.

[75] K. Moustafa Kamel Ehab, P. Gerard, J.-B. Colliat, and T. J. Massart. Modelling stress-induced permeability alterations in sandstones using CT scan-based representations of the pore space morphology. *International Journal of Rock Mechanics and Mining Sciences*, 150, 2022.

[76] E. Nadal, F. Chinesta, P. Díez, F. Fuenmayor, and F. Denia. Real time parameter identification and solution reconstruction from experimental data using the Proper Generalized Decomposition. *Computer Methods in Applied Mechanics and Engineering*, 296:113–128, Nov. 2015.

[77] C. Nasika, P. Díez, P. Gerard, T. J. Massart, and S. Zlotnik. Towards real time assessment of earthfill dams via Model Order Reduction. *Finite Elements in Analysis and Design*, 199:103666, Feb. 2022.

[78] F. Negri, A. Manzoni, and D. Amsallem. Efficient model reduction of parametrized systems by matrix discrete empirical interpolation. *Journal of Computational Physics*, 303:431–454, Dec. 2015.

[79] M. Nuth and L. Laloui. Advances in modelling hysteretic water retention curve in deformable soils. *Computers and Geotechnics*, 35(6):835–844, Nov. 2008.

[80] S. Oh and N. Lu. Uniqueness of the Suction Stress Characteristic Curve under Different Confining Stress Conditions. *Vadose Zone Journal*, 13(5):1–10, May 2014.

[81] S. Olivella, A. Gens, J. Carrera, and E. Alonso. Numerical formulation for a simulator (CODE_bright) for the coupled analysis of saline media. *Engineering Computations*, 13(7):87–112, Nov. 1996.

[82] O. Ortega-Gelabert, S. Zlotnik, J. C. Afonso, and P. Díez. Fast Stokes Flow Simulations for Geophysical-Geodynamic Inverse Problems and Sensitivity Analyses Based On Reduced Order Modeling. *Journal of Geophysical Research: Solid Earth*, 125(3):25, Mar. 2020.

[83] S. Pagani, A. Manzoni, and A. Quarteroni. Numerical approximation of parametrized problems in cardiac electrophysiology by a local reduced basis method. *Computer Methods in Applied Mechanics and Engineering*, 340:530–558, Oct. 2018.

[84] I. Papaioannou and D. Straub. Reliability updating in geotechnical engineering including spatial variability of soil. *Computers and Geotechnics*, 42:44–51, May 2012.

[85] I. Papaioannou and D. Straub. Learning soil parameters and updating geotechnical reliability estimates under spatial variability – theory and application to shallow foundations. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 11(1):116–128, Jan. 2017.

[86] J. Paton and N. Semple. Investigation of the stability of an earth dam subjected to rapid drawdown including details of pore pressure recorded during a controlled drawdown test. In *Pore Pressure and Suction in Soils*, pages 85–90. Butterworth, London, 1961.

[87] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, and D. Cournapeau. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[88] B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz. Localized Discrete Empirical Interpolation Method. *SIAM Journal on Scientific Computing*, 36(1):A168–A192, Jan. 2014.

[89] K.-K. Phoon and F. H. Kulhawy. Characterization of geotechnical variability. *Canadian Geotechnical Journal*, 36(4):612–624, 1999.

[90] B. Pichler, R. Lackner, and H. A. Mang. Back analysis of model parameters in geotechnical engineering by means of soft computing. *International Journal for Numerical Methods in Engineering*, 57(14):1943–1978, Aug. 2003.

[91] N. M. Pinyol, E. E. Alonso, and S. Olivella. Rapid drawdown in slopes and embankments. *Water Resources Research*, 44(5), May 2008.

[92] Y. Qiu and D. C. Sego. Laboratory properties of mine tailings. *Canadian Geotechnical Journal*, 38(1):183–190, Feb. 2001.

[93] A. Quarteroni, A. Manzoni, and F. Negri. *Reduced Basis Methods for Partial Differential Equations*, volume 92 of *UNITEXT*. Springer International Publishing, Cham, 2016.

[94] A. Quarteroni and G. Rozza, editors. *Reduced Order Methods for Modeling and Computational Reduction*. Springer International Publishing, Cham, 2014.

[95] R. Rackwitz. Reviewing probabilistic soils modelling. *Computers and Geotechnics*, 26(3-4):199–223, Apr. 2000.

[96] N. S. A. Rashid, M. Askari, T. Tanaka, J. Simunek, and M. T. van Genuchten. Inverse estimation of soil hydraulic properties under oil palm trees. *Geoderma*, 241-242:306–312, Mar. 2015.

[97] C. Rechea, S. Levasseur, and R. Finno. Inverse analysis techniques for parameter identification in simulation of excavation support systems. *Computers and Geotechnics*, 35:331–345, 05 2008.

[98] M. Rewienski and J. White. A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(2):155–170, Feb. 2003.

[99] M. Rewieński and J. White. Model order reduction for nonlinear dynamical systems based on trajectory piecewise-linear approximations. *Linear Algebra and its Applications*, 415(2-3):426–454, June 2006.

[100] G. Rozza, D. B. P. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Archives of Computational Methods in Engineering*, 15(3):1–47, Sept. 2007.

[101] B. Saad and H. Mitri. Hydromechanical Analysis of Upstream Tailings Disposal Facilities. *Journal of Geotechnical and Geoenvironmental Engineering*, 137(1):27–42, Jan. 2011.

[102] S. Sakurai and K. Takeuchi. Back analysis of measured displacements of tunnels. *Rock Mechanics and Rock Engineering*, 16:173–180, 1983.

[103] A. P. S. Selvadurai and A. P. Suvorov. *Thermo-Poroelasticity and Geomechanics*. Cambridge University Press, 2016.

[104] R. Sevilla, L. Borchini, M. Giacomini, and A. Huerta. Hybridisable discontinuous Galerkin solution of geometrically parametrised Stokes flows. *Computer Methods in Applied Mechanics and Engineering*, 372:113397, Dec. 2020.

[105] D. Sheng. Review of fundamental principles in modelling unsaturated soil behaviour. *Computers and Geotechnics*, 38(6):757–776, Sept. 2011.

[106] L. H. Silva Rotta, E. Alcântara, E. Park, R. G. Negri, Y. N. Lin, N. Bernardo, T. S. G. Mendes, and C. R. Souza Filho. The 2019 Brumadinho tailings dam collapse: Possible cause and impacts of the worst human and environmental disaster in Brazil. *International Journal of Applied Earth Observation and Geoinformation*, 90:102119, Aug. 2020.

[107] R. Storn and K. Price. Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[108] A. Szostak, A. Chrzanowski, and M. Massiéra. Use of geodetic monitoring measurements in solving geomechanical problems in structural and mining engineering. page 9, Santorini, Greece, 2003.

[109] K. Terzaghi. Evaluation of coefficients of subgrade reaction. *Géotechnique*, 5(4):297–326, 1955.

[110] A. F. Toorman, P. J. Wierenga, and R. G. Hills. Parameter estimation of hydraulic properties from one-step outflow data. *Water Resources Research*, 28(11):3021–3028, Nov. 1992.

[111] J. Toromanovic, H. Mattsson, S. Knutsson, and J. Laue. Parameter identification for an embankment dam using noisy field data. In *Proceedings of the Institution of Civil Engineers - Geotechnical Engineering*, pages 1–42, Jan. 2020.

[112] S. M. Vaezinejad, S. M. Marandi, and E. Salajegheh. Inverse modelling of leakage through earth dams (case study: Baft dam, Iran). *Geotechnical Research*, 5(4):218–230, Dec. 2018.

[113] J. C. van Dam, J. N. M. Stricker, and P. Droogers. Inverse Method for Determining Soil Hydraulic Functions from One-Step Outflow Experiments. *Soil Science Society of America Journal*, 56(4):1042–1050, July 1992.

[114] J. F. M. van Doren, R. Markovinović, and J.-D. Jansen. Reduced-order optimal control of water flooding using proper orthogonal decomposition. *Computational Geosciences*, 10(1):137–158, Mar. 2006.

[115] M. T. van Genuchten. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils1. *Soil Science Society of America Journal*, 44(5):892, 1980.

[116] J.-F. Vanden Berghe, J.-C. Ballard, J.-F. Wintgens, and B. List. Geotechnical Risks Related to Tailings Dam Operations. In *Tailings and Mine Waste Conference (2011 : Vancouver, B.C.)*, page 11, 2011.

[117] S. Vardakos, M. Gutierrez, and C. Xia. Parameter identification in numerical modeling of tunneling using the Differential Evolution Genetic Algorithm (DEGA). *Tunnelling and Underground Space Technology*, 28:109–123, Mar. 2012.

[118] P. J. Vardon, K. Liu, and M. A. Hicks. Reduction of slope stability uncertainty based on hydraulic measurement via inverse analysis. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 10(3):223–240, July 2016.

[119] P. Vermeulen, A. Heemink, and C. Te Stroet. Reduced models for linear groundwater flow models using empirical orthogonal functions. *Advances in Water Resources*, 27(1):57–69, Jan. 2004.

[120] A. G. Villavicencio, P. Breul, C. Bacconnet, D. Boissier, and A. R. Espinace. Estimation of the Variability of Tailings Dams Properties in Order to Perform Probabilistic Assessment. *Geotechnical and Geological Engineering*, 29(6):1073–1084, Nov. 2011.

[121] J.-P. Wang, P.-Z. Zhuang, J.-Y. Luan, T.-H. Liu, Y.-R. Tan, and J. Zhang. Estimation of unsaturated hydraulic conductivity of granular soils from particle size parameters. *Water*, 11(9), 2019.

[122] W. Whalley, G. Matthews, and S. Ferraris. The effect of compaction and shear deformation of saturated soil on hydraulic conductivity. *Soil and Tillage Research*, 125:23–29, Sept. 2012.

[123] C. X. Yang, Y. H. Wu, T. Hon, and X.-T. Feng. Application of extended Kalman filter to back analysis of the natural stress state accounting for measuring uncertainties. *International Journal for Numerical and Analytical Methods in Geomechanics*, 35(6):694–712, Apr. 2011.

[124] H.-Q. Yang, X. Chen, L. Zhang, J. Zhang, X. Wei, and C. Tang. Conditions of Hydraulic Heterogeneity under Which Bayesian Estimation is More Reliable. *Water*, 12(1):160, Jan. 2020.

[125] H.-Q. Yang, L. Zhang, and D.-Q. Li. Efficient method for probabilistic estimation of spatially varied hydraulic properties in a soil slope based on field responses: A Bayesian approach. *Computers and Geotechnics*, 102:262–272, Oct. 2018.

[126] D. W. Zachmann, P. C. DuChateau, and A. Klute. The Calibration of the Richards Flow Equation for a Draining Column by Parameter Identification. *Soil Science Society of America Journal*, 45(6):1012–1015, Nov. 1981.

[127] C. Zhang, J. Chai, J. Cao, Z. Xu, Y. Qin, and Z. Lv. Numerical Simulation of Seepage and Stability of Tailings Dams: A Case Study in Lixi, China. *Water*, 12(3):742, Mar. 2020.

[128] L. Zhang, Y. Zheng, L. Zhang, X. Li, and J. Wang. Probabilistic model calibration for soil slope under rainfall: effects of measurement duration and frequency in field monitoring. *Géotechnique*, 64(5):365–378, Apr. 2014.

[129] B. D. Zhao, L. L. Zhang, D. S. Jeng, J. H. Wang, and J. J. Chen. Inverse Analysis of Deep Excavation Using Differential Evolution Algorithm. *International Journal for Numerical and Analytical Methods in Geomechanics*, 39(2):115–134, Feb. 2015.

[130] H. Zhao, B. Chen, and S. Li. Determination of geomaterial mechanical parameters based on back analysis and reduced-order model. *Computers and Geotechnics*, 132:104013, Apr. 2021.

[131] H.-b. Zhao and S. Yin. Geomechanical parameters identification by particle swarm optimization and support vector machine. *Applied Mathematical Modelling*, 33(10):3997–4012, Oct. 2009.

[132] C.-B. Zhou, W. Liu, Y.-F. Chen, R. Hu, and K. Wei. Inverse modeling of leakage through a rockfill dam foundation during its construction stage using transient flow model, neural network and genetic algorithm. *Engineering Geology*, 187:183–195, Mar. 2015.

[133] S. Zlotnik, P. Dıez, D. Modesto, and A. Huerta. Proper Generalized Decomposition of a geometrically parametrized heat problem with geophysical applications. page 31, 2014.

[134] J. Šimůnek, M. T. van Genuchten, and O. Wendroth. Parameter Estimation Analysis of the Evaporation Method for Determining Soil Hydraulic Properties. *Soil Science Society of America Journal*, 62(4):894–905, July 1998.