



Perspectives of physics-based machine learning strategies for geoscientific applications governed by partial differential equations

Denise Degen¹, Daniel Caviedes Voullième^{2,3}, Susanne Buiter^{4,5}, Harrie-Jan Hendricks Franssen^{3,6}, Harry Vereecken^{3,6}, Ana González-Nicolás², and Florian Wellmann^{1,7}

¹Computational Geoscience, Geothermics and Reservoir Geophysics (CG3), RWTH Aachen University, Mathieustraße 30, 52074 Aachen, Germany

²Forschungszentrum Jülich GmbH, Jülich Supercomputing Centre (JSC), Simulation and Data Lab. Terrestrial Systems, Wilhelm-Johnen-Straße, 52425 Jülich, Germany

³Forschungszentrum Jülich GmbH, Agrosphere, IBG-3, Wilhelm-Johnen-Straße, 52425 Jülich, Germany

⁴RWTH Aachen University, Tectonics and Geodynamics (TAG), Lochnerstraße 4-20, 52064 Aachen, Germany

⁵Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

⁶Centre for High-Performance Computing Terrestrial Systems, Geoverbund ABC/J, Wilhelm-Johnen-Straße, 52425 Jülich, Germany

⁷Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems (IEG), Am Hochschulcampus 1, 44801 Bochum, Germany

Correspondence: Denise Degen (denise.degen@cgre.rwth-aachen.de)

Received: 21 December 2022 – Discussion started: 27 March 2023

Revised: 25 September 2023 – Accepted: 4 November 2023 – Published: 19 December 2023

Abstract. An accurate assessment of the physical states of the Earth system is an essential component of many scientific, societal, and economical considerations. These assessments are becoming an increasingly challenging computational task since we aim to resolve models with high resolutions in space and time, to consider complex coupled partial differential equations, and to estimate uncertainties, which often requires many realizations. Machine learning methods are becoming a very popular method for the construction of surrogate models to address these computational issues. However, they also face major challenges in producing explainable, scalable, interpretable, and robust models. In this paper, we evaluate the perspectives of geoscience applications of physics-based machine learning, which combines physics-based and data-driven methods to overcome the limitations of each approach taken alone. Through three designated examples (from the fields of geothermal energy, geodynamics, and hydrology), we show that the non-intrusive reduced-basis method as a physics-based machine learning

approach is able to produce highly precise surrogate models that are explainable, scalable, interpretable, and robust.

1 Introduction

Applications in geosciences dealing with land surface and subsurface environments cover a broad range of disciplines and Earth system compartments comprising lithosphere (including deep reservoirs), groundwater, soil, and the land surface. Subsurface environment applications typically deal with the management of natural resources such as subsurface water stored in soils and aquifers (e.g., water supply, contamination) as well as energy (e.g., geothermal energy, hydrocarbons) that require an in-depth understanding of the governing physical, chemical, and biological principles. In addition, subsurface environments are heterogeneous across a variety of spatial scales, and processes in these environments are typically complex, highly nonlinear, and time-dependent. Spatial scales may range from the nanoscale to the basin and

continental scale. Despite the enormous variation in applications in terms of scale and processes, there are underlying common principles. Many of these processes are governed by partial differential equations (PDEs) and are embedded in continuum physics approaches (Bauer et al., 2021; Bergen et al., 2019). In this study, we focus on subsurface applications from the fields of geodynamics, hydrology, and geothermal energy to showcase various methods that are applied in applications which are characterized by data sparsity and difficult accessibility for sampling and measuring states and fluxes as well as for mapping the spatial organization of key subsurface properties.

We focus on continuum PDEs and examine how to accelerate the numerical solution of PDEs that arise from geoscientific applications combining methods from the field of projection-based model order reduction (Benner et al., 2015) and machine learning. Our decision to focus on the aspect of accelerating the solution of PDEs stems from the fact that this is essential in geoscientific applications that aim at improving the understanding of the Earth system with its associated physical processes. An improved understanding requires extensive parameter estimation studies, sensitivity analyses, and uncertainty quantification, which are computationally expensive for many geoscientific PDE-based models (Degen et al., 2022c). This problem is further increased by the circumstance that the governing PDEs are becoming more and more complex as our understanding of the physical processes and their coupling increases. As an example, geothermal studies moved from a purely thermal system to a coupled thermo–hydro–mechanical and, depending on the scale, also a chemical system (e.g., Gelet et al., 2012; Jacquey and Cacace, 2020b; Kohl et al., 1995; Kolditz et al., 2012; O’Sullivan et al., 2001; Stefansson et al., 2020; Taron and Elsworth, 2009). Recent studies include more advanced considerations of the mechanical components, yielding nonlinear hyperbolic PDEs (e.g., Gelet et al., 2012; Jacquey and Cacace, 2017, 2020b, a; Poulet et al., 2014).

The more complex PDEs allow a better characterization of the physical processes in the subsurface. However, they also lead to several challenges.

- *Challenge 1.* In geoscientific applications, one primary interest is to determine which parameters (e.g., rock and fluid properties) influence the state distribution (e.g., temperature, pressure, stress) to what extent. This question can be addressed through sensitivity analyses, which are categorized into local and global sensitivity analysis (SA) (Wainwright et al., 2014). Local SAs are computationally fast and require only a few simulations. However, they neither return information on the parameter correlation nor allow the exploration of the entirety of the parameter space (Wainwright et al., 2014; Sobol, 2001). Global SAs overcome these issues but are computationally challenging because they demand numer-

ous simulations (Degen et al., 2021b; Degen and Cacace, 2021).

- *Challenge 2.* Geoscientific and geophysical investigations are often subject to various sources of uncertainty. To assess these uncertainties, stochastic descriptions of the PDEs or probabilistic inverse methods such as Markov chain Monte Carlo are commonly employed. However, these methods have in common that they require solving many forward evaluations, yielding an even more computationally demanding problem (Rogelj and Knutti, 2016; Stewart and Lewis, 2017).
- *Challenge 3.* Often, PDE-based systems are required for predictions and real-time applications (Bauer et al., 2021; Sabeena and Reddy, 2017). Hence, there is the need to solve PDEs fast. This imposes further difficulties since the simulations for fully coupled nonlinear hyperbolic PDEs are too demanding to allow such fast applications. Therefore, they currently have only limited applicability in many geoscientific research areas (Bauer et al., 2021; Bergen et al., 2019; Willcox et al., 2021).

These challenges have one important aspect in common: they require multiple or fast simulations. If a single simulation already requires a long simulation time, then addressing these challenges becomes difficult to prohibitive.

In this work, we focus on the utility of surrogate models as a way to speed up PDE-based simulations of the physical state in geoscientific problems. In particular, we determine how surrogate models (sometimes also referred to as meta models) can be employed to obtain precise predictions at a reduced computational cost. With the increasing interest in machine learning and the availability of various physics-driven approaches, it is becoming apparent that we need to clarify which requirements we have for our surrogate models. Machine learning has become very popular and powerful in application fields where a vast number of data are available (Baker et al., 2019; Willcox et al., 2021). However, Baker et al. (2019) also stated in their report to the United States Department of Energy that machine learning models need to be scalable, interpretable, generalizable, and robust to be attractive for scientific applications, which is also stated in Willcox et al. (2021). This is important not only for scientific machine learning models but also for surrogate models in general.

- *Requirement 1.1, physical scalability.* The system resolutions are constantly increasing since one wants to resolve models with higher and higher resolutions in space and time (Bauer et al., 2021; van Zelst et al., 2022). This increase in numerical system size, which is, for instance, noticeable by the increase in degrees of freedom, yields a computationally demanding problem. As an example in geodynamic simulations, one is

interested both in large- and small-scale deformations simultaneously. For the time component, variations of the order of milliseconds (seismic cycle) and variations of the order of millions of years (basin- and plate-scale geodynamics) are of interest. In hydrological problems, capturing the spatial small-scale heterogeneity of the surface and subsurface, the short-term nature of meteorological forcing (hours to days), and the long-term nature of subsurface fluxes (tens of years and longer) poses similar challenges (Blöschl and Sivapalan, 1995; Condon et al., 2021; Fan et al., 2019; Paniconi and Putti, 2015). This results in computationally expensive problems (van Zelst et al., 2022), which often require high-performance computing (HPC) infrastructures even for solving a single forward problem.

- *Requirement 1.2, performance scalability.* Porting applications to HPC infrastructures is another major requirement, which requires the rewrite of major parts of the software packages (Alexander et al., 2020; Artigues et al., 2019; Baig et al., 2020; Bauer et al., 2021; Bertagna et al., 2019; Lawrence et al., 2018; Gan et al., 2020; Grete et al., 2021; Hokkanen et al., 2021). Therefore, one important aspect is to evaluate how physics-based machine learning methods can help in the transition to modern HPC infrastructures.
- *Requirement 2, interpretability.* Another main requirement is to produce surrogate models that maintain the characteristic of PDE solutions and the original structure of the governing principles (i.e., that they map from model parameters to state information). Related to this is the consideration of nonlinear problems (Grepl, 2012; Hesthaven and Ubbiali, 2018; Reichstein et al., 2019; van Zelst et al., 2022), which, in the field of geosciences, is often related to the consideration of mechanical or complex flow effects. Such nonlinear problems result in extensive computation times because of extra iterations in the numerical solve (van Zelst et al., 2022).
- *Requirement 3, generalizability.* Furthermore, it is important that the approaches apply to various disciplines and are not restricted to single applications only. To enable a transfer between disciplines it is important to provide accessible solutions. Often software packages are developed in-house and are not available to the entire community.
- *Requirement 4, robustness.* The last requirement is robustness. It is critical that the solutions produced by the surrogate models are consistent for different levels of accuracy. As an example, it is not desired to have a surrogate model that predicts lower accuracy for a diffusion-dominated solution and slightly increased accuracy for an advection-dominated solution.

The paper is structured as follows: we present the PDEs considered in this study first (see Sect. 2). Afterwards, we present the concept of data-driven, physics-based approaches and physics-based machine learning in the same section. In Sect. 3, we present three designated case studies, from the field of geothermal energy, geodynamics, and hydrology, to illustrate the potential of physics-based machine learning. This is further emphasized in Sect. 4, where we evaluate how physics-based machine learning can address various challenges commonly occurring in the field of geosciences. We conclude the paper in Sect. 5.

2 Surrogate modeling methods

In the following, we introduce different surrogate model techniques to address the challenges presented in Sect. 4. Thereby, the focus is on inverse applications, uncertainty quantification methods, global sensitivity analyses, and other methods concerned with parameter estimation.

Surrogate models can be constructed with physics- and data-driven approaches (Fig. 1a, b). Physics-driven approaches preserve the governing equations, meaning that they, for instance, conserve mass, momentum, and energy in the same way as the original discretized version. They also maintain the characteristic that for a given set of model parameters (e.g., material properties), they produce information about the state variables (e.g., temperature, pressure) at, for instance, every node of the discretized model. But they are limited in the complexity that they can capture (Fig. 1a). We focus here on hyperbolic and/or nonlinear PDEs as these pose computational challenges in physics-based methods, as outlined above. We will not cover linear elliptic and linear parabolic PDEs for which we advise using pure physics-based approaches. These can provide error bounds in addition to the surrogate model, allowing us to obtain objective guarantees of the approximation quality. In Table 1, we list the PDEs used throughout the paper and their respective classes.

Within the physics-driven approaches, we focus on techniques that allow a retrieval of the entire state. Projection-based model order reduction techniques, such as proper orthogonal decomposition (POD), balanced truncation, and the reduced-basis method, belong to this class (Benner et al., 2015). We discuss physics-based approaches in detail in Sect. 2.1. Machine learning techniques are a common example of data-driven approaches (Baker et al., 2019; Bauer et al., 2021; Willcox et al., 2021). They have the advantage of capturing the full complexity of the model at hand but lose the information about the physical system (Fig. 1b). Examples of other data-driven techniques are kriging (Gaussian processes), polynomial chaos expansion, and response surfaces (e.g., Miao et al., 2019; Mo et al., 2019; Navarro et al., 2018). Data-driven methods are described in Sect. 2.2. The last category is physics-based machine learning, which com-

Table 1. List of partial differential equations and their respective classes used throughout the paper.

PDE	Class
Heat transfer equation	Nonlinear, hyperbolic
Heat transfer equation with shear heating	Nonlinear, hyperbolic
Time-dependent Richard's equation	Nonlinear, elliptic–parabolic

binas aspects of both physics-based and data-driven techniques to overcome the limitations of the individual approaches (Fig. 1c). A detailed overview of this technique is provided in Sect. 2.3.

2.1 Physics-based surrogate modeling and model order reduction

Model order reduction (also called reduced-order modeling) has been extensively used in various application fields, such as groundwater (Ghasemi and Gildin, 2016; Gosses et al., 2018; Rousset et al., 2014) and thermal studies (Rizzo et al., 2017). The idea is to find a low-dimensional representation of the original high-dimensional problem, while maintaining the input–output relationship. This low-dimensional representation allows a fast computation, enabling, in turn, probabilistic inverse methods and other forward-intensive algorithms. Commonly known methods of projection-based model order reduction techniques are proper orthogonal decomposition (POD) (Benner et al., 2015; Hesthaven et al., 2016; Volkwein, 2011), the reduced-basis (RB) method (Hesthaven et al., 2016; Prud'homme et al., 2002; Quarteroni et al., 2015), and balanced truncation (Antoulas, 2005). In the following, we will briefly introduce the POD and RB methods since they are relevant for the remaining case studies. For a more detailed overview of projection-based model order reduction, we refer to Benner et al. (2015).

POD is the probably most widely used projection-based model order reduction technique. For the construction of the reduced model, we first calculate snapshots. Snapshots are state distributions for different parameters (e.g., rock properties) and/or different time steps. Afterwards, we perform a singular value decomposition. The reduced basis is obtained by keeping the singular vectors corresponding to the largest singular values. The resulting orthonormal POD basis is turned into the reduced model through a Galerkin projection (Benner et al., 2015; Hesthaven et al., 2016; Volkwein, 2011). We categorize the POD as a physics-based method since through the projection the PDE is translated from the space of the spatial coordinates to the material properties. In the case that all POD modes are used this leads to an equivalent representation. The Galerkin projection needs access to the stiffness matrix and load vector and hence direct access to the discretized version of the PDE. Note that a POD can also be directly applied to, for instance, image data that have not been constructed by PDEs. However, in this case,

no Galerkin projection would be applied. In the geoscience community, the POD method has been used in, for instance, thermal applications (Rousset et al., 2014) and groundwater studies (e.g., Ghasemi and Gildin, 2016; Gosses et al., 2018; Rizzo et al., 2017; Vermeulen et al., 2004; Vermeulen and Heemink, 2006).

Another physics-based method is the RB method. The idea of RB is similar to the principle of POD. The main difference is the construction of the reduced basis. In RB the selection of the snapshots is often performed via a greedy algorithm as the sampling strategy. The greedy algorithm chooses the solution that adds the highest information content at each iteration (Bertsimas and Tsitsiklis, 1997; Veroy et al., 2003). Afterwards, the orthonormalized snapshots are added to the reduced basis and the reduced model is again obtained via a Galerkin projection (Benner et al., 2015; Hesthaven et al., 2016; Prud'homme et al., 2002; Quarteroni et al., 2015). Since this version of the RB method does not calculate the snapshots in advance but calculates them “on the fly” it performs slightly better for higher-dimensional parameter spaces (Jung, 2012). However, for the efficient construction of the reduced basis, an error bound or estimator is required. Currently, no efficient error bounds for hyperbolic PDEs exist, limiting the applicability of the method. The RB method is a physics-based method since, as described for the POD above, it is a projection method, and the Galerkin method requires the discretized version of the PDE.

A note of caution: the combination of the POD and the Galerkin projection step is sometimes referred to as POD (Swischuk et al., 2019) only and sometimes as POD-Galerkin (Busto et al., 2020).

Both the POD and the RB method have limitations when considering complex geophysical problems. Both methods aim at an affine decomposable problem, meaning that our problem is decomposable into a parameter-dependent and parameter-independent part. This requirement is naturally given for linear but not for nonlinear problems. Methods such as the empirical interpolation method (Barrault et al., 2004) exist to extend the methods for the nonlinear case, but their performance for high-order nonlinearities is limited.

Our choice to mostly rely on physics-based methods stems from the fact that understanding the physical processes is a central component in geosciences. Physics-driven methods are ideal for predictions and risk assessments, but they also have the major disadvantage of being computationally too demanding for complex coupled subsurface processes.

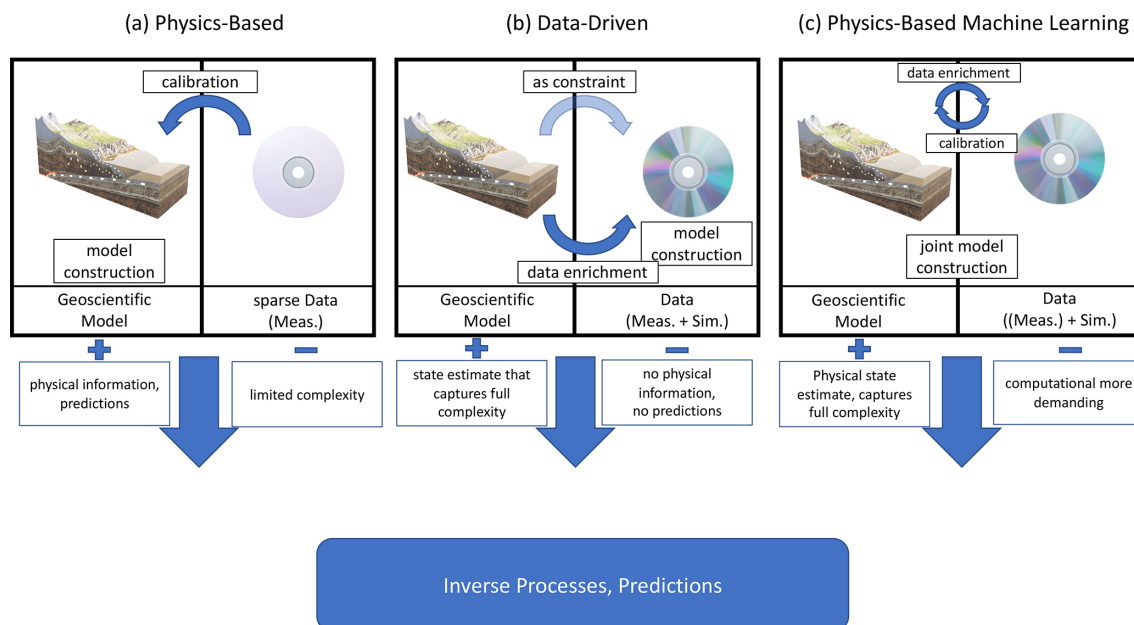


Figure 1. Schematic comparison of physics-based, data-driven, and physics-based machine learning methods for the construction of surrogate models. “Meas.” indicates measured (e.g., well data, surface observations) and “Sim.” indicates simulated. The boxes with a + list advantages, and the boxes with a – list disadvantages of the method above.

In the literature, this can often be observed indirectly. For example, in geothermal applications basin-scale applications usually consider a hydrothermal system, whereas reservoir-scale applications also incorporate the mechanical and/or chemical processes (e.g., Freymark et al., 2019; Gelet et al., 2012; Jacquy and Cacace, 2017, 2020b, a; Kohl et al., 1995; Kolditz et al., 2012; Koltzer et al., 2019; O’Sullivan et al., 2001; Poulet et al., 2014; Stefansson et al., 2020; Taron and Elsworth, 2009; Taron et al., 2009; Vallier et al., 2018; Wellmann and Reid, 2014). Another example is magnetotellurics (MT). The inversion of MT data is very costly; hence, most inversions are performed in 1D or 2D (e.g., Chen et al., 2012; Conway et al., 2018; Jones et al., 2017; Rosas-Carbajal et al., 2014). 3D MT inversion is challenging even with HPC infrastructures (Rochlitz et al., 2019; Schaa et al., 2016). How to accelerate solving the Maxwell equations through the reduced-basis method and simultaneously achieve efficient error estimates or bounds has been shown fundamentally in, e.g., Chen et al. (2010), Hammerschmidt et al. (2015), Hess and Benner (2013a), Hess and Benner (2013b), and Hess et al. (2015). The potential of the reduced-basis method is well illustrated in Manassero et al. (2020), where the authors perform an adaptive Markov chain Monte Carlo run for an MT application using the reduced-basis method.

2.2 Data-driven techniques

In applications where we have a huge amount of data (e.g., earthquake and seismic applications) and only limited knowledge of the physical processes, data-driven approaches such

as machine learning are superior to purely physics-based methods (Bergen et al., 2019; Raissi et al., 2019). Machine learning approaches are powerful in discovering low-dimensional patterns and in reducing the dimensionality of the model. Hence, they would be ideal for surrogate models. A review of the state and potential of machine learning techniques for the solid Earth community is provided in Bergen et al. (2019) focusing on data-rich applications such as the analyses of earthquakes. The applied methods in this field comprise, for instance, (deep) neural networks, random forest, decision trees, and support vector machines. However, open questions remain about the reliability, the explainability of the models, the robustness, and the rigorousness of how much the underlying assumptions have been tested and validated (Baker et al., 2019; Willcox et al., 2021). This is critical in applications such as risk assessment, where we need to provide guarantees and confidence intervals for our predictions. Another issue is that data-driven approaches such as machine learning are problematic in terms of robustness and have no proven convergence if applied to the so-called “small data” regime (Raissi et al., 2019). Raissi et al. (2019) define the small data regime as a regime where the collection of data becomes too expensive and thus conclusions have to be drawn on incomplete datasets.

This is especially critical for the application field of solid Earth, where we are classically concerned with a lack of data. Hence, this field does not allow straightforward applicability of data-driven approaches. Furthermore, in this field, we need to perform predictions and risk assessments. Therefore, black-box approaches, such as machine learning, pose a ma-

challenge in how to provide guarantees and how to obtain interpretable models (Baker et al., 2019; Willcox et al., 2021).

Some data-driven methods assume that we are interested in a quantity of interest and not in the entire state distribution. Possible quantities of interest are the value of the state variable at a certain location in space or the amount of substance that is accumulated over time in a certain region. Note that this class of methods allows a violation of the governing conservation laws since it commonly uses statistical methods for the construction of the surrogate model (Bezerra et al., 2008; Cressie, 1990; Khuri and Mukhopadhyay, 2010). Typical methods used in the field of geosciences are kriging, polynomial chaos expansion, and response surfaces (Baş and Boyacı, 2007; Bezerra et al., 2008; Frangos et al., 2010; Khuri and Mukhopadhyay, 2010; Miao et al., 2019; Mo et al., 2019; Myers et al., 2016; Navarro et al., 2018). We will not further discuss this class of methods since in this paper we focus on applications that are not only interested in a quantity of interest but also in the entire state distribution. For an overview of data-driven machine learning techniques, we refer to Jordan and Mitchell (2015), Kotsiantis et al. (2007), and Mahesh (2020).

2.3 Physics-based machine learning

Data-driven methods are a nonideal choice in the Earth sciences because they do not provide an understanding of the underlying subsurface processes (Reichstein et al., 2019). Furthermore, they require more data than physics-based methods (Willcox et al., 2021), which is relevant because some applications fields such as geodynamics and geothermal applications are characterized by data sparsity. However, physics-based methods are also not applicable for complex coupled processes since they do not capture the full complexity of the problem.

This is the reason why we introduce the concept of physics-based machine learning. Note that we discuss the potential of these methods for applications where primarily physical knowledge is available (in the form of PDEs) with very sparse to no data. The discussion would be different in a data-dominated application. Physics-based machine learning uses a combination of data-driven and physics-based techniques, as schematically shown in Fig. 1. For solving PDEs Willard et al. (2020) distinguish various techniques, comprising the categories of physics-guided loss functions, physics-guided architectures, and hybrid models. These various approaches are already compared in several papers (e.g., Faroughi et al., 2022; Swischuk et al., 2019; Willard et al., 2020). However, they focus on applications where substantially more data are available than in most subsurface applications. Therefore, we shift the focus to applications with very sparse datasets. This impacts the potential of physics-based machine learning methods differently, depending on the paradigm used to combine physics- and data-based meth-

ods. So, instead of presenting numerous of these techniques in detail, as already done in the aforementioned papers, we want to discuss the different paradigms that exist. Here, we identify two end-member cases: (i) physics-guided loss functions and (ii) hybrid models.

In physics-guided loss function methods, the idea is to use a data-driven approach and incorporate the physics in the loss function to force the system to fulfill the physical principles. They have a better performance for the small data regime since the physical constraints regularize the system and they constrain the size of admissible solutions (Raissi et al., 2019). Within the field of physics-guided loss functions various different approaches are available (e.g., Bar-Sinai et al., 2019; De Bézenac et al., 2019; Dwivedi and Srinivasan, 2020; Geneva and Zabarar, 2020; Karumuri et al., 2020; Meng and Karniadakis, 2020; Meng et al., 2020; Pang et al., 2019, 2020; Peng et al., 2020; Raissi et al., 2019; Shah et al., 2019; Sharma et al., 2018; Wu et al., 2020; Yang et al., 2018; Yang and Perdikaris, 2018; Yang et al., 2020; Zhu et al., 2019). For the following, we focus on the method of physics-informed neural networks (PINNs) as an example.

The second class of physics-based machine learning techniques that we present in more detail are hybrid models. In this technique, the physics are not added as a constraint in the loss function, but instead we perform both a physics-based and a data-driven approach. Often the physics-based method is executed first and the output serves as an input to the machine learning approach (Willard et al., 2020; Hesthaven and Ubbiali, 2018; Swischuk et al., 2019). Also, for this class of methods, various methods are available (Hesthaven and Ubbiali, 2018; Malek and Beidokhti, 2006; Swischuk et al., 2019; Wang et al., 2019); here we focus on the non-intrusive RB method.

Note that both methods act as examples to better illustrate the different concepts used for combining physics- and data-based approaches. Numerous other methods exist such as the physics-encoded neural network (PeNN) (Chen et al., 2018a; Li et al., 2020; Rao et al., 2021), physics-encoded recurrent-convolutional neural network (PeRCNN) (Rao et al., 2021), the Fourier neural operator (FNO) (Li et al., 2020), and DeepONets (Lu et al., 2021). We will not discuss these in detail but shortly explain their relations to the two paradigms, represented by PINNs and the non-intrusive RB method.

2.3.1 Physics-guided loss functions and physics-informed neural networks

PINNs overcome the shortcomings of machine learning techniques by including physics in the system. The principle is that in applications where we lack data we often have other sources of information, including domain expert knowledge and physical laws (e.g., conservation and empirical laws). By including these additional sources into machine learning the problem is regularized and the set of admissible solutions greatly reduced, yielding significantly improved perfor-

mance of the algorithm (Raissi et al., 2019). The idea of including physical knowledge is well known for Gaussian processes. However, Gaussian processes have only limited applicability for nonlinear problems (Raissi et al., 2017; Raissi and Karniadakis, 2018). Therefore, PINNs rely on neural networks which are known for their good performance in the estimation of nonlinear functions (Raissi et al., 2019).

In Fig. 2, we show the setup of a PINN schematically. The first step is to create a (deep) neural network with space and time as inputs. The network generates a model for the state variable u . Through automatic differentiation, we obtain the spatial and temporal derivatives, which, in turn, are used in the calculation of the PDE. We learn the shared parameters of the network and the PDE by minimizing a joint loss function as presented in Eq. (1) (Raissi et al., 2019):

$$\begin{aligned} \text{MSE} &= \text{MSE}_u + \text{MSE}_f, \\ \text{where } \text{MSE}_u &= \frac{1}{N_u} \sum_{i=1}^{N_u} |u(t_u^i, x_u^i) - u^i|^2, \\ \text{and } \text{MSE}_f &= \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, x_f^i)|^2. \end{aligned} \quad (1)$$

Here, MSE stands for mean squared error, $u(t, x)$ denotes the variable with the training data $\{t_u^i, x_u^i, u^i\}_{i=1}^{N_u}$, and $\{t_f^i, x_f^i\}_{i=1}^{N_f}$ represents the collocation points for our PDE $f(x, t)$. The subscript u refers to the constraints for the boundary and initial conditions and the subscript f to the PDE constrains. Consequently, the MSE_u minimizes the difference between the data u and the simulated data at the collocation points (for the boundary and initial conditions). The MSE_f enforces the structure of the PDE at selected collocation points. Through the terms $\frac{1}{N_u}$ and $\frac{1}{N_f}$, we can weight the components of the loss function depending on, for instance, the amount of available data and physical knowledge.

PINNs are frequently used due to their high flexibility and straightforward implementation. They are used as a state estimation technique since they can recover the entire state from a given set of observations. PINNs face problems in high-frequency and multiscale feature scenarios (Wang et al., 2022). Further disadvantages of PINNs are that the PDE is used as a constraint (among other constrains) and is enforced at a limited number of selected points only. Chuang and Barba (2022) point out that PINNs also have disadvantages concerning their computational efficiency. Due to the usage of automatic differentiation many residuals need to be evaluated, leading to high computational costs. Additionally, in their tested computational fluid dynamics applications, losses close to zero needed to be minimized, yielding performance and convergences issues. This problem is also expected for a wide range of geoscientific applications. Lastly, Chuang and Barba (2022) reported that PINNs did not reproduce the desired physical solution in all of their benchmark examples.

Therefore, the applicability of PINNs should be investigated prior to their use.

Numerous variations of PINNs are available, and we provide an overview of the following:

- B-PINNs (provide a built-in uncertainty quantification, Yang et al., 2020)
- fPINNs (for fractional PDEs, Pang et al., 2019)
- nPINNs (for a nonlocal universal Laplace operator, Pang et al., 2020)
- PPINNs (offer a time parallelization, Meng et al., 2020)
- sPINNs (for stochastic PDEs, Chen et al., 2019)
- XPINNs (offer an efficient space–time decomposition, Jagtap and Karniadakis, 2020)

PINNs directly provide an estimate of the state and do not preserve the input–output relationship (e.g., from model parameters to state information) of our original physical problem. Hence, we no longer have a model that takes rock and fluid properties as input and produces estimates of, for instance, the temperature or pressure as our output. Additionally, PINNs assume that observation data might be sparse but are available. In many geoscientific applications, we have an extremely sparse dataset and sometimes no direct observation data. Just as an example, if we want to investigate the formation of a sedimentary basin, we have only direct measurements for the present state. For the past, we have at best indirect measurements at hand. Consequently, we do not have enough data to utilize the PINN. In order to compensate, an enrichment of the dataset by, for instance, numerical simulations is needed. However, these numerical simulations are the results of PDEs. So, the natural question would be why we do not use a method that builds upon the principles of PDEs rather than using the PDEs as a constraint. Therefore, we do not see benefits in employing PINNs for applications that have nearly no data, which is the case for the geothermal and geodynamic applications we consider in this paper. For subsurface hydrological data the situation is better since exhaustive data for the land surface are available. However, the data are commonly of low quality (e.g., coarse, with systematic biases and measurement gaps). This yields similar problems for the PINNs as in the “no data” case. Hence, we focus in the remaining part of the paper on another physics-based machine learning technique, namely the non-intrusive RB method, which we present in the next section.

2.3.2 Hybrid methods – non-intrusive reduced-basis method

The non-intrusive RB method originates from the model order reduction community. Model order reduction and machine learning have huge similarities. For example, proper

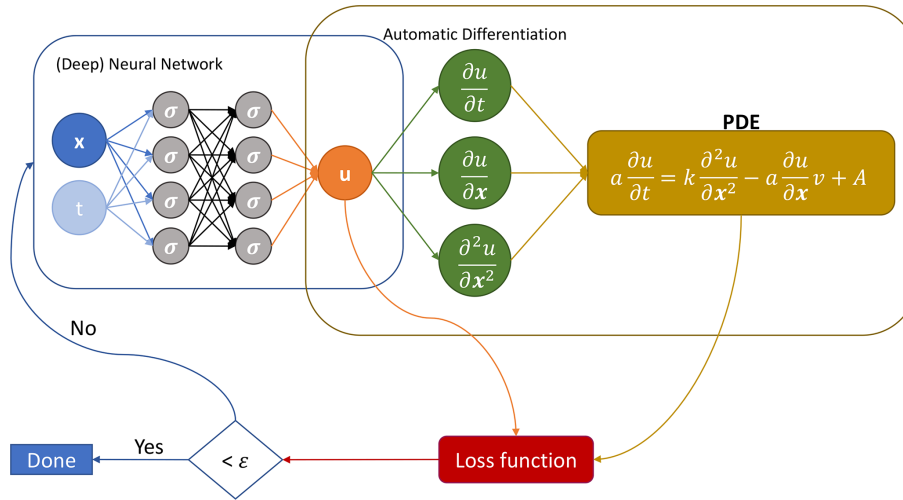


Figure 2. Schematic representation of a physics-informed neural network.

orthogonal decomposition is very closely connected to principal component analysis (Swischuk et al., 2019). Swischuk et al. (2019) pointed out that the difference between the methods has historical reasons. Model order reduction methods have been developed by the scientific computing community to reduce the high-dimensional character of typical scientific applications. In contrast, machine learning originates from the computer science community aiming at generating low-dimensional representations through black-box approaches (Swischuk et al., 2019).

The first method that we presented, physics-informed neural networks, can be seen as a machine-learning method, where we use physics to constrain our system. On the other hand, the non-intrusive RB approach can be seen as a modification of the model order reduction techniques of POD and RB using data-driven approaches. The major difference between the methods is the question of what we learn from the algorithm. For PINNs, we are learning the operators themselves. In contrast, the non-intrusive RB method derives a mapping between the input parameters (e.g., rock properties) and the output (e.g., temperature and pressure distribution) (Swischuk et al., 2019).

The idea behind model order reduction is that the reduced solution $u_L(\mathbf{x}; \boldsymbol{\mu})$ can be expressed as the sum of the product of the basis functions $\psi_l(\mathbf{x})$ and the reduced coefficients $u_{rb}^{(l)}(\boldsymbol{\mu})$. Here, \mathbf{x} defines the spatial coordinates and $\boldsymbol{\mu}$ the parameters. The procedure is described in Eq. (2) (Hesthaven and Ubbiali, 2018):

$$u_L(\mathbf{x}; \boldsymbol{\mu}) = \sum_{l=1}^L u_{rb}^{(l)}(\boldsymbol{\mu}) \psi_l(\mathbf{x}). \quad (2)$$

So, for the construction of the surrogate model, two steps are required. First the basis functions are determined, and afterwards the reduced coefficients are conventionally determined by a Galerkin projection. In Sect. 2.1, we present

two projection-based model order reduction techniques. The disadvantage of the presented methods is that they have only limited applicability to nonlinear PDEs (Hesthaven and Ubbiali, 2018; Wang et al., 2019). Furthermore, they require access to the decomposed stiffness matrix and load vector. However, this retrieval is sometimes prevented by the forward solver (i.e., in commercial software packages) (Peherstorfer and Willcox, 2016). Limitations regarding both nonlinearity and the software arise from this intrusive Galerkin projection.

To overcome the problem with the projection recent advances have replaced the intrusive projection with a non-intrusive approximation (Hesthaven and Ubbiali, 2018; Wang et al., 2019). Hesthaven and Ubbiali (2018) and Wang et al. (2019) propose a (deep) neural network for learning the reduced coefficient. Other approaches including interpolation and machine learning methods have been explored for this purpose (Swischuk et al., 2019). Swischuk et al. (2019) compare four machine learning methods, namely neural networks, multivariate polynomial regression, k -nearest neighbors, and decision trees, for the determination of the reduced coefficients in aerodynamic and structural applications. Audouze et al. (2009, 2013) and Wirtz et al. (2015) use Gaussian process regression to determine the reduced coefficients. Another strategy is proposed in Mainini and Willcox (2015), where the reduced coefficients are learned by an adaptive mixture of self-organizing maps and response surfaces.

Due to the different methods for obtaining the reduced coefficients, the commonly used greedy sampling strategy of the intrusive RB method (Hesthaven et al., 2016; Veroy et al., 2003) has to be modified. For the selection of basis functions, we can either use a POD (Hesthaven and Ubbiali, 2018; Wang et al., 2019) or a modified greedy sampling strategy (Chen et al., 2018b). In the case of the POD, which is used in this paper, the snapshots (i.e., simulation results for

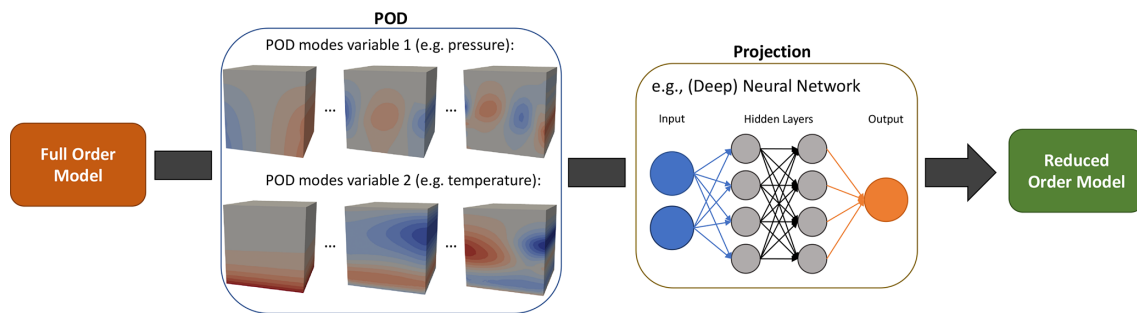


Figure 3. Schematic representation of the non-intrusive reduced-basis method with a proper orthogonal decomposition sampling strategy.

different model parameters) serve as an input and the basis functions as an output. For the second stage, the projection is via a machine learning method. The labels are the different model parameters (e.g., permeability, thermal conductivity), the training data the matrix product of the snapshots and the basis functions, and the outputs the reduced coefficients. Taking the matrix product of snapshots and basis functions means that the dimension of spatial degrees of freedom does not enter into the training of the machine learning model, which has great implications regarding the cost of training these models, as we will detail in Sect. 3.5.

Note that non-intrusive model order reduction is not restricted to the inference of the output states via the input parameters. As an example, Peherstorfer and Willcox (2016) developed a data-driven operator inference non-intrusive method. However, this method is restricted to low-dimensional nonlinearities for computational reasons (Peherstorfer and Willcox, 2016). As we aim to demonstrate the potential of physics-based machine learning, we chose as an example two techniques: one closer to the community of computer science and the other closer to the community of scientific computing. We do not aim to provide an overview of all possible methods.

Again, we add a note of caution with regard to the terminology. The non-intrusive RB method consists of a sampling and projection step. If we use, for instance, the POD as the sampling method and the NN as the projection method, then this is often referred to as POD-NN (Hesthaven and Ubiali, 2018) instead of non-intrusive RB. In this paper, we use various machine learning techniques for the projection step. Since this has no general impact on the applicability of physics-based machine learning we use the more general term of non-intrusive RB to avoid unnecessary confusion.

2.3.3 Other physics-based machine learning methods

As discussed before, many physics-based machine learning methods are available, all with their own advantages and disadvantages. Nonetheless, we can identify generally different approaches regarding how to integrate physics into data-driven approaches; see also Willcox et al. (2021). These diverse approaches yield different potential for addressing cur-

rent challenges faced in subsurface applications, as we will present in Sect. 4. So, in the following, we briefly present other physics-based machine learning methods and how they relate to the two paradigms presented in detail.

The idea behind PeNNs is to hard-encode the prior physical knowledge in the architecture of neural networks. This can be achieved through various methods (Chen et al., 2018a; Li et al., 2020; Rao et al., 2021). One possibility is the PeRCNN, where an optional physics convolution layer is implemented. This means that the method is in between PINNs and the non-intrusive RB method. It is able to overcome some common problems encountered with PINNs since it does not implement physics as one of many possible constraints. However, the method is still a statistical interpolation method and therefore cannot guarantee a preservation of the physics. Regarding the categories that are used in Willard et al. (2020), this method would fall under the category of physics-guided architectures.

Another method that is associated with PeNNs involves FNOs (Faroughi et al., 2022; Li et al., 2020). This method uses Fourier layers instead of the typical layers used in a neural network. Conceptually this is close to the non-intrusive RB method. In the Fourier layers the data are decomposed, similar to the POD step of the non-intrusive RB method. However, the main difference is that FNO is limited to sine and cosine functions. Neural operators are known to require large training datasets (Lu et al., 2022), which is critical in the applications presented here since the generation of the data is often very costly. Furthermore, they are limited when it comes to complex 3D problems (Lu et al., 2022). Another difference to the non-intrusive RB method is that again the approach is embedded into a data-driven scheme. Something similar is the case for DeepONets (Lu et al., 2021), which overcome the requirement of structured data that FNOs have but show conceptually similar limitations as FNOs.

2.3.4 Differences in physics-based machine learning

We provide examples from the field of hydrology and water resource management since machine learning is currently actively being studied in these research fields (Shen et al., 2021; Sit et al., 2020). This is fueled by the promise of ML to un-

ravel information and provide hydrological insights (Nearing et al., 2021) in what can be considered increasingly rich(er) datasets, especially for surface hydrology (in comparison to subsurface hydrology and other subsurface geoscientific applications, such as geodynamics). Typical applications of ML include runoff and flood forecasting, drought forecasting, water quality, and subsurface flows, among many others (see Ardabili et al., 2020; Bentivoglio et al., 2021; Mohammadi, 2021; Sit et al., 2020; Zounemat-Kermnai et al., 2021, for extended reviews on the topic). Many of these applications are based on artificial neural networks (ANNs) and deep learning tool sets which have become accessible to hydrologists thanks to the arrival of well-established and GPU-enabled libraries. It is currently clear that ML offers a wide range of possibilities in hydrology, such as faster-than-real-time forecasting, extracting information from citizen data, parameter inversion, and uncertainty quantification. The idea of ML surrogates is increasingly being exploited in hydrological research (e.g., Bermúdez et al., 2018; Liang et al., 2019; Tran et al., 2021; Zahura et al., 2020; Zanchetta and Coulibaly, 2022).

Physics-based ML is still rare and incipient in hydrology, with efforts largely motivated by overcoming the black-box nature of traditional ML techniques (Jiang et al., 2020; Young et al., 2017) and the need to include physical constraints and theory (Nearing et al., 2021). The techniques to inform ML of the physics, as well as how much of the physics is embedded into the problem, depend on the application and the underlying deterministic model, ranging from ordinary partial differential equations (ODEs) (Jiang et al., 2020) to classical lumped cascade hydrological models (Young et al., 2017), subsurface hydraulic parameter estimation based on nonlinear diffusion equations (Tartakovsky et al., 2020), the Darcy equation (He et al., 2020) and Richards equation (Bandai and Ghezzehei, 2021), and overland flow (Maxwell et al., 2021). It is therefore clear that there is growing momentum and clear potential gains in exploring physics-based ML in hydrological sciences, motivated by methodological, theoretical, and computational factors.

The term physics-based machine learning is used in an increasing number of scientific works; however, the methodologies employed vastly differ. Therefore, we would like to briefly elaborate on the differences between these techniques. Note that this explanation is purely meant to illustrate the usage of the term physics-based machine learning throughout the remaining part of this paper. To our knowledge, there is no unique definition of that term. Furthermore, we will only illustrate the differences between the methodologies presented here. We do not aim to provide a complete overview of all physics-based ML techniques but rather illustrate the various concept associated with the technique in general.

What might be sometimes defined as physics-informed or physics-based approaches are applications whereby the surrogate model is constructed with classical data-driven ML

techniques but the data originate from a physical model instead of observation data. In this paper, the definition of physics-based ML is purely based on the methodology. Since the former application is methodology-wise the same as a data-driven approach this would be defined here as a data-driven and not a physics-based ML technique.

The two physics-based ML techniques presented here are PINNs and the non-intrusive RB method. Both methods have similar aims but are vastly different in where and how ML techniques and physics are utilized. PINNs use the physics as an additional constraint in the loss function and do not preserve the original structure of the problem. In contrast, the non-intrusive RB method extracts the main physical characteristics and weights them using an ML technique, while maintaining the input–output relationship.

3 Benchmark studies

Combining both data- and physics-driven techniques allows for robust and reliable (surrogate) models that enable probabilistic inverse processes, which yield a significant improvement in the understanding of the Earth system. For an illustration of the potential that physics-based machine learning has for the community, we present in the following three benchmark studies from the field of geothermal, geodynamical, and hydrological applications. Afterwards, we use these studies to illustrate how physics-based machine learning can help to overcome the challenges introduced at the beginning of the paper.

3.1 General workflow

Before we present three case studies in the next section, we present a general workflow (Fig. 4) and point out some common pitfalls for the data-driven part of the methodology. Although the procedure is developed with the projection part of the non-intrusive RB method in mind, it applies to a wide range of data-driven approaches. For the following workflow, we assume that a geometrical and physical model is already available. Hence, we assume that the model is already tested for single forward simulations.

1. At the beginning of the procedure, we need to define the training and validation set. For its generation, several aspects have to be taken into account.
 - a. The sampling strategy is crucial to ensure a small and reliable training set. We recommend using the Latin hypercube sampling method. Regular sampling schemes with equal spacings should be used with caution. They bear the danger of always having the same ratio between the various input parameters. In this case, the predictions will only be accurate if that ratio is maintained. To ensure that the sampling method is not negatively impacting the

prediction, we strongly recommend using a different sampling strategy for the validation and training dataset. In our studies, we use randomly chosen parameter combinations for the validation dataset.

- b. Another important aspect is the size of the training and validation set. The size of the training dataset cannot be determined in general since it is dependent on the underlying complexity of the parameter space. However, as a control check, the training dataset needs to be significantly larger than the number of basis functions obtained via the POD method. The size of the validation dataset should be at least 10 % of the training dataset.
2. After generating the training and validation dataset and before applying the physics-based machine learning method, it is important to preprocess the data. The preprocessing involves the following steps.
 - a. The first is the scaling of the input parameters. Common scaling methods are the normal and the min–max transform. In this study, we use normal transformed parameters. This means that the mean of each of the input parameters is zero and that they have a standard deviation of 1.
 - b. Also crucial is the preprocessing of the actual training and validation data. This consists of two steps.
 - i. The first step is the removal of static effects such as the initial conditions. As an example, we often have an initial temperature distribution that is a magnitude larger than the changes induced by the varying rock properties. In order to capture the changes induced by the rock properties it is important to remove this “masking” effect.
 - ii. The other important step is the scaling of the data. Again, common scaling methods are the normal and the min–max transform. In this paper, we employed, where necessary, the min–max transform.
 3. After preparing the data, we can perform the physics-based machine learning method. When using methods such as neural networks, the search for optimal hyperparameters (parameters that determine the architecture of the machine learning model) can be extremely time-consuming. This is a cost that should not be underestimated since it can prolong the construction time of the surrogate model significantly, which majorly impacts the efficiency of the concept. To considerably speed up the hyperparameter search, we advise using a Bayesian optimization scheme with hyperbands (BOHB) (Falkner et al., 2018). The idea behind BOHB is to accelerate the convergence rate of the Bayesian optimization for the tuning of the hyperparameters with

the hyperbands. Furthermore, the methodology yields a scalable approach suited for parallel computing. For further details, we refer to Falkner et al. (2018).

3.2 Geothermal example

For the geothermal benchmark study, we consider a three-dimensional model with three equally spaced horizontal layers (Fig. 5). The problem is nondimensional, with an extent of 1 in the x , y , and z direction. The spatial mesh is a tetrahedral mesh with 11 269 nodes.

As the physical model, we take a fully coupled hydrothermal simulation into account, for which we consider the heat transfer equation (Nield and Bejan, 2017):

$$\frac{\partial ((\rho c)_m T)}{\partial t} = \nabla \cdot (\lambda \nabla T - (\rho c T)_f v) + H,$$

$$\text{with } (\rho c)_m = (1 - \phi) \rho_s c_s + \phi \rho_f c_f,$$

$$\text{and } \lambda_m = (1 - \phi) \lambda_s + \phi \lambda_f. \quad (3)$$

Here, ρ is the density, c the specific heat capacity, T the temperature, t the time, λ the thermal conductivity, H the radiogenic heat production, and ϕ the porosity. The subscripts m, s, and f stand for the porous medium, the solid properties, and the fluid properties, respectively. For the fluid discharge v , we consider Darcy’s law (Nield and Bejan, 2017):

$$v = -\frac{k}{\mu} (\nabla p + \rho_f g \nabla z), \quad (4)$$

where k denotes the permeability, μ the viscosity, p the pressure, g the gravity acceleration, and z the vertical coordinate component. In this section, we consider natural convection resulting in a temperature-dependent density of the following form (Nield and Bejan, 2017):

$$\rho_f = \rho_{\text{ref}} (1 - \alpha (T - T_{\text{ref}})), \quad (5)$$

where the subscript ref denotes the reference properties and α the thermal expansion coefficient. Furthermore, we consider the Boussinesq assumption (Nield and Bejan, 2017):

$$B = \alpha \Delta T. \quad (6)$$

Note that we consider the nondimensional representation to investigate the relative importance of the material properties and to improve the efficiency. This means all properties and variables in the following are presented in their dimensionless form. Since nondimensionalization is not the focus of this paper, we only present the nondimensional equations themselves and not their derivations. For a detailed derivation of the nondimensional equations refer to Degen (2020) and Huang (2018). The nondimensional heat and fluid flow

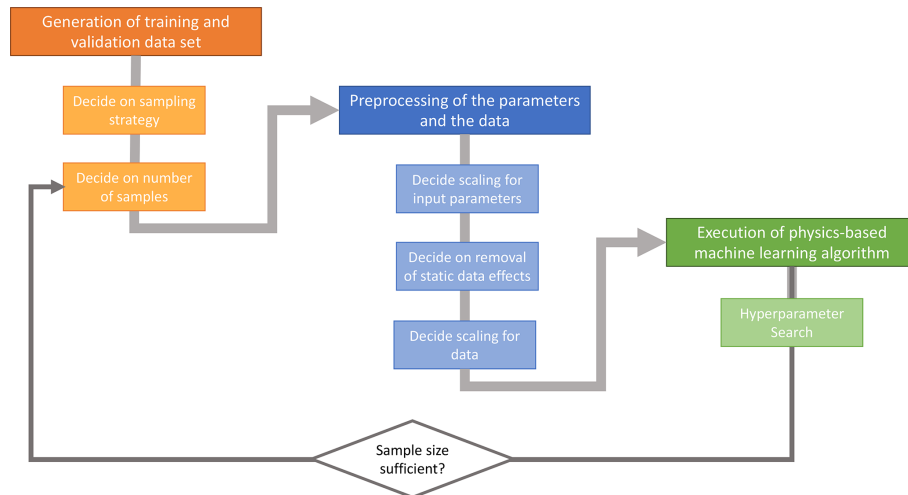


Figure 4. Schematic representation of the workflow to efficiently generate surrogate models by using physics-based machine learning techniques.

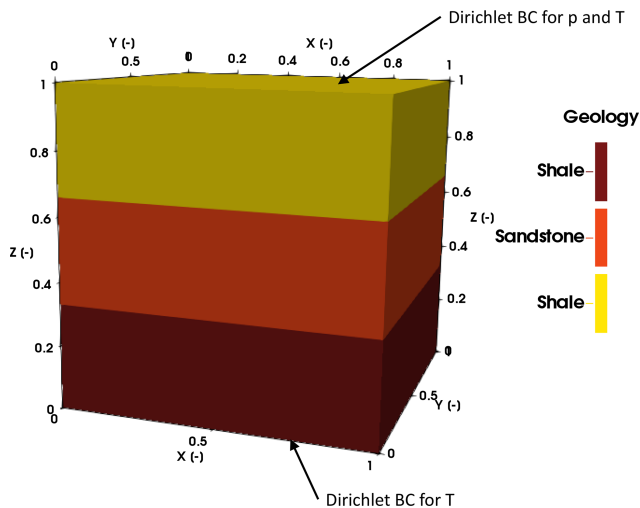


Figure 5. Representation of the geological model used for the geothermal benchmark study. In addition to the geological layers the boundary conditions are also indicated.

can be expressed as

$$\begin{aligned}
 \nabla \cdot \mathbf{q}^* &= 0, \\
 \mathbf{q}^* &= -\nabla^* p^* + \sqrt{Ra} T^* \hat{\mathbf{j}}, \\
 \frac{\partial T^*}{\partial t} + \sqrt{Ra} \mathbf{q}^* \cdot \nabla^* T^* - \nabla^{*2} T^* &= 0, \\
 \text{with } Ra &= \frac{\rho_{\text{ref}} g \alpha k l_{\text{ref}} \Delta T}{\mu \kappa}, \\
 \text{and } \kappa &= \frac{\lambda_m}{\rho_{\text{ref}} c_f}.
 \end{aligned} \tag{7}$$

Here, the asterisk denotes the nondimensional variables, $\hat{\mathbf{j}}$ the unit vector, and l_{ref} the reference length.

For the numerical modeling, we use the finite-element software DwarfElephant (Degen et al., 2020a, b), which is based on the MOOSE framework (Permann et al., 2020).

The pressure distribution has a Dirichlet boundary condition of zero at the top and the remaining boundary conditions are no-flow boundaries. For the temperature, we apply Dirichlet boundary conditions at the top and the base of the model, where the top boundary condition has a value of zero and the base one has a value of 1. The remaining boundaries have zero Neumann boundary conditions. For the time stepping, we consider an equal spacing between 0 and 0.8, resulting in 15 time steps. Note that the time is dimensionless. The initial condition is zero everywhere in the model domain.

During the study, we assume that the top and base layers are low-permeability layers with a variation range of 1 to 4 for the Rayleigh number (Ra) and the middle layer is a highly permeable layer with a variation range of 36 to 64 for Ra . For these parameter ranges, we construct two training datasets: one with 50 samples and one with 100 samples. Throughout the entire paper, we use a Latin hypercube sampling strategy for the training datasets and a random sampling strategy for the validation dataset. In the case of the geothermal example, we have a validation dataset consisting of 20 samples.

For this benchmark example, we investigate how to construct reliable surrogate models for the pressure and temperature distribution resulting from the nonlinear and hyperbolic PDE (Eq. 3). We employ the non-intrusive RB method with a Gaussian process regression (GPR) as the machine learning method during the projection step since the variations tested here are linear. We use the GPR method during the projection step since it is known to perform well for linear variations and has fewer hyperparameters that need to be determined, yielding a more efficient construction of the surrogate model.

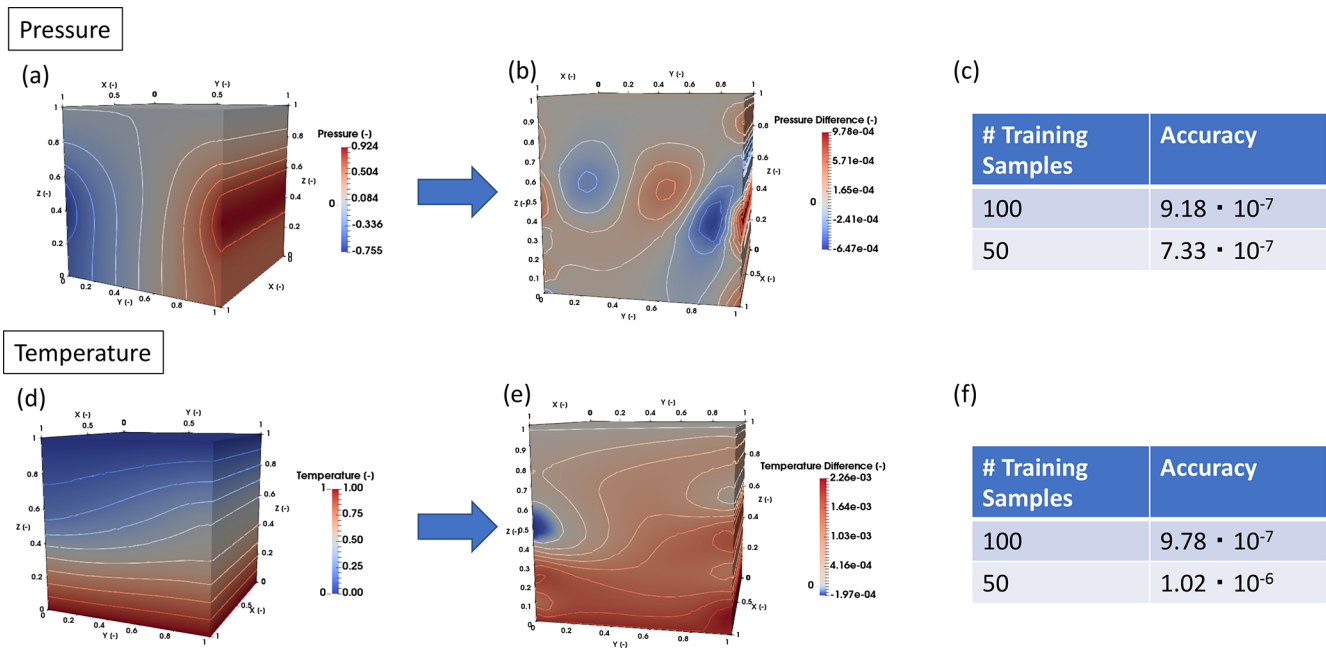


Figure 6. Representation of the accuracy for the pressure and temperature distribution of the geothermal benchmark study. The top three panels show the model for the pressure response; panel (a) contains the original pressure distribution of the finite-element model, panel (b) shows the error between the reduced and full model, and panel (c) lists the accuracy of the reduced pressure response model for different training sample sizes. Analogously, the three bottom panels display the temperature response; panel (d) shows the temperature response of the finite-element model, panel (e) illustrates the error between the reduced and full model, and panel (f) displays the accuracy of the reduced temperature response model for varying training sample sizes.

We start the discussion with the surrogate model for the pressure distribution (top three panels of Fig. 6). The pressure distribution for the full finite-element model is shown in Fig. 6a. As mentioned in the description of the non-intrusive RB method, the method consists of two steps: the POD step (where we determine the basis functions) and the projection step (where the weights are determined). Overall, we desire to achieve an accuracy of 10^{-3} ; this accuracy is chosen with respect to typical errors of pressure measurements. To reach this accuracy, we require five basis functions, which are partly shown in Fig. 3. The error distribution for one randomly chosen sample from the validation dataset is shown in Fig. 6b. Here, we can see that the error distributions follow the pattern of the higher-order modes obtained by the POD method. This is in accordance with our expectations since we truncated the basis after reaching the desired accuracy. Consequently, we lose information about the higher-order modes. In Fig. 6c, we compare the accuracies for the training datasets with 100 and 50 samples, and we reach our desired accuracy for 50 simulations.

Similar results are obtained for the temperature distribution (Fig. 6d), where we require seven basis functions to obtain an accuracy of 10^{-3} . Note that in this example, we chose the same tolerances for the pressure and temperature model. However, they do not need to be the same. Hence, the tolerances can be adjusted to the varying measurement

accuracies of, for instance, pressure and temperature measurements. Also, for the temperature model we observe that the errors (Fig. 6e) follow the distribution of the higher-order POD modes and that 50 simulations in the training set are sufficient to reach the desired tolerance (Fig. 6f).

The benefits of the method become apparent when we have a closer look at the computation times. The finite-element model requires 85 s to solve for the pressure and temperature for all time steps. On the other hand, the reduced model requires 1 ms for solving either the pressure or the temperature for a single time step. Note that in the case of the reduced model, we can solve for individual time steps and state variables and do not need to solve for the entire time period if we are, for instance, only interested in the final time step. But even if we want to solve for both variables for all time steps, we are nearly 3 orders of magnitude faster with the reduced model. From previous studies using the intrusive RB method, this speed-up is expected to be higher for real case models because the complexity of these models increases more slowly than the degrees of freedom (Degen et al., 2020a; Degen and Cacace, 2021; Degen et al., 2021b).

3.3 Geodynamic example

For the numerical modeling of the geodynamic benchmark, we use the thermomechanical finite-element code SULEC v.

6.1 (Buiter and Ellis, 2012; Ellis et al., 2011; Naliboff et al., 2017). The software solves the conservation of momentum,

$$\nabla \cdot \sigma' - \nabla p + \rho g = 0, \quad (8)$$

and the incompressible continuity equation,

$$\nabla \cdot v = 0, \quad (9)$$

where σ' is the deviatoric stress tensor, p is the pressure, ρ is the density, g is the gravitational acceleration, and v is the velocity. The density variations are described analogous to the geothermal example (Eq. 5). Also, the heat transfer is similar to the geothermal description (Eq. 3), with the main difference that we obtain shear heating as an additional source term:

$$\rho c \left(\frac{\partial T}{\partial t} + v \cdot \nabla T \right) = \nabla \cdot \lambda \nabla T + H + 2\sigma'_2 \cdot \dot{\epsilon}'_2. \quad (10)$$

Here, σ'_2 and $\dot{\epsilon}'_2$ are the second invariant of the deviatoric stress and strain tensors, respectively.

For this study, we consider the development of a salt dome (Fig. 7). The linear viscous salt rises buoyantly upward through linear viscous sediment layers. The rise is initiated by a sinusoidal perturbation of the salt–sediment interface with a 40 m amplitude. Note that all Neumann boundary conditions (Fig. 7) are no-flow boundary conditions. All velocity Dirichlet boundary conditions have a value of 0 m s^{-1} . Furthermore, the temperature at the top of the model has a fixed value of 0°C and at the base a value of 140°C . The discharge temperature Dirichlet boundary condition at the lateral sides has a value of $0 \text{ m}^3 \text{ s}^{-1}$. The model extends 8 km in the x direction and 4 km in the y direction. For the spatial discretization of the 2D benchmark study, we use quadratic cells with a resolution of $40 \text{ m} \times 40 \text{ m}$, resulting in a structured mesh with 20 301 nodes. SULEC solves for velocities on the nodes, whereas pressure is constant over the cells and obtained in Uzawa iterations (Pelletier et al., 1989).

For the time stepping, we use fixed time steps of 5 kyr and simulated from 0 to 6.25 Ma. In contrast to the previous example, we consider the dimensional case here. Nondimensional forward simulations are more efficient for physics-based machine learning methods. However, many simulations in the field of geosciences are often conducted in dimensional form. Hence, we want to illustrate how these simulations need to be processed to ensure an efficient construction of the surrogate models. Note that we describe the general workflow in Sect. 3.1; we will only briefly highlight the most important steps for the geodynamic case study here.

For the given example, we allow a variation of the thermal conductivity of the sediments between 2.5 and $3.0 \text{ W m}^{-1} \text{ K}^{-1}$ and for the salt between 5.0 and $8.0 \text{ W m}^{-1} \text{ K}^{-1}$. This means we assume that the thermal conductivity of the salt has higher uncertainty than that of the sediments. We construct a training set of 50 samples with a

Latin hypercube sampling strategy, and the validation dataset consists of 10 randomly chosen samples. The remaining model parameters stay constant throughout all simulations and are displayed in Table 2. We consider linear viscous materials that are tracked with tracers. For the benchmark example, we have in total 182 709 tracers at the start, yielding a minimum of 16 and a maximum of 20 tracers per cell (which are maintained by a tracer population control). The tracers are moved using a second-order Runge–Kutta scheme.

In Fig. 8, we present the result of the finite-element forward model and the accuracy of the surrogate model. Figure 8a shows the temperature distribution of the full model ranging from 0 to 140°C . We observe that the changes induced by the variations of the thermal conductivities (at the interface between the salt dome and the sediments) are much smaller than the initial temperature distribution. Furthermore, the initial temperature distribution is independent of any variations in the thermal conductivity. Hence it is the same for all forward simulations. Therefore, we train the model only on the temperature variations, which we additionally scale between 0 and 1 by calculating the minimum and maximum temperature values of the entire training dataset. The scaled temperature distribution used within the algorithm is shown in Fig. 8b. If visualizations of the absolute temperature distribution are desired they are obtained by adding the initial condition after calculating the reduced forward model.

Figure 8c displays the difference between the full and reduced model, where the highest errors are of the order of 0.1°C . This means that the errors induced by the approximation are lower than those of typical temperature measurements. We observe the highest errors at the flanks, in the lower part of the model. This matches our expectations since we have the largest changes at the interface between the salt diapir and the surrounding sediments. Furthermore, the temperature distribution at the flanks is more diffusive in contrast to the very sharp changes at the top of the salt dome, which adds to the higher errors of the reduced model at these locations. In total, we required 18 basis functions to reach the given accuracy. Note that the number of basis functions is relatively high because of the rather abrupt changes that we need to capture.

Analogously to the geothermal example, we evaluate the gain in computation time. The average computation time over 50 finite-element simulations is 1.05 h (including input–output processes) on two Intel Xeon Platinum 8160 CPUs (24 cores, 2.1 GHz, 192 GB of RAM) using the PARallel DIrect SOLver (PARDISO) (Schenk and Gärtner, 2004). In contrast, the average computation time over 50 reduced simulations is about 2 ms for a single time step. Note that with the reduced model, we can perform predictions for individual time steps. That this is of general interest can be seen directly in this case study. Although we calculate the solution for all 1250 time steps, we output 14 only. Note that only these 14 time steps have been used to construct the surrogate model.

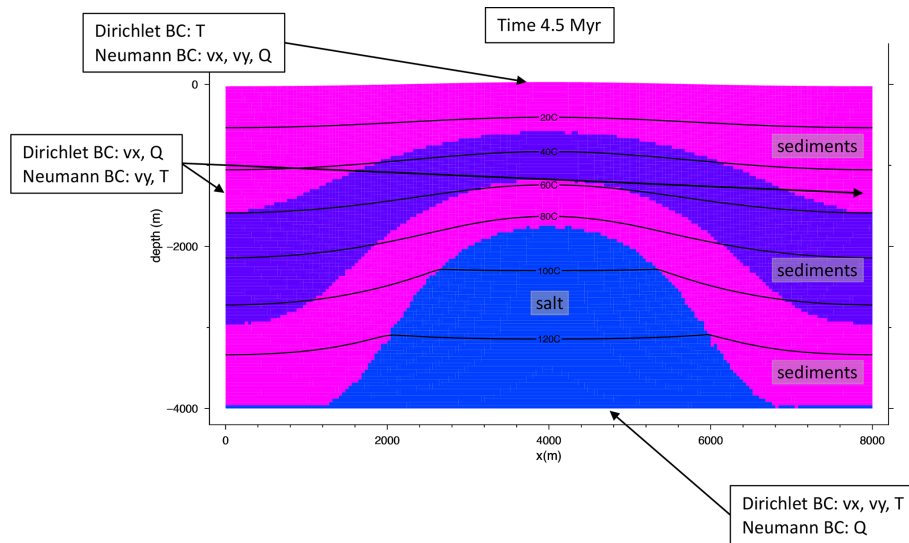


Figure 7. Geological model of the salt diapir at 4.5 Myr. The temperature isolines are plotted for a realization from the training dataset.

Table 2. Model parameters of the geodynamic benchmark example.

Parameter	Value	
	Sediments	Salt
Density	2300 kg m ³	2160 kg m ³
Viscosity	3.00 × 10 ¹⁹ Pa s	2.67 × 10 ¹⁸ Pa s
Gravitational acceleration	−9.81 m s ^{−2} (in y direction only)	
Specific heat capacity	833 J kg ^{−1} K ^{−1}	
Radiogenic heat production	0 W m ^{−3}	
Thermal expansion coefficient	0 K ^{−1}	

However, the model is trained for spatial and temporal variations. Hence, we can also retrieve intermediate time steps. The obtained speed-ups range between 5 orders of magnitude (interested in all 14 display time steps) and 6 orders of magnitude (interested in a single time step).

To construct the surrogate model, we require in total 18 basis functions to reach our POD tolerance of 10^{−3}, as displayed in Fig. 9. Having a closer look at the basis functions, we observe that the “low-frequency” information is presented in the first basis functions and the “high-frequency” information in the last basis functions. This nicely shows the analogy between the non-intrusive RB method and the Fourier decomposition.

For the determination of the hyperparameters, we use Bayesian optimization with hyperbands (as described in Sect. 3.1). To save memory, the training was performed using only every second node. All results and accuracies presented here are calculated using all nodes in space. The hyperparameters obtained after optimization are presented in Table 3. With these hyperparameters, we obtain an error of 2.21 × 10^{−6} for the scaled training dataset and an error of 1.99 × 10^{−6} for the scaled validation dataset.

3.4 Hydrological example

In this section, we present a proof-of-concept hydrological application inspired by a well-known two-dimensional infiltration problem on a domain with soil heterogeneity, originally proposed by Forsyth et al. (1995). The setup is illustrated in Fig. 10. The domain has impervious lateral and bottom boundaries and four zones with different soils, represented through different permeabilities. A uniform initial pressure of $h = -100$ cm was set and an infiltration flux q is prescribed on the top left corner of the domain.

Although this is evidently a benchmark problem, it offers many of the complexities which arise in subsurface flow simulation, such as heterogeneity and sharp infiltration fronts, and therefore nicely serves as a proof of concept.

The physical model for variably saturated flow in a porous medium is the Richards equation (Richards, 1931):

$$\frac{\partial \theta}{\partial t} = \nabla [\mathbf{K}(h) \nabla (h + z)], \tag{11}$$

where h is pressure, θ is the volumetric soil water content, K is hydraulic conductivity, t is the time, and z is the vertical

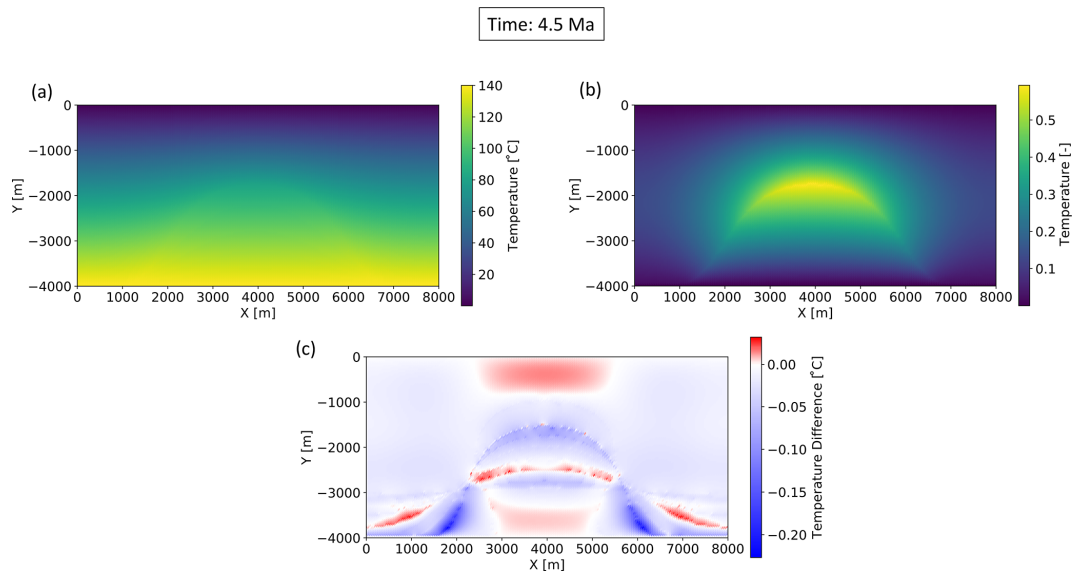


Figure 8. Comparison of the full finite-element and the reduced model. (a) The temperature distribution of the finite-element solution, (b) the scaled finite-element solution as utilized in the training of the reduced model, and (c) the difference between the reduced and full solution.

Table 3. Hyperparameters of the neural network for the geodynamic benchmark example. Note that hl denotes hidden layers.

Hyperparameter	Value
Number of hidden layers	3
Number of neurons per hidden layer	427 (hl 1), 383 (hl 2), 313 (hl 3)
Number of epochs	22 438
Learning rate	4.68×10^{-4}
Batch size	216
Loss function	Sigmoid
Optimizer	Adam

coordinate. A soil model is necessary to provide closure relationships, for which, for this case, we use the Mualem model for the water retention curve $\theta(h)$ and the van Genuchten model for $K(\theta(h))$ (van Genuchten, 1980). The resulting mathematical model is a highly nonlinear diffusion equation, formally an elliptic–parabolic PDE, which also shows very sharp front propagation akin to that of advection problems (Caviedes-Voullième et al., 2013).

To numerically solve Eq. (11) we rely on the well-established hydrological model ParFlow (Kuffour et al., 2020). ParFlow solves the Richards equation via a backward Euler finite-difference scheme and uses multigrid-preconditioned Newton–Krylov methods to solve the resulting system. It is massively parallelized and, in fact, recently ported to GPUs (Hokkanen et al., 2021). For the infiltration problem here, we do not leverage ParFlow’s HPC capabilities, since the time to solution of this problem is only a few seconds on a single CPU core. In contrast, we ran all realizations of both training sets in parallel in an HPC node in the JUWELS system at the Jülich Supercomputing Centre.

In this proof-of-concept exercise, we first construct a training set for the non-intrusive RB method. A total of 100 different combinations of permeabilities for the four materials in the domain are considered, as well as various inflow rates q . A Latin hypercube strategy was used for sampling the five-dimensional parameter space. A validation set of 20 samples (with a random sampling strategy) was also generated. Importantly, permeability values follow a lognormal distribution, whereas the infiltration flux follows a uniform distribution. The original permeabilities proposed by Forsyth et al. (1995) were selected as the mean value for the lognormal distributions with a standard deviation of 1. The uniform distribution for the infiltration flux was centered around 4 cm d^{-1} and a standard deviation of 0.5.

Similar to the previous case studies, we use the non-intrusive RB method to construct a surrogate model that predicts the hydraulic head for every parameter combination in the above-defined ranges at every time step. For the construction of the surrogate model, we require 124 basis functions. This number is significantly larger than in the previous examples. This is related to the more pronounced nonlinearity

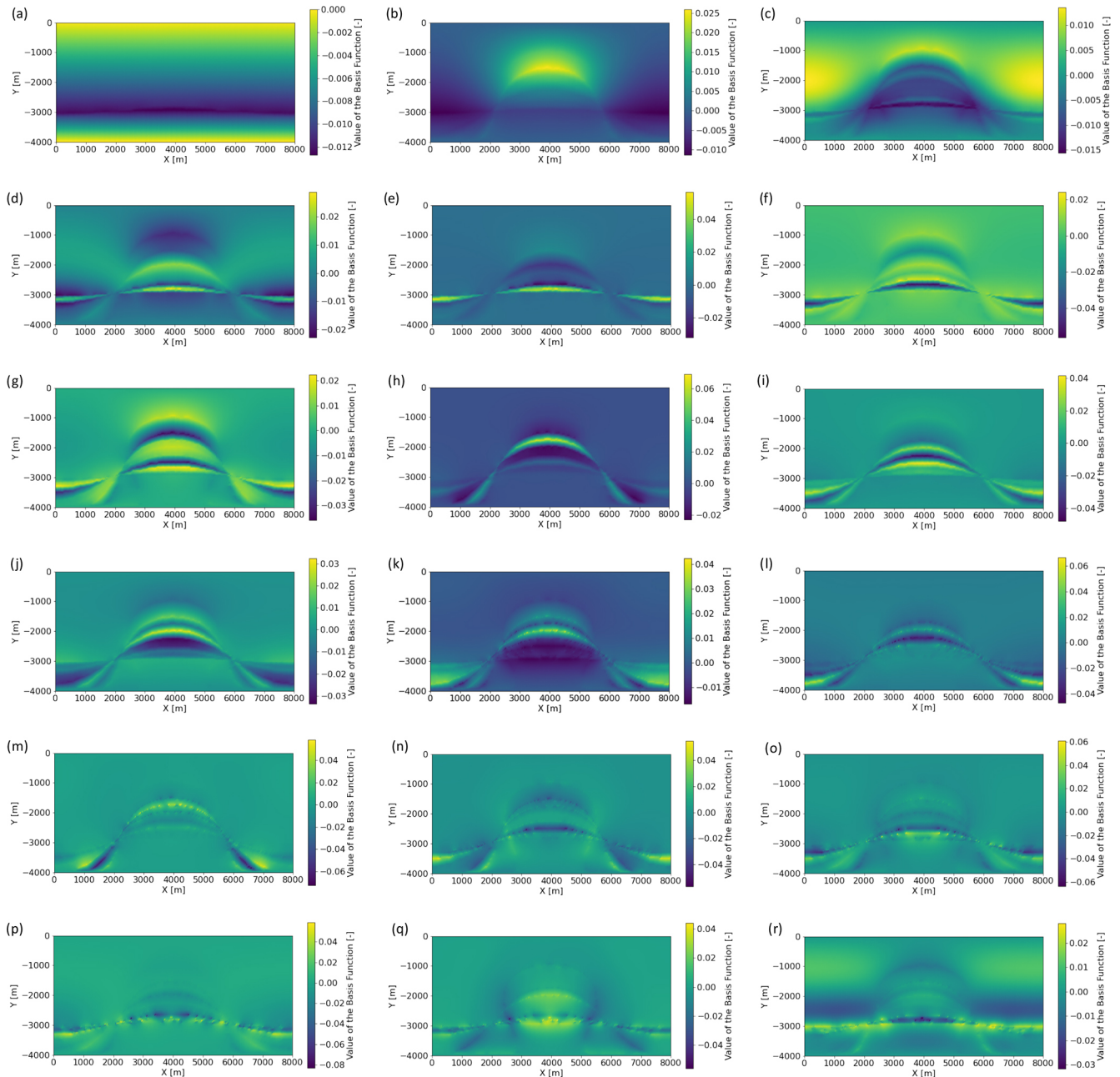


Figure 9. Illustration of the 18 basis functions of the surrogate model, where (a) shows the first basis function and (r) the last basis function.

of the solutions, which differs for the various time periods. Nonetheless, we obtain solutions in less than 3 ms (calculated over 100 iterations), yielding a speed-up of 3 orders of magnitude (if we want to predict the solution for a single time step). With the hyperparameters presented in Table 4, we achieve an accuracy of 8.71×10^{-7} for the scaled training set and $2.89 \cdot 10^{-5}$ for the validation dataset.

The errors for the training and validation dataset are global errors. In the next step, we look at the spatial distribution of the approximation errors. Therefore, we pick an arbitrary so-

lution from the validation dataset and compare the solution for the last time step (Fig. 11). Figure 11a and b look visually the same, which is the reason we focus the presentation on the difference plot displayed in Fig. 11c. Here, we observe that no part of the model is underrepresented and that the errors are the highest in areas where there is strong spatial heterogeneity of the permeability fields and of the infiltration flux (most notably around zone 4), both of which contribute to strong nonlinear gradients in hydraulic head. The magni-

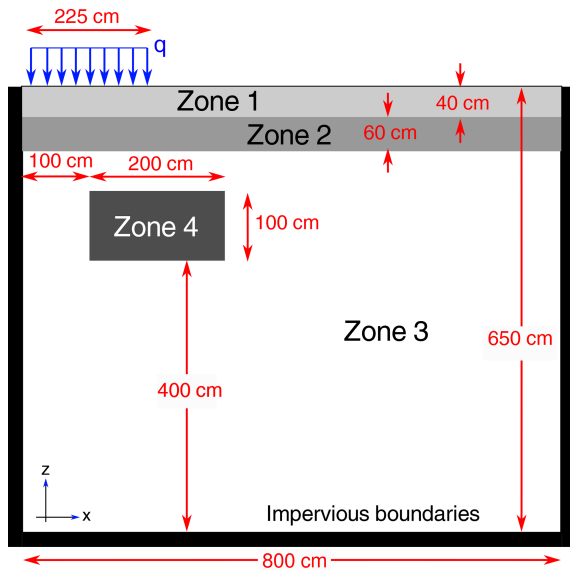


Figure 10. Sketch of the infiltration problem for the proof-of-concept hydrological application. Adapted from Forsyth et al. (1995).

Table 4. Hyperparameters of the neural network for the hydrological benchmark example. Note that hl denotes hidden layers.

Hyperparameter	Value
Number of hidden layers	6
Number of neurons per hidden layer	88
Number of epochs	26 750
Learning rate	3.08×10^{-4}
Batch size	152
Loss function	Sigmoid
Optimizer	Adam

tude of the differences is below 2 cm, which for all practical purposes is excellent agreement.

To ensure that the behavior is the same over the entire time period, we pick four random time steps and repeat the procedure (Fig. 12). Note that for an easier comparison between the various temporal responses, we plot the solution over the number of nodes and no longer in the 2D representation from above. We observe that the approximation quality is good over the entire time period and that the errors tend to decrease with progressing time steps, which is again related to the nonlinearity.

3.5 Physics-based machine learning versus data-driven machine learning

In the previous section, we demonstrated how the non-intrusive RB method can be used to construct reliable and efficient surrogate models through three designated benchmark examples. The focus of this paper is to illustrate the perspec-

tive of physics-based machine learning for subsurface geoscientific applications and how this perspective might differ for various approaches existing for physics-based machine learning. Therefore, we extend the previous sections by constructing surrogate models through neural networks for all three examples. We chose neural networks for this purpose because this enables a comparison between a physics-based and a data-driven machine learning approach, while at the same time also showing the differences between the two paradigms presented for physics-based machine learning, as we detail later on.

One advantage associated with physics-based machine learning vs. data-driven approaches is the reduction in the amount of training data required by reducing the number of admissible solutions through physical knowledge (Faroughi et al., 2022; Raissi et al., 2019). Due to the simplicity of the presented examples this could not be observed, and we obtained similar global errors for both the non-intrusive RB and the NN surrogate models. Nonetheless, it has been shown in, for instance, Santos et al. (2022) that for more complex applications significantly fewer data are required for physics-based approaches.

Another important aspect becomes apparent by focusing on the local error distributions presented in Fig. 13. Generally, the error distributions show a noisy behavior for the data-driven approach, whereas they exhibit a smooth behavior for the non-intrusive RB method. Furthermore, we can observe in the areas marked with A and B that boundary conditions are not preserved. Not only are they not preserved but also show one of the highest errors in the entire surrogate model, which is related to the general challenge of optimizing for loss function values close to zero (Chuang and Barba, 2022). Furthermore, we observe a sharp line in area C, where regions of higher and lower error values occur adjacent to each other, which can not be explained through physical processes. Similar observations of error distributions not corresponding to physical effects are observed for areas D and E. In contrast, the error behavior of the non-intrusive RB method is clearly related to physical processes and changes in material properties, as detailed in the previous sections. This observation is not only relevant for the comparison of data-driven and physics-based machine learning approaches but also for distinguishing between the various physics-based machine learning methods. Methods that use physics-guided loss functions (e.g., PINNs) will exhibit the same error behavior as the physics only enter in the loss functions, and physics-guided architecture will also suffer from this as long as the outer layer still operates as a neural network.

Finally, in Table 5, we compare the construction times for the surrogate models (for both the non-intrusive RB method and the NN), the evaluation times for a single surrogate model run, and the original computation times for the full-dimensional simulations. The first observation is that the construction time of the surrogate models is considerably smaller for the non-intrusive RB method than for the NN.

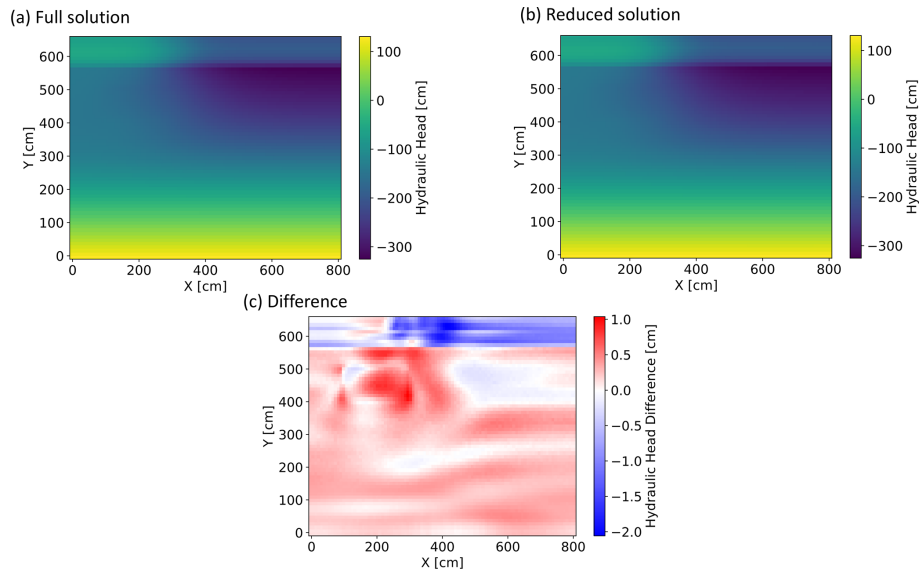


Figure 11. Comparison of the prediction accuracy of the surrogate model for subsurface hydrology. We show in (a) the full solution of the Richards equation, in (b) the prediction from the surrogate model, and in (c) the difference between the full and reduced solutions for the time step 30 of an arbitrarily chosen solution from the validation dataset.

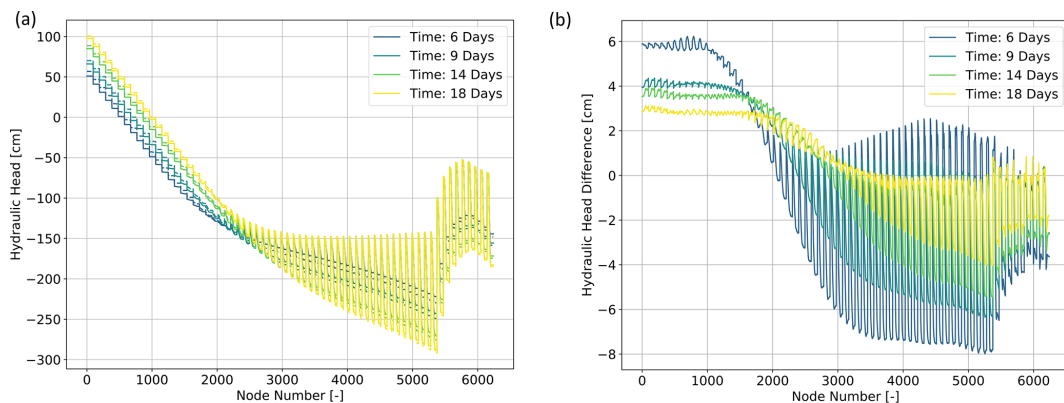


Figure 12. Comparison of the prediction accuracy of the surrogate model for various time steps. We show in (a) the response and in (b) the differences between the full and reduced solutions. In panel (a) the full solutions are denoted by solid lines and the reduced solutions by dashed lines.

This is related to the dimension of the training dataset entering the machine learning part. For the non-intrusive RB method, we first determine the basis functions and determine only the coefficient over the machine learning technique. This means that the training data have the dimension $N_s \times N_{\text{bfs}}$, where N_s is the number of snapshots and N_{bfs} is the number of basis functions. In contrast, the dimension of the training dataset for the NN is $N_s \times N_{\text{dofs}}$, where N_{dofs} denotes the number of spatial degrees of freedom (e.g., nodes, elements). Typically the number of basis functions is orders of magnitude smaller than the degrees of freedom, yielding a reduction in the construction time. The evaluation cost of the surrogate models is comparable for both the non-intrusive RB method and the NN. Furthermore, the evaluation times

of the surrogate model are between 3 and 6 orders of magnitude lower than the full-order evaluations. This demonstrates the benefits of surrogate models for applications where either results are required in real time or numerous evaluations are necessary. Note that again the discussed aspect of the increased construction time applies not only to NNs but also to methods such as PINNs.

To conclude, even in applications where we do not have the added benefit of the reduction in the amount of simulation required, the non-intrusive RB method performs better than data-driven alternatives, while yielding a lower computational cost.

Another point to note is that in this paper we looked at parameterized partial differential equations, assuming the pa-

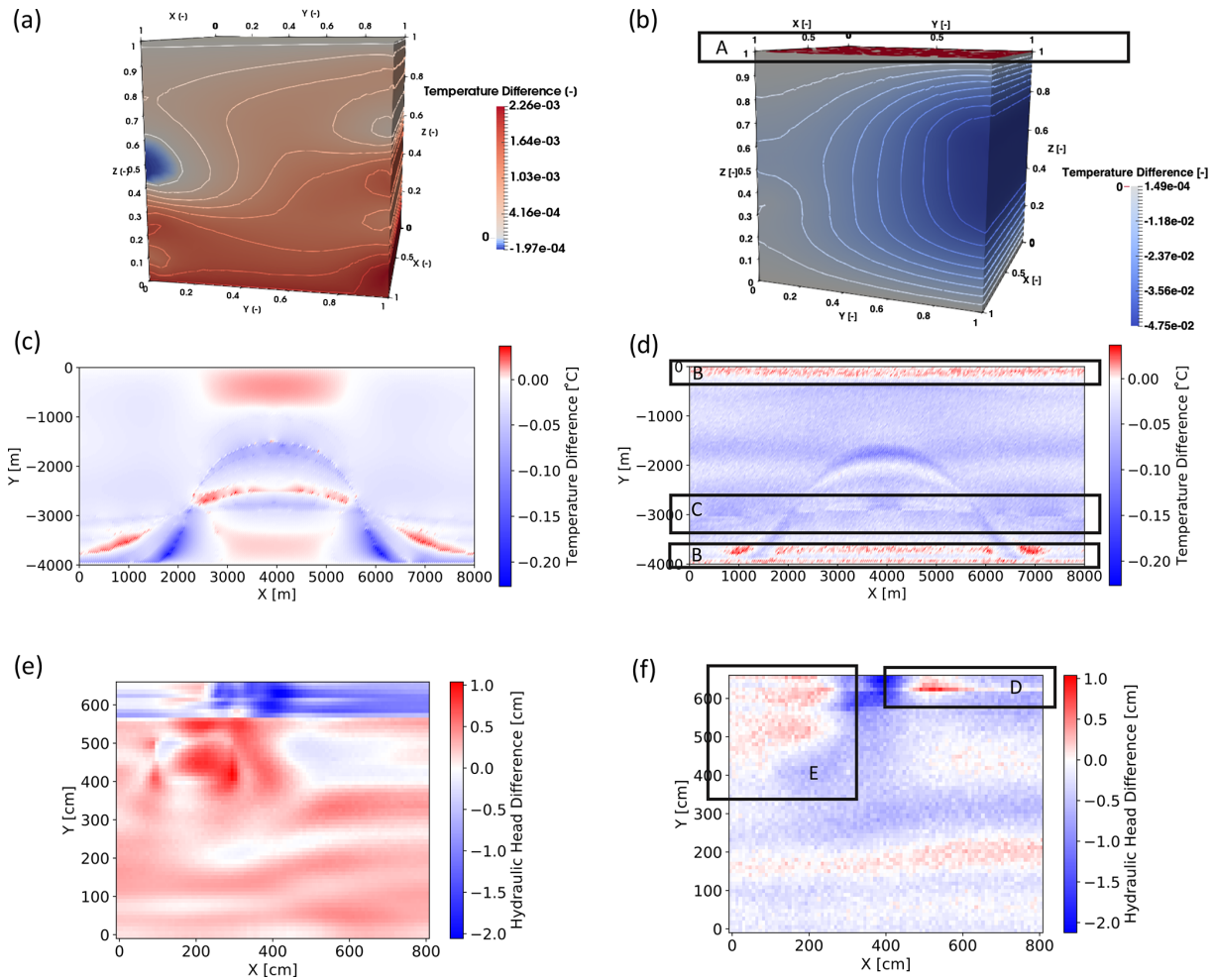


Figure 13. Comparison of the local error distribution for the surrogate models of the geothermal example (a) non-intrusive RB method and (b) NN, the geodynamic example (c) non-intrusive RB method and (d) NN, and the hydrological example (e) non-intrusive RB method and (f) NN.

Table 5. Comparison of the computational cost for the three benchmark examples considering both physics-based and data-driven machine learning methods. Note that the construction time excludes the time required for the generation of the training data. The hyperparameters of the NN are identical to the ones described for the non-intrusive RB method. For the geothermal example, we use the hyperparameters provided for the geodynamic example since the non-intrusive RB method used a Gaussian process regression, and we perform the construction only for the temperature.

	Geothermal		Geodynamic		Hydrological	
	NI-RB	NN	NI-RB	NN	NI-RB	NN
Surrogate construction time [min]	0.58	84	7	125	20	115
Surrogate evaluation time [ms]	1	3	2	3	3	2
Full model solve [min]	1.4		63		0.05	

parameter of interest is a material parameter. This assumption is purely exemplary. PINNs sampling the input from the state itself can incorporate changing initial and boundary conditions through the sampling. Similar considerations are valid for the non-intrusive RB method. Here, additional param-

eters are included to take variations in the initial and boundary conditions into account. Examples are provided in, for instance, Degen (2020), Degen and Cacace (2021), Degen et al. (2022b), and Grepl (2005).

4 Challenges and requirements

We introduced three challenges and four requirements common to many geoscientific applications in the Introduction. In the following, we elaborate on how physics-based machine learning, in particular the non-intrusive RB method, can help to address these challenges and fulfill the requirements.

4.1 Challenge 1: sensitivity analysis

Global sensitivity analyses determine the relevant physical parameters for the respective processes and their correlation (Sobol, 2001). Therefore, they enhance our understanding of the subsurface but are computationally demanding, making them prohibitive for large-scale models. The non-intrusive RB method addresses this computational challenge through the construction of surrogate models. As pointed out by Razavi et al. (2012) and Song et al. (2015) the usage of surrogate models in sensitivity analyses is critical. It is critical because if the surrogate model does not represent the high-dimensional model appropriately, we face the high risk of producing biased sensitivity analyses and deriving the wrong conclusions. Although we agree with this statement, we think it is useful to distinguish between different surrogate modeling techniques highlighting the importance of a surrogate model fulfilling the four listed requirements. As shown for its intrusive counterpart, the results of a sensitivity analysis do not change significantly with the surrogate model accuracy (Degen and Cacace, 2021).

The non-intrusive RB method maintains the input–output relationship of the PDEs, in contrast to, for instance, PINNs. This means that in the classical formulation PINNs sample from spaces in time and space and produce as output state information at any arbitrary location of the model. These considerations are for a fixed set of model parameters (e.g., material properties), making the method unsuitable for sensitivity analysis, where the model parameters need to be changed. This is a disadvantage not only of PINNs but also of all other methods that follow the idea of going from state information to state information, while maintaining the same material properties. Note that recent developments (Zhang et al., 2022) allow a variation of the material properties. However, the disadvantage is that the material properties embedded in the loss function change throughout the training process, meaning that an unphysical correlation between hyperparameters and physical parameters is introduced.

In contrast, the non-intrusive RB method yields surrogates that do not change the general characteristics of the original problem. Consequently, we can address the challenge of performing computationally expensive global sensitivity analysis by using surrogate models constructed by the non-intrusive RB method. This reduces the computational cost by several orders of magnitude, making it a suitable technique to also ensure the feasibility of global sensitivity studies for

large-scale models (Degen et al., 2021b; Degen and Cacace, 2021; Degen et al., 2021a, 2022b).

4.2 Challenge 2: uncertainty

Another application field where the non-intrusive RB method offers great potential is the field of uncertainty quantification. Here, the method can be used to generate fast surrogate models as a replacement for the computationally expensive high-dimensional problem (Degen et al., 2021b, 2022c, a). Uncertainty quantification (Iglesias and Stuart, 2014) determines the uncertainties of the input parameters and has great potential in reducing the risk associated with installations. Uncertainties can be further reduced by finding the optimal position of new measurement locations by employing optimal experimental design methods (Alexanderian, 2021). Also, for this technique, the non-intrusive RB method is suitable to construct surrogate models as shown in Degen et al. (2022a). An important consideration in the usage of surrogate models for uncertainty quantification is the reliability of the surrogate. The surrogate needs to be representative of the original high-dimensional model because otherwise the uncertainty quantification is biased, analog to the discussion in the previous section. Therefore, we require at least error estimators for the surrogate model and in the ideal case error bounds, i.e., guarantees of the model accuracy. In contrast to other surrogate modeling techniques, the non-intrusive RB method further has the advantage of preserving the input–output relationship and therefore maintains the structure of the original model.

Adaptive MCMC approaches are also interesting, where a smaller reduced basis is constructed which is adapted during the MCMC analysis. The idea of constructing an adaptable reduced basis is presented in Qian et al. (2017), where the authors create a reduced basis that is valid for a trust region. An adaptive MCMC algorithm for MT applications has been presented in Manassero et al. (2020), which follows up on the ideas of Ortega-Gelabert et al. (2020).

To conclude, as for the sensitivity analyses the computational challenge is addressed by using surrogate models. Note that as before methods that map from model parameters to state information are more suitable within an uncertainty quantification framework.

4.3 Challenge 3: real time

In many geoscientific applications, we need to obtain results in real time. Examples are forecasting applications often employed in the fields of hydrology and climate (Bauer et al., 2021; Sabeena and Reddy, 2017), as well as monitoring and/or risk analyses (Asante-Duah, 2021; Kumar et al., 2021; Yu et al., 2018). The latter analysis is often important for fields such as geothermal energy and natural hazards (Kumar et al., 2021; Liu and Ramirez, 2017; Yu et al., 2018). In the field of geothermal energy that concerns the topic of induced seismicity, there is a need to have fast and reliable predic-

tions to employ necessary security measures. However, with increasingly complex models that require more and more computation time, this leads to major challenges. In order to address these challenges, a possibility might be to think of employing physically simplified models, which could result in additional error sources that might bias further analyses. Another possibility is to employ data-driven approaches; however, how the quality of these models can be guaranteed is an open research question (Baker et al., 2019; Willcox et al., 2021). Therefore, physics-based machine learning techniques, such as the non-intrusive RB method that fulfill the four listed requirements, have great potential since the surrogate models produced are computationally fast to compute and at the same time maintain the general characteristics of the original problem.

Consequently, the challenge of producing real-time predictions is addressable by the use of surrogate models. We generally see greater potential in methods that follow the conceptual ideas of the non-intrusive RB method since they allow for both parameter and state estimation. Nonetheless, in the context of state estimation and in situations in which some physical knowledge is available, methods that map from state-to-state information (such as PINNs) are advantageous since they allow also a direct incorporation of the measurement data. However, these are not the target applications of this paper.

Looking at all three presented challenges, we see great potential in the non-intrusive RB method, which combines physics-based and data-driven approaches. This avoids simplifications on the physical level and at the same time yields explainable, generalizable, scalable, and robust models (Willcox et al., 2021), as we will detail in the next sections.

4.4 Requirement 1.1: physical scalability

To improve our understanding of the subsurface, we often want to construct models with an increasingly higher resolution in space and time and also concerning the parameter representation. To give a few examples, in the field of geodynamics we typically aim to investigate different scales in both time and space. On the one hand, we are interested in large-scale spatial deformations related to plate tectonics; on the other hand, we want to investigate how microstructures form. For the time component, we face the same issues of being interested in seismic cycles (of the order of milliseconds) and basin dynamics (of the order of millions of years) (Bauer et al., 2021; van Zelst et al., 2022). Having a single model that crosses over all scales is computationally infeasible, even with adaptive time stepping and spatial discretization schemes. Furthermore, we need to additionally consider the complexity of the parameter space. On a larger scale, local heterogeneities are often negligible and we can use homogenization approaches (Regenauer-Lieb et al., 2013) and often simpler physical formulations. On a smaller scale, these

concepts are not applicable and we need to consider the full complexity.

One way to address this problem is a multilevel approach, a concept that works independently of physics-based machine learning. However, physics-based machine learning and other surrogate modeling techniques can considerably accelerate the procedure, as we show later on. The advantages of multilevel approaches to improve convergence have been pointed out by, for instance, Bauer et al. (2021). We can, for example, combine models with different spatial and temporal resolutions, which yields faster convergences. This faster convergence is achieved by starting the solve on the coarsest level first. Once we obtain results for this level, we move to the next level which has a higher resolution. However, since we already have a good starting guess, we need significantly fewer iterations on this level in comparison to the number of iterations required if we directly solved on the finer level. Multilevel or multi-fidelity approaches not only imply different spatial and temporal resolutions but also allow for the employment of different physical descriptions (adjusted to the resolution) and different methodologies (Pehrerstorfer et al., 2018).

Physics-based machine learning can improve the computational time on several levels. First of all, we can provide surrogate models for the coarse levels through methods such as the non-intrusive RB method. The non-intrusive RB method performs very well in the reduction of the spatial and temporal degrees of freedom but is inefficient if the dimension of the parameter space becomes too large. However, other methods might be less efficient in the reduction of space and time but more efficient in dealing with high-dimensional parameter spaces (Benner et al., 2015). Multilevel approaches allow combining these techniques. Hence, we can employ different methodologies on each of these levels tailored to the different requirements. On the finest level, we can even employ the original full-dimensional problem.

4.5 Requirement 1.2: performance scalability

In the following, we elaborate on how physics-based machine learning approaches impact the field of HPC. Therefore, we first present the current issues we face and afterwards explain how the methods presented here help to address these aspects.

Previously, we illustrated that the complexity of the considered physical processes and their associated coupling processes are constantly increasing. This has a high impact on the computational resources since more complex scenarios yield a higher computational demand, hence often requiring HPC infrastructures. At the same time, we observe the end of the scalability laws such as Dennard scaling (Frank et al., 2001) and Moore's law (Bondyopadhyay, 1998).

A paradigm shift is underway as a consequence of this. The hardware infrastructure has been moving more and more to graphics processing units (GPUs) and is becoming more

heterogeneous than before. This poses major challenges for most areas of geosciences (Bauer et al., 2021), as we can no longer rely on hardware developments to cover the required higher computational demand (end of the free-lunch era), but we need to invest in adapting the algorithms and codes themselves (Bauer et al., 2021). This is a nontrivial and potentially time-consuming process that often requires rather specialized knowledge. Moreover, the specific problems can vary greatly for the different geoscientific communities. For example, the climate community relies on very large, well-established community models. Although efforts are underway to port them to GPUs (Bauer et al., 2021), the size and complexity of the codes, and potentially even their inherent structure, make this an ambitious and costly task (Lawrence et al., 2018). Often, more than one intensive kernel may exist and the low-hanging computational hotspot to port is not obvious (Gan et al., 2020). A criticism raised by Bauer et al. (2021) is that software developments are primarily driven by scientific relevance, neglecting computer science aspects in terms of efficiency, which further complicates the transition to future HPC infrastructures. The situation is different for other geoscientific communities since they have smaller – and possibly more manageable – codes, which can therefore be more flexibly adapted, or more recent codes, which may even already leverage novel technologies. In either case, porting tasks are still not trivial, but with the help of current developments to leverage GPUs, such as CUDA and OpenACC, as well as domain-specific languages (DSLs) – e.g., Kokkos (Edwards et al., 2014; Trott et al., 2021), RAJA (Beckingsale et al., 2019), Alpaka (Zenker et al., 2016), GridTools (Afanasyev et al., 2021), ClawDSL (Clement et al., 2018) – the resources and expertise required can be reduced or at least compartmentalized and better distributed in time. We can also provide cross-domain platforms such as MOOSE and Firedrake to synchronize the developments in computer and domain sciences. Moreover, it is becoming increasingly relevant to consider co-design approaches (Germann, 2021), in which domain science applications (and developers) engage with library, back-end, and even hardware developers in pointing out where significant (progressive and disruptive) advances can be made without constant adaptation from scientific applications.

Other aspects that we would like to address here are the discretization of time (Bauer et al., 2021) and how to adapt beyond the current HPC infrastructures. This includes future proofing against architectural shifts, vendors, and languages. To name one example, we require strategies to adapt to GPUs from various manufacturers (e.g., upcoming AMD and Intel GPU accelerated systems). Solutions such as CUDA are optimized for Nvidia hardware, and comparing the performance of CUDA against OpenACC (or CUDA through portability layers such as Kokkos) yields indications that this transition still poses challenges (Artigues et al., 2019; Baig et al., 2020). These considerations do not even cover the problems

that might arise from tensor processing units (TPUs). Overall, the heterogeneous computing infrastructures of the future and asynchronous processes pose major and partly unpredictable challenges, emphasizing the need for flexible, open-source software projects.

There are various possibilities to address these issues. The first is naturally the intensive training of researchers in the field of HPC, including, for instance, advanced programming training sessions, which is a long learning process. However, here we will focus on how physics-based machine learning and other methodologies can help to fulfill these requirements.

Therefore, let us briefly recap the physical problem. In fields such as climate, groundwater, geodynamical, and geothermal applications we often have problems that are governed by partial differential equations (PDEs). The porting of PDE solvers is an ongoing research question (Alexander et al., 2020; Bertagna et al., 2019; Grete et al., 2021; Hokkanen et al., 2021) and no general solution is available yet. We continue to face an immediate need to adjust to new infrastructures consisting of both central processing units (CPUs) and GPUs, and many codes are still not ported. An alternative approach to harness HPC resources is the usage of physics-based machine learning methods to leverage heterogeneous computations. The need to use GPUs has also been recognized in climate science (Bauer et al., 2021), where the use of machine learning techniques wherever it may be possible without a loss of quality is seen as a way forward. Similar arguments are present in other related communities where porting and machine learning are seen as ways to address the computational demands in fluid mechanics (Dauxois et al., 2021) and in hydrodynamics where they are perceived as possible surrogates for forecasting (Morales-Hernández et al., 2020).

In the following, we explain how physics-based machine learning methods help to optimally use HPC infrastructures, starting with PINNs. PINNs are directly portable to GPUs since they consist of a deep neural network. The calculation of the PDE inside the loss function also offers portability since the derivatives are calculated via automatic differentiation (Raissi et al., 2019). Additionally, the non-intrusive RB method is partly portable to GPUs insofar as the (deep) neural network that replaces the classical Galerkin projection can be directly ported. The calculation of the PDEs and hence the generation of the training data are not easily portable. However, note that this task is embarrassingly parallel since all evaluations are independent of each other. Hence, we can take full advantage of a given HPC system. This is not the case for the inverse processes for which we use the surrogate models. Methods such as Markov chain Monte Carlo are only partly parallelizable since they are dependent on previous results, further emphasizing the usefulness of the methodologies presented here. The presented schemes propose an intermediate solution to solve the paradigm of the new architectural constraints. It can also easily incorporate any new ad-

vances for porting PDEs to GPUs. Of course, limitations can still arise when the size and cost of forward simulations for the training set are too large, which reverts the issue to porting PDE solvers. However, once that is achieved, physics-based machine learning workflows directly benefit.

The advantage of such workflows is that GPU support for most of the algorithms required currently exists in environments such as TensorFlow, PyTorch (as described in Sect. 4.7), and DSLs. With the software packages presented here and environments ranging from machine learning libraries over cross-domain finite-element solvers to DSLs, we aim to provide a future-proof concept that can also adapt to new GPU systems (including vendors other than Nvidia) and new architectures such as TPUs. In the ideal case, software built upon these frameworks requires only minimal changes, if at all, to adapt to the architectural changes.

The high portability of these workflows is in line with the general proposition in Bauer et al. (2021), where the authors emphasize the need to use utmost flexible software for adaptability with short-lived hardware infrastructures. We strongly highlight the fact that the life span of scientific software is usually substantially longer than that of hardware. Consequently, software sustainability must be at the center of geoscientific model development. In the context of a changing hardware landscape, it is therefore essential to future-proof codes so that their applicability to new hardware (and software) paradigms is ensured, contributing to their sustainability. Ideally, it will not be necessary to repeat substantial porting efforts in the future. This future proofing is nontrivial and intertwined with ongoing porting efforts, and to achieve these porting efforts we should strive to achieve performance portability, productivity, and maintainability. For this, a good separation of concerns (between geoscientific code, algorithms, and hardware-specific code) is necessary, for which again solutions such as DSLs are interesting, as well as ever more popular approaches (Artigues et al., 2019; Evans et al., 2021; Lawrence et al., 2018). Although there are several paths to this, we argue that leveraging well-established accelerated environments to build physics-based machine learning models acting as PDE solver surrogates is one approach.

From the numerics and algorithmic point of view, there are also opportunities to improve scalability and performance, some of which may also be better suited to upcoming hardware, but for which new implementations in specific codes are inescapable. These include ideas such as mesh adaptivity (e.g., Chen et al., 2021; Kevlahan, 2021; Gerhard et al., 2015; Özgen-Xian et al., 2020; Piggott et al., 2005), higher-order local schemes (e.g., Abdi et al., 2017; Käser and Dumbser, 2006; Gassner and Winters, 2021; Kärnä et al., 2018), local time stepping (e.g., Dazzi et al., 2018), and parallelization in time (Lions et al., 2001; Maday, 2008). These different strategies can also be combined into even more sophisticated algorithms exploiting several features (e.g., Caviedes-Voullième et al., 2020; Dawson et al., 2013; Rannabauer et al., 2018). Mesh adaptivity seeks to reduce computational

load, compress data, and reduce communications while keeping accuracy, all of which contribute to scalability. Higher-order schemes which preserve locality (e.g., discontinuous Galerkin) increase the load on arithmetic where accelerators excel and allow for a reduction of the number of elements (which potentially reduces memory and communication costs) for a target accuracy in comparison to lower-order schemes. Local time stepping explicit schemes allow for asynchronous updates, which enhances parallelization and may allow reducing costly reduction operations.

Similarly, parallelization in time can strongly relax the fundamental constraint of one time step depending on the previous one (Bauer et al., 2021; Carraro et al., 2015; Fisher and Gürol, 2017; Hamon et al., 2020). Interestingly, some physics-based machine learning techniques provide parallel-in-time schemes (Meng et al., 2020). Here, we highlight the parareal algorithm, first developed by Lions et al. (2001). The current stage of the algorithm is presented in Maday (2008). The algorithm follows domain decomposition approaches that divide the domain into smaller subdomains. Domain decomposition methods then solve the problem iteratively over the subdomains using different processors. The parareal algorithm divides the global time evolution into a set of independent smaller time intervals. For the iterative algorithm, the method relies on a predictor–corrector approach that yields generally fast convergences (Maday, 2008). A parareal implementation for PINNs, called PPINNs has been already developed in Meng et al. (2020). Similarly, the parallel full approximation scheme in space and time (PFASST) (Emmett and Minion, 2012, 2014; Hamon et al., 2020) falls into the same class as parareal algorithms.

4.6 Requirement 2: interpretability

The next requirement that needs to be addressed is interpretability. For PDE-based applications, we often face the need to provide predictions. However, guaranteeing the correctness of these predictions with surrogate models is challenging (Willcox et al., 2021), especially for nonlinear applications. Many surrogate model techniques that work well for linear examples do not work in the nonlinear setting (Grepl, 2012). Some physics-based methods, such as the reduced-basis method, which yield interpretable models, are extendable to the nonlinear case but are significantly impacted in their efficiency when encountering high degrees of nonlinearity (Hesthaven and Ubbiali, 2018). Neural networks are known for approximating nonlinear problems well (Raissi et al., 2019) but require a high amount of data, do not preserve physical principles, and are in general not explainable.

Both of these aspects are critical since applications in the fields of, for instance, geothermal energy and geodynamics do not have a sufficient amount of data. Furthermore, we need to conduct predictions and risk assessments, which are critical with data-driven schemes since their quality is related to the quality and amount of training data. These aspects

are the reason why the mentioned application fields heavily rely on physical models. We see greater potential for methods such as the non-intrusive RB method that incorporate the physics into the surrogate model construction instead of PINNs since the non-intrusive RB method results in explainable surrogate models.

To clarify, PINNs are a complete black-box approach. That means we provide our input data (e.g., sample points in space and time) and obtain predictions for the entire state. So, the method can reconstruct the state for non-sampled points in space and time. In contrast to the non-intrusive RB method, we provide our input data (e.g., rock and fluid properties) and obtain the corresponding state output. This state output consists of two parts, the basis functions (determined by the sampling method, e.g., the POD method) and the weights of the basis functions (determined by the machine learning method). This has several advantages.

1. The POD modes correspond to the different physical processes (as demonstrated in the case studies in this paper).
2. The decomposition of our original problem into basis function time weights is mathematically identical (Hesthaven and Ubbiali, 2018).
3. The machine learning method only investigates the weights.

The last point is especially important for nonlinear examples. Although a PDE might be nonlinear, that does not necessarily mean that the variations of the state are highly nonlinear because of different model parameters. The RB method focuses on the variations of the state caused by changes in the model parameters. So, it depends on this complexity and not on the spatial and temporal distributions.

4.7 Requirement 3: generalizability

In the following, we want to emphasize the importance of generalizability. Approaches to construct suitable surrogate models need to be applicable across different disciplines and not restricted to single applications only. In this paper, we focus on PDE-based systems and demonstrated the suitability of the non-intrusive RB method for three geoscientific applications with partly significantly varying conditions and requirements. The non-intrusive RB methods works efficiently for all three examples, which cover a wide range of the general challenges encountered in geoscience applications and many other PDE-dominated domains.

Connected to the aspect of generalizability is also the question of accessibility, meaning the open-access and open-source availability of research and software. Accessibility is a topic that has great importance in general, especially in the context of interdisciplinarity. The importance of accessibility was already recognized in Bergen et al. (2019), who point out that the recent rapid growth of geoscientific applications

using machine learning is associated with the availability of easy-to-use machine learning toolboxes.

Machine learning methods are mainly developed by the computer science community, whereas model order reduction methods (the second set of methods presented here) originate from the scientific computing and applied mathematical community (Swischuk et al., 2019). Geosciences belong to neither of these communities. However, it heavily relies on the methodological developments within these communities to further advance. This opens the path to several subsequent challenges since the terminology and the purposes of the various communities are vastly different.

Consequently, if new algorithms are solely presented in the form of papers, expert knowledge in several fields is required. This is a major challenge since it makes the exploration of different methodologies extremely time-consuming, especially for changing applications. Therefore, it is of great importance to have flexible access to a broad range of methods. Note that with the term flexible, we mean the implementation of algorithms that are not tied to a specific forward solver. Examples of such accessible software frameworks are available for many scientific fields. For inverse methods, packages as Dakota (Adams et al., 2020), Raven (Rabiti et al., 2020), and various designated Python libraries such as SciPy (Virtanen et al., 2020), SALib (Herman and Usher, 2017), and pyMC (Patil et al., 2010) are available. Naturally, this does not mean that no prior knowledge about the methods is required. A general understanding of the methods is necessary to ensure reliable operation and performance of the respective software packages. The step of accessibility should solely simplify the applicability and testing of the different methods for scientific purposes.

In the following, we want to clarify how we understand the term accessibility and what impacts it has on the development of research software. Accessible software means open-source software. However, accessibility means more than open-source. Following the classification of black-box models in Peherstorfer and Willcox (2016), we also define software packages as inaccessible if the complexity of the code prevents an understanding of the methodology and if there is no detailed documentation on the package. Regarding this last point, we want to highlight the difference between educational and scientific software packages. Educational software packages are often written in scripting and/or parsed languages, such as Python, and use simple structures for the implementation (e.g., Ballarin et al., 2017). This eases the learning of the procedures but comes at the cost of lower efficiency. In contrast, scientific software packages are often written in compiled language packages, such as C++, and/or using highly object-oriented structures (e.g., Adams et al., 2020; Rabiti et al., 2020). This complicates the learning process but allows the efficient execution of the methods for real-case applications. In no way do we want to state that only educational software packages are accessible and should be further developed. The point that we want to make here is

that scientific software packages additionally require extensive documentation and tutorials to compensate for the complex nature of the code, as presented in, for example, Adams et al. (2020), Herman and Usher (2017), Patil et al. (2010), Rabiti et al. (2020), and Virtanen et al. (2020). However, scientific software packages that lack these additional resources fall under the category of inaccessible software packages in our opinion.

With this last aspect, we already covered the main impacts of accessibility on software development. From the point of view of understanding, scripting and/or parsed languages are preferential. However, from the efficiency point of view, compiled languages would be preferential. Having access to the methods in various compute languages is naturally the ideal case. Here, the parsed implementation serves as an entry point to improve the understanding, whereas the compiled implementation is used later on for actual studies. However, developing the code in several languages or structures is time-consuming and often neither possible nor feasible. Therefore, the abovementioned compensation by documentation seems to be a suitable compromise from our point of view. Also, note that accessibility is not a responsibility of the developing communities (such as computer sciences) alone but a joint responsibility of all communities.

For machine learning methods various high-performance Python packages exist, such as Keras (Chollet, 2015), MXNet (Chen et al., 2015), PyTorch (Paszke et al., 2019), TensorFlow (Abadi et al., 2015), Theano (Al-Rfou et al., 2016), DeepXDE (implementation of PINNs and various modifications, Lu et al., 2019), HeAT (Götz et al., 2020), and SciANN (implementation of PINNs based on Keras/TensorFlow, Haghghat and Juanes, 2021).

The variety of available packages shows the potential for easy usability of machine learning in geosciences but also presents one major danger. The software packages are rapidly developing and, for some, the developments have stopped (Al-Rfou et al., 2016). Hence, software that relies on these packages should be developed with the possibility of interchanging these libraries. Clearly, this is a nontrivial task but is nevertheless important to consider.

Regarding open-source finite-element solvers, several packages are available (in various compute languages): e.g., FEniCS (Alnæs et al., 2015), Firedrake (Rathgeber et al., 2016), and the MOOSE framework (Permann et al., 2020). Note that we need these solvers to generate the input data for physics-based machine learning methods.

4.8 Requirement 4: robustness

The last requirement we address here is the aspect of robustness. Hence, the question of how to ensure that the model does not change if, for instance, the accuracy or the amount of data used for the surrogate model construction is increased. For the application of data-driven methods to PDE-based systems, that can be rather challenging. If we, for in-

stance, generate training sets from numerical simulations and use only spatial and temporal sample points of the solutions to construct our surrogates (as done for PINNs), we need to be careful that we include all physical effects. To give an example, we need to ensure that for a geothermal study we do not extract samples only in the conductive regime but also in the advective regime. This is an aspect that does not occur for the non-intrusive RB method since it constructs the surrogate model directly on the solution space. For the non-intrusive RB method (geothermal example in Sect. 3.2) we construct two surrogate models with different training dataset sizes. Although we see a difference in the overall accuracy of the model, both models return the same general characteristics of the physical problem, demonstrating the robustness. This can be further highlighted by looking at the intrusive counterpart of the method. Degen and Cacace (2021) show that the results of the sensitivity analysis are not impacted by the model accuracy, which demonstrates that the general characteristic is the same for all models.

5 Conclusions

We present the great potential of physics-based machine learning for geoscientific systems governed by partial differential equations through three designated examples from the fields of geothermal energy, geodynamical systems, and hydrology. These three applications fields were chosen to demonstrate the potential across different geoscientific applications. We address three challenges here: enabling efficient (i) global sensitivity analyses, (ii) uncertainty quantification, and (iii) real-time applications for geoscientific problems dominated by PDEs. Furthermore, it is essential that the techniques used for enabling these efficient analyses fulfill the requirements of being scalable, interpretable, generalizable, and robust.

Bergen et al. (2019) pointed out the importance of providing an optimal predictive understanding of the processes for different geosystems. Predictability is a great challenge in ML techniques if based purely on data. In contrast, methods that incorporate our physical knowledge have great advantages in the geosciences. We see particular potential for the non-intrusive RB method since it maintains the structure of the original physical problem, a viewpoint that is also supported by Reichstein et al. (2019) and Willcox et al. (2021).

The presented methods allow for real-time or faster-than-real-time updates of models. This makes them interesting for applications and concepts such as digital twins. Digital twins are digital copies of the original process with real-time capabilities to assist in, for instance, decision- and policy-making (Bauer et al., 2021). However, we are aware that the geoscience community is rather skeptical concerning black-box approaches such as ML (Bauer et al., 2021). For this reason, it is important to ensure transparent qualitative and quantitative measures (Baker et al., 2019), which are a big challenge

for ML in general. In this regard, we would like to point out again the non-intrusive RB method, which is in contrast to the other methodologies based on a mathematically rigorous proven method. The intrusive version has limited applicability and efficiency for nonlinear (Hesthaven and Ubbiali, 2018) and hyperbolic PDEs (e.g., Benner et al., 2015; Hesthaven et al., 2016). In the original formulation of the RB method, a Galerkin projection is used instead of the ML techniques presented here. In consequence, this means that by performing simplifications to the physical problem, we are always able to fall back on a proven scheme for quality assurance purposes for elliptic and parabolic PDEs.

Current research on the mathematical foundations of physics-based machine learning also includes efficient error estimators for hyperbolic PDEs, as well as extension to high-dimensional parameter spaces. Further developments in these directions will allow for ever wider applicability of the methods in the field of geosciences.

Code and data availability. The training and validation datasets, their associated model parameters, and the non-intrusive RB and neural network code for the construction of all surrogate models are published in the following Zenodo repository (<https://doi.org/10.5281/zenodo.8369108>, Degen et al., 2023). For the construction of the geothermal dataset the software package DwarfElephant (Degen et al., 2020a, b) has been used. This software is based on the finite-element solver MOOSE (Permann et al., 2020) and is freely available on Zenodo (<https://doi.org/10.5281/zenodo.4074777>). For the generation of the hydrology dataset, we used ParFlow v3.10 (available on Zenodo, <https://doi.org/10.5281/zenodo.6413322>, Smith et al., 2022) and for the geodynamic dataset SULEC v6.1. The presented geodynamic benchmark example is conceptually simple and reproducible in other open-source software packages such as Aspect v.2.4 (available on Zenodo, <https://doi.org/10.5281/zenodo.6903424>, Bangerth et al., 2022) with the data provided in the paper.

Author contributions. All authors discussed and interpreted the presented work. DD carried out the simulations and drafted the paper. The “Hydrological example” and “High-performance computing” sections were jointly drafted by DCV and DD. AGN configured, ran, and post-processed the hydrological simulations. The salt dome example was set up by SB. Furthermore, all authors read and approved the final paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes ev-

ery effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The authors gratefully acknowledge the funding provided by Exploratory Research Space Prep Funds through the project “High-Performance Computing in the Geosciences” (project number: PF-JARA-SDS009). Furthermore, we thankfully acknowledge the funding provided by the BMBF BioökonomieREVIER funding scheme with its “BioRevierPlus” project (funding ID: 031B1137EX).

Financial support. This research has been supported by the RWTH Aachen University (grant no. PF-JARA-SDS009) and the BMBF (grant no. 031B1137EX).

This open-access publication was funded by the RWTH Aachen University.

Review statement. This paper was edited by James Kelly and Paul Ullrich and reviewed by two anonymous referees.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, <https://www.tensorflow.org/> (last access: 24 September 2023), 2015.
- Abdi, D. S., Wilcox, L. C., Warburton, T. C., and Giraldo, F. X.: A GPU-accelerated continuous and discontinuous Galerkin non-hydrostatic atmospheric model, *Int. J. High Perform. C.*, 33, 81–109, <https://doi.org/10.1177/1094342017694427>, 2017.
- Adams, B., Bohnhoff, W., Dalbey, K., Ebeida, M., Eddy, J., Eldred, M., Hooper, R., Hough, P., Hu, K., Jakeman, J., Khalil, M., Maupin, K., Monschke, J., Ridgway, E., Rushdi, A., Seidl, D., Stephens, J., Swiler, L., and Winokur, J.: DAKOTA, A Multi-level Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.12 User’s Manual, Sandia National Laboratories, Tech. Rep., SAND2020-12495, 2020.
- Afanasyev, A., Bianco, M., Mosimann, L., Osuna, C., Thaler, F., Vogt, H., Fuhrer, O., VandeVondele, J., and Schulthess, T. C.: GridTools: A framework for portable weather and climate applications, *Software X*, 15, 100707, <https://doi.org/10.1016/j.softx.2021.100707>, 2021.
- Alexander, F., Almgren, A., Bell, J., Bhattacharjee, A., Chen, J., Colella, P., Daniel, D., DeSlippe, J., Diachin, L., Draeger, E., Dubey, A., Dunning, T., Evans, T., Foster, I., Francois, M., Germann, T., Gordon, M., Habib, S., Halappanavar, M., Hamilton, S., Hart, W., Huang, Z. H., Hungerford, A., Kasen, D.,

- Kent, P. R. C., Kolev, T., Kothe, D. B., Kronfeld, A., Luo, Y., Mackenzie, P., McCallen, D., Messer, B., Mniszewski, S., Oehmen, C., Perazzo, A., Perez, D., Richards, D., Rider, W. J., Rieben, R., Roche, K., Siegel, A., Sprague, M., Steefel, C., Stevens, R., Syamlal, M., Taylor, M., Turner, J., Vay, J.-L., Voter, A. F., Windus, T. L., and Yelick, K.: Exascale applications: skin in the game, *Philos. T. Roy. Soc. A*, 378, 20190056, <https://doi.org/10.1098/rsta.2019.0056>, 2020.
- Alexanderian, A.: Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: a review, *Inverse Problems*, 37, 043001, <https://doi.org/10.1088/1361-6420/abe10c>, 2021.
- Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., and Wells, G. N.: The FEniCS project version 1.5, *Archive of Numerical Software*, 3, 9–23, <https://doi.org/10.11588/ans.2015.100.20553>, 2015.
- Al-Rfou, R., Alain, G., Almahairi, A., et al.: Theano: A Python framework for fast computation of mathematical expressions, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1605.02688>, 2016.
- Antoulas, A. C.: *Approximation of Large-Scale Dynamical Systems*, Society for Industrial and Applied Mathematics, ISBN 978-0-89871-658-0, eISBN 978-0-89871-871-3, <https://doi.org/10.1137/1.9780898718713>, 2005.
- Ardabili, S., Mosavi, A., Dehghani, M., and Várkonyi-Kóczy, A. R.: Deep Learning and Machine Learning in Hydrological Processes Climate Change and Earth Systems a Systematic Review, in: *Lecture Notes in Networks and Systems*, Springer International Publishing, 52–62, https://doi.org/10.1007/978-3-030-36841-8_5, 2020.
- Artigues, V., Kormann, K., Rampp, M., and Reuter, K.: Evaluation of performance portability frameworks for the implementation of a particle-in-cell code, *Concurrency and Computation Practice and Experience*, 32, e5640, <https://doi.org/10.1002/cpe.5640>, 2019.
- Asante-Duah, D. K.: *Hazardous waste risk assessment*, CRC Press, <https://doi.org/10.1201/9781003070009>, 2021.
- Audouze, C., De Vuyst, F., and Nair, P.: Reduced-order modeling of parameterized PDEs using time-space-parameter principal component analysis, *Int. J. Numer. Meth. Eng.*, 80, 1025–1057, 2009.
- Audouze, C., De Vuyst, F., and Nair, P. B.: Nonintrusive reduced-order modeling of parametrized time-dependent partial differential equations, *Numer. Meth. Part. D. E.*, 29, 1587–1628, 2013.
- Baig, F., Gao, C., Teng, D., Kong, J., and Wang, F.: Accelerating Spatial Cross-Matching on CPU-GPU Hybrid Platform With CUDA and OpenACC, *Frontiers in Big Data*, 3, 14, <https://doi.org/10.3389/fdata.2020.00014>, 2020.
- Baker, N., Alexander, F., Bremer, T., Hagberg, A., Kevrekidis, Y., Najm, H., Parashar, M., Patra, A., Sethian, J., Wild, S., Willcox, K., and Lee, S.: Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence, *Tech. Rep.*, 1478744, <https://doi.org/10.2172/1478744>, 2019.
- Ballarin, F., Sartori, A., and Rozza, G.: RBniCS-reduced order modelling in FEniCS, <http://mathlab.sissa.it/rbnics> (last access: 23 September 2023), 2017.
- Bandai, T. and Ghezzehei, T. A.: Physics-Informed Neural Networks With Monotonicity Constraints for Richardson-Richards Equation: Estimation of Constitutive Relationships and Soil Water Flux Density From Volumetric Water Content Measurements, *Water Resour. Res.*, 57, e2020WR027642, <https://doi.org/10.1029/2020wr027642>, 2021.
- Bangerth, W., Dannberg, J., Fraters, M., Gassmoeller, R., Glerum, A., Heister, T., Myhill, R., and Naliboff, J.: ASPECT v2.4.0 (v2.4.0), Zenodo [code], <https://doi.org/10.5281/zenodo.6903424>, 2022.
- Barrault, M., Maday, Y., Nguyen, N. C., and Patera, A. T.: An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations, *C. R. Math.*, 339, 667–672, 2004.
- Bar-Sinai, Y., Hoyer, S., Hickey, J., and Brenner, M. P.: Learning data-driven discretizations for partial differential equations, *P. Natl. Acad. Sci. USA*, 116, 15344–15349, 2019.
- Baş, D. and Boyacı, I. H.: Modeling and optimization I: Usability of response surface methodology, *J. Food Eng.*, 78, 836–845, 2007.
- Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., and Wedi, N. P.: The digital revolution of Earth-system science, *Nature Computational Science*, 1, 104–113, 2021.
- Beckingsale, D. A., Burmark, J., Hornung, R., Jones, H., Killian, W., Kunen, A. J., Pearce, O., Robinson, P., Ryu-jin, B. S., and Scogland, T. R.: RAJA: Portable Performance for Large-Scale Scientific Applications, in: *2019 IEEE/ACM International Workshop on Performance, Portability and Productivity in HPC (P3HPC)*, 71–81, <https://doi.org/10.1109/p3hpc49587.2019.00012>, 2019.
- Benner, P., Gugercin, S., and Willcox, K.: A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems, *SIAM Rev.*, 57, 483–531, <https://doi.org/10.1137/130932715>, 2015.
- Bentivoglio, R., Isufi, E., Jonkman, S. N., and Taormina, R.: Deep Learning Methods for Flood Mapping: A Review of Existing Applications and Future Research Directions, *Hydrol. Earth Syst. Sci. Discuss. [preprint]*, <https://doi.org/10.5194/hess-2021-614>, 2021.
- Bergen, K. J., Johnson, P. A., Maarten, V., and Beroza, G. C.: Machine learning for data-driven discovery in solid Earth geoscience, *Science*, 363, e2020WR027642, <https://doi.org/10.1126/science.aau0323>, 2019.
- Bermúdez, M., Ntegeka, V., Wolfs, V., and Willems, P.: Development and Comparison of Two Fast Surrogate Models for Urban Pluvial Flood Simulations, *Water Resour. Manage.*, 32, 2801–2815, <https://doi.org/10.1007/s11269-018-1959-8>, 2018.
- Bertagna, L., Deakin, M., Guba, O., Sunderland, D., Bradley, A. M., Tezaur, I. K., Taylor, M. A., and Salinger, A. G.: HOMMEXX 1.0: a performance-portable atmospheric dynamical core for the Energy Exascale Earth System Model, *Geosci. Model Dev.*, 12, 1423–1441, <https://doi.org/10.5194/gmd-12-1423-2019>, 2019.
- Bertsimas, D. and Tsitsiklis, J. N.: *Introduction to linear optimization*, vol. 6, Athena Scientific Belmont, MA, 1997.
- Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S., and Escalreira, L. A.: Response surface methodology (RSM) as a tool for optimization in analytical chemistry, *Talanta*, 76, 965–977, 2008.
- Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, *Hydrol. Process.*, 9, 251–290, <https://doi.org/10.1002/hyp.3360090305>, 1995.
- Bondyopadhyay, P. K.: Moore’s law governs the silicon revolution, *P. IEEE*, 86, 78–81, 1998.

- Buiter, S. and Ellis, S.: SULEC: Benchmarking a new ALE finite-element code, in: EGU General Assembly Conference Abstracts, 7528, <https://ui.adsabs.harvard.edu/abs/2012EGUGA..14.7528B/abstract> (last access: 8 December 2023), 2012.
- Busto, S., Stabile, G., Rozza, G., and Vázquez-Cendón, M. E.: POD–Galerkin reduced order methods for combined Navier–Stokes transport equations based on a hybrid FV-FE solver, *Comput. Math. Appl.*, 79, 256–273, 2020.
- Carraro, T., Geiger, M., Rorkel, S., and Rannacher, R.: *Multiple Shooting and Time Domain Decomposition Methods*, Springer, ISBN 978-3-319-23320-8, 2015.
- Caviedes-Voullième, D., García-Navarro, P., and Murillo, J.: Verification, conservation, stability and efficiency of a finite volume method for the 1D Richards equation, *J. Hydrol.*, 480, 69–84, 2013.
- Caviedes-Voullième, D., Gerhard, N., Sikstel, A., and Müller, S.: Multiwavelet-based mesh adaptivity with Discontinuous Galerkin schemes: Exploring 2D shallow water problems, *Adv. Water Resour.*, 138, 103559, <https://doi.org/10.1016/j.advwatres.2020.103559>, 2020.
- Chen, J., Hoversten, G. M., Key, K., Nordquist, G., and Cumming, W.: Stochastic inversion of magnetotelluric data using a sharp boundary parameterization and application to a geothermal site, *Geophysics*, 77, E265–E279, 2012.
- Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K.: Neural ordinary differential equations, *Adv. Neur. In.*, 31, ISBN 9781510884472, 2018a.
- Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C., and Zhang, Z.: MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1512.01274>, 2015.
- Chen, W., Hesthaven, J. S., Junqiang, B., Qiu, Y., Yang, Z., and Tihao, Y.: Greedy nonintrusive reduced order model for fluid dynamics, *AIAA Journal*, 56, 4927–4943, 2018b.
- Chen, X., Duan, J., and Karniadakis, G. E.: Learning and meta-learning of stochastic advection-diffusion-reaction systems from sparse measurements, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1910.09098>, 2019.
- Chen, Y., Hesthaven, J. S., Maday, Y., and Rodríguez, J.: Certified reduced basis methods and output bounds for the harmonic Maxwell’s equations, *SIAM J. Sci. Comput.*, 32, 970–996, 2010.
- Chen, Y., Simon, K., and Behrens, J.: Extending legacy climate models by adaptive mesh refinement for single-component tracer transport: a case study with ECHAM6-HAMMOZ (ECHAM6.3-HAM2.3-MOZ1.0), *Geosci. Model Dev.*, 14, 2289–2316, <https://doi.org/10.5194/gmd-14-2289-2021>, 2021.
- Chollet, F.: Keras, <https://keras.io> (last access: 23 September 2023), 2015.
- Chuang, P.-Y. and Barba, L. A.: Experience report of physics-informed neural networks in fluid simulations: pitfalls and frustration, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2205.14249>, 2022.
- Clement, V., Ferrachat, S., Fuhrer, O., Lapillonne, X., Osuna, C. E., Pincus, R., Rood, J., and Sawyer, W.: The CLAW DSL, in: PASC ’18: Proceedings of the Platform for Advanced Scientific Computing Conference, ACM, Basel Switzerland, 2–4 July 2018, <https://doi.org/10.1145/3218176.3218226>, 2018.
- Condon, L. E., Kollet, S., Bierkens, M. F., Fogg, G. E., Maxwell, R. M., Hill, M. C., Fransen, H.-J. H., Verhoef, A., Van Loon, A. F., Sulis, M., and Abesser, C.: Global groundwater modeling and monitoring: Opportunities and challenges, *Water Resour. Res.*, 57, e2020WR029500, <https://doi.org/10.1029/2020WR029500>, 2021.
- Conway, D., Simpson, J., Didana, Y., Rugari, J., and Heinson, G.: Probabilistic magnetotelluric inversion with adaptive regularisation using the no-U-turns sampler, *Pure Appl. Geophys.*, 175, 2881–2894, 2018.
- Cressie, N.: The origins of kriging, *Math. Geol.*, 22, 239–252, 1990.
- Dauxois, T., Peacock, T., Bauer, P., Caulfield, C. P., Cenedese, C., Gorré, C., Haller, G., Ivey, G. N., Linden, P. F., Meiburg, E., Pinardi, N., Vriend, N. M., and Woods, A. W.: Confronting Grand Challenges in environmental fluid mechanics, *Physical Review Fluids*, 6, 020501, <https://doi.org/10.1103/physrevfluids.6.020501>, 2021.
- Dawson, C., Trahan, C. J., Kubatko, E. J., and Westerink, J. J.: A parallel local timestepping Runge–Kutta discontinuous Galerkin method with applications to coastal ocean modeling, *Comput. Method. Appl. M.*, 259, 154–165, <https://doi.org/10.1016/j.cma.2013.03.015>, 2013.
- Dazzi, S., Vacondio, R., Palù, A. D., and Mignosa, P.: A local time stepping algorithm for GPU-accelerated 2D shallow water models, *Adv. Water Resour.*, 111, 274–288, <https://doi.org/10.1016/j.advwatres.2017.11.023>, 2018.
- De Bézenac, E., Pajot, A., and Gallinari, P.: Deep learning for physical processes: Incorporating prior scientific knowledge, *J. Stat. Mech-Theory E.*, 2019, 124009, <https://doi.org/10.1088/1742-5468/ab3195>, 2019.
- Degen, D. and Cacace, M.: Effects of transient processes for thermal simulations of the Central European Basin, *Geosci. Model Dev.*, 14, 1699–1719, <https://doi.org/10.5194/gmd-14-1699-2021>, 2021.
- Degen, D., Veroy, K., and Wellmann, F.: Certified reduced basis method in geosciences, *Comput. Geosci.*, 24, 241–259, 2020a.
- Degen, D., Veroy, K., and Wellmann, F.: cgreachen/DwarfElephant: DwarfElephant 1.0, Zenodo [code], <https://doi.org/10.5281/zenodo.4074777>, 2020b.
- Degen, D., Spooner, C., Scheck-Wenderoth, M., and Cacace, M.: How biased are our models? – a case study of the alpine region, *Geosci. Model Dev.*, 14, 7133–7153, <https://doi.org/10.5194/gmd-14-7133-2021>, 2021a.
- Degen, D., Veroy, K., Freymark, J., Scheck-Wenderoth, M., Poulet, T., and Wellmann, F.: Global sensitivity analysis to optimize basin-scale conductive model calibration – A case study from the Upper Rhine Graben, *Geothermics*, 95, 102143, <https://doi.org/10.1016/j.geothermics.2021.102143>, 2021b.
- Degen, D., Cacace, M., and Wellmann, F.: 3D multi-physics uncertainty quantification using physics-based machine learning, *Scientific Reports*, 12, 17491, <https://doi.org/10.1038/s41598-022-21739-7>, 2022a.
- Degen, D., Veroy, K., Scheck-Wenderoth, M., and Wellmann, F.: Crustal-scale thermal models: Revisiting the influence of deep boundary conditions, *Environ. Earth Sci.*, 81, 88, <https://doi.org/10.1007/s12665-022-10202-5>, 2022b.
- Degen, D., Veroy, K., and Wellmann, F.: Uncertainty quantification for basin-scale geothermal conduction models, *Scientific*

- Reports, 12, 4246, <https://doi.org/10.1038/s41598-022-08017-2>, 2022c.
- Degen, D., Caviedes Voullième, D., Buitter, S., Hendriks Franssen, H.-J., Vereecken, H., González-Nicolás, A., and Wellmann, F.: Non-Intrusive Reduced Basis Code for Geothermal, Geodynamic, and Hydrology Benchmarks, Zenodo [code and data set], <https://doi.org/10.5281/zenodo.8369108>, 2023.
- Degen, D. M.: Application of the reduced basis method in geophysical simulations: concepts, implementation, advantages, and limitations, PhD thesis, Rheinisch-Westfälische Technische Hochschule Aachen, Aachen, <https://doi.org/10.18154/RWTH-2020-12042>, 2020.
- Dwivedi, V. and Srinivasan, B.: Solution of Biharmonic Equation in Complicated Geometries With Physics Informed Extreme Learning Machine, *J. Comput. Inf. Sci. Eng.*, 20, 061004, <https://doi.org/10.1115/1.4046892>, 2020.
- Edwards, H. C., Trott, C. R., and Sunderland, D.: Kokkos: Enabling manycore performance portability through polymorphic memory access patterns, *J. Parallel Distr. Com.*, 74, 3202–3216, <https://doi.org/10.1016/j.jpdc.2014.07.003>, 2014.
- Ellis, S., Little, T., Wallace, L., Hacker, B., and Buitter, S.: Feedback between rifting and diapirism can exhume ultrahigh-pressure rocks, *Earth Planet. Sc. Lett.*, 311, 427–438, 2011.
- Emmett, M. and Minion, M.: Toward an efficient parallel in time method for partial differential equations, *Comm. App. Math. Com. Sc.*, 7, 105–132, 2012.
- Emmett, M. and Minion, M. L.: Efficient implementation of a multi-level parallel in time algorithm, in: *Domain Decomposition Methods in Science and Engineering XXI*, Springer, 359–366, ISBN 978-3-319-05788-0, 2014.
- Evans, T. M., Siegel, A., Draeger, E. W., Deslippe, J., Francois, M. M., Germann, T. C., Hart, W. E., and Martin, D. F.: A survey of software implementations used by application codes in the Exascale Computing Project, *Int. J. High Perform. C.*, 36, 109434202110289, <https://doi.org/10.1177/10943420211028940>, 2021.
- Falkner, S., Klein, A., and Hutter, F.: BOHB: Robust and efficient hyperparameter optimization at scale, in: *International Conference on Machine Learning*, PMLR, Stockholm, Sweden, 10–15 July 2018, 1437–1446, <https://proceedings.mlr.press/v80/falkner18a.html> (last access: 22 September 2023), 2018.
- Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., Hazenberg, P., McNamara, J., Pelletier, J., Perket, J., Rouholahnejad-Freund, E., Wagener, T., Zeng, X., Beighley, E., Buzan, J., Huang, M., Livneh, B., Mohanty, B. P., Nijssen, B., Safeeq, M., Shen, C., van Verseveld, W., Volk, J., and Yamazaki, D.: Hillslope Hydrology in Global Change Research and Earth System Modeling, *Water Resour. Res.*, 55, 1737–1772, <https://doi.org/10.1029/2018wr023903>, 2019.
- Faroughi, S. A., Pawar, N., Fernandes, C., Das, S., Kalantari, N. K., and Mahjour, S. K.: Physics-guided, physics-informed, and physics-encoded neural networks in scientific computing, arXiv [preprint], <https://doi.org/10.48550/arXiv.2211.07377>, 2022.
- Fisher, M. and Gürol, S.: Parallelization in the time dimension of four-dimensional variational data assimilation, *Q. J. Roy. Meteor. Soc.*, 143, 1136–1147, 2017.
- Forsyth, P. A., Wu, Y. S., and Pruess, K.: Robust numerical methods for saturated-unsaturated flow with dry initial conditions in heterogeneous media, *Adv. Water Resour.*, 18, 25–38, 1995.
- Frangos, M., Marzouk, Y., Willcox, K., and van Bloemen Waanders, B.: Surrogate and reduced-order modeling: a comparison of approaches for large-scale statistical inverse problems, Chapt. 7, John Wiley & Sons. Print ISBN 9780470697436, Online ISBN 9780470685853, <https://doi.org/10.1002/9780470685853>, 2010.
- Frank, D. J., Dennard, R. H., Nowak, E., Solomon, P. M., Taur, Y., and Wong, H.-S. P.: Device scaling limits of Si MOSFETs and their application dependencies, *P. IEEE*, 89, 259–288, 2001.
- Freyermark, J., Bott, J., Cacace, M., Ziegler, M., and Scheck-Wenderoth, M.: Influence of the Main Border Faults on the 3D Hydraulic Field of the Central Upper Rhine Graben, *Geofluids*, 2019, 7520714 <https://doi.org/10.1155/2019/7520714>, 2019.
- Gan, L., Fu, H., and Yang, G.: Translating novel HPC techniques into efficient geoscience solutions, *Journal of Computational Science*, 52, 101212, <https://doi.org/10.1016/j.jocs.2020.101212>, 2020.
- Gassner, G. J. and Winters, A. R.: A Novel Robust Strategy for Discontinuous Galerkin Methods in Computational Fluid Mechanics: Why? When? What? Where?, *Frontiers in Physics*, 8, 500690, <https://doi.org/10.3389/fphy.2020.500690>, 2021.
- Gelet, R., Loret, B., and Khalili, N.: A thermo-hydro-mechanical coupled model in local thermal non-equilibrium for fractured HDR reservoir with double porosity, *J. Geophys. Res.-Sol. Ea.*, 117, B07205, <https://doi.org/10.1029/2012JB009161>, 2012.
- Geneva, N. and Zabarar, N.: Modeling the dynamics of PDE systems with physics-constrained deep autoregressive networks, *J. Comput. Phys.*, 403, 109056, <https://doi.org/10.1016/j.jcp.2019.109056>, 2020.
- Gerhard, N., Caviedes-Voullième, D., Müller, S., and Kesserwani, G.: Multiwavelet-Based Grid Adaptation with Discontinuous Galerkin Schemes for Shallow Water Equations, *J. Comput. Phys.*, 301, 265–288, <https://doi.org/10.1016/j.jcp.2015.08.030>, 2015.
- Germann, T. C.: Co-design in the Exascale Computing Project, *Int. J. High Perform. C.*, 35, 503–507, <https://doi.org/10.1177/10943420211059380>, 2021.
- Ghasemi, M. and Gildin, E.: Model order reduction in porous media flow simulation using quadratic bilinear formulation, *Comput. Geosci.*, 20, 723–735, 2016.
- Gosses, M., Nowak, W., and Wöhling, T.: Explicit treatment for Dirichlet, Neumann and Cauchy boundary conditions in POD-based reduction of groundwater models, *Adv. Water Resour.*, 115, 160–171, 2018.
- Götz, M., Debus, C., Coquelin, D., Krajsek, K., Comito, C., Knechtges, P., Hagemeyer, B., Tarnawa, M., Hanselmann, S., Siggel, M., Basermann, A., and Streit, A.: HeAT – a Distributed and GPU-accelerated Tensor Framework for Data Analytics, in: *Proceedings of the 19th IEEE International Conference on Big Data*, IEEE, Atlanta, USA, 10–13 December 2020, 276–288, <https://doi.org/10.1109/BigData50022.2020.9378050>, 2020.
- Grepl, M.: Reduced-basis Approximation and A Posteriori Error Estimation for Parabolic Partial Differential Equations, PhD thesis, Massachusetts Institute of Technology, <https://dspace.mit.edu/handle/1721.1/32387> (last access: 24 September 2023), 2005.

- Grepl, M. A.: Model order reduction of parametrized nonlinear reaction–diffusion systems, *Comput. Chem. Eng.*, 43, 33–44, 2012.
- Grete, P., Glines, F. W., and O’Shea, B. W.: K-Athena: A Performance Portable Structured Grid Finite Volume Magnetohydrodynamics Code, *IEEE T. Parallel Distrib. Syst.*, 32, 85–97, <https://doi.org/10.1109/tpds.2020.3010016>, 2021.
- Haghighat, E. and Juanes, R.: SciANN: A Keras/TensorFlow wrapper for scientific computations and physics-informed deep learning using artificial neural networks, *Comput. Method. Appl. M.*, 373, 113552, <https://doi.org/10.1016/j.cma.2020.113552>, 2021.
- Hammerschmidt, M., Herrmann, S., Pomplun, J., Zschiedrich, L., Burger, S., and Schmidt, F.: Reduced basis method for Maxwell’s equations with resonance phenomena, in: *Optical Systems Design 2015: Computational Optics*, SPIE, vol. 9630, 138–151, <https://doi.org/10.1117/12.2190425>, 2015.
- Hamon, F. P., Schreiber, M., and Minion, M. L.: Parallel-in-time multi-level integration of the shallow-water equations on the rotating sphere, *J. Comput. Phys.*, 407, 109210, <https://doi.org/10.1016/j.jcp.2019.109210>, 2020.
- He, Q., Barajas-Solano, D., Tartakovsky, G., and Tartakovsky, A. M.: Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport, *Adv. Water Resour.*, 141, 103610, <https://doi.org/10.1016/j.advwatres.2020.103610>, 2020.
- Herman, J. and Usher, W.: SALib: An open-source Python library for Sensitivity Analysis, *J. Open Source Softw.*, 2, 97, <https://doi.org/10.21105/joss.00097>, 2017.
- Hess, M. W. and Benner, P.: Fast evaluation of time–harmonic Maxwell’s equations using the reduced basis method, *IEEE T. Microw. Theory*, 61, 2265–2274, 2013a.
- Hess, M. W. and Benner, P.: Reduced basis generation for Maxwell’s equations by rigorous error estimation, in: *19th International Conference on the Computation of Electromagnetic Fields (COMPUMAG 2013)*, Budapest, Hungary, 30 June–4 July 2013, PD2–10, <https://doi.org/10.1109/TMAG.2014.2299393>, 2013b.
- Hess, M. W., Grundel, S., and Benner, P.: Estimating the inf-sup constant in reduced basis methods for time-harmonic Maxwell’s equations, *IEEE T. Microw. Theory*, 63, 3549–3557, 2015.
- Hesthaven, J. and Ubbiali, S.: Non-intrusive reduced order modeling of nonlinear problems using neural networks, *J. Comput. Phys.*, 363, 55–78, <https://doi.org/10.1016/j.jcp.2018.02.037>, 2018.
- Hesthaven, J. S., Rozza, G., and Stamm, B.: *Certified Reduced Basis Methods for Parametrized Partial Differential Equations*, SpringerBriefs in Mathematics, Springer, ISBN 978-3-319-22469-5, 2016.
- Hokkanen, J., Kollet, S., Kraus, J., Herten, A., Hrywniak, M., and Pleiter, D.: Leveraging HPC accelerator architectures with modern techniques – hydrologic modeling on GPUs with ParFlow, *Computat. Geosci.*, 25, 1579–1590, <https://doi.org/10.1007/s10596-021-10051-4>, 2021.
- Huang, P.: Investigating different formulations for hydrothermal convection in geothermal systems, in: *Proceedings World Geothermal Congress 2020 Reykjavik, Iceland, 26 April–2 May 2020* https://www.researchgate.net/profile/Po_Wei_Huang4/publication/ (last access: 20 September 2023), 2018.
- Iglesias, M. and Stuart, A. M.: *Inverse Problems and Uncertainty Quantification*, SIAM News, 2–3, <https://homepages.warwick.ac.uk/~masdr/BOOKCHAPTERS/stuart19c.pdf> (last access: 21 September 2023), 2014.
- Jacquey, A. and Cacace, M.: Modelling fully-coupled Thermo-Hydro-Mechanical (THM) processes in fractured reservoirs using GOLEM: a massively parallel open-source simulator, in: *EGU General Assembly Conference Abstracts*, Vienna, Austria, April 2017, vol. 19, 15721, <https://ui.adsabs.harvard.edu/abs/2017EGUGA..1915721J/abstract> (last access: 22 September 2023), 2017.
- Jacquey, A. B. and Cacace, M.: Multiphysics modeling of a brittle-ductile lithosphere: 2. Semi-brittle, semi-ductile deformation and damage rheology, *J. Geophys. Res.-Sol. Ea.*, 125, e2019JB018475, <https://doi.org/10.1029/2019JB018475>, 2020a.
- Jacquey, A. B. and Cacace, M.: Multiphysics Modeling of a Brittle-Ductile Lithosphere: 1. Explicit Visco-Elasto-Plastic Formulation and Its Numerical Implementation, *J. Geophys. Res.-Sol. Ea.*, 125, e2019JB018474, <https://doi.org/10.1029/2019JB018474>, 2020b.
- Jagtap, A. D. and Karniadakis, G. E.: Extended Physics-Informed Neural Networks (XPINNs): A Generalized Space-Time Domain Decomposition Based Deep Learning Framework for Nonlinear Partial Differential Equations, *Commun. Comput. Phys.*, 28, 2002–2041, 2020.
- Jiang, S., Zheng, Y., and Solomatine, D.: Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning, *Geophys. Res. Lett.*, 47, e2020GL088229, <https://doi.org/10.1029/2020gl088229>, 2020.
- Jones, A. G., Afonso, J. C., and Fulla, J.: Geochemical and geophysical constraints on the dynamic topography of the Southern African Plateau, *Geochem. Geophys. Geosy.*, 18, 3556–3575, 2017.
- Jordan, M. I. and Mitchell, T. M.: Machine learning: Trends, perspectives, and prospects, *Science*, 349, 255–260, 2015.
- Jung, N.: *Error Estimation for Parametric Model Order Reduction and Its Application* (Doctoral dissertation, Technische Universität München), <https://mediatum.ub.tum.de/?id=1080163> (last access: 18 September 2023), 2012.
- Karumuri, S., Tripathy, R., Billionis, I., and Panchal, J.: Simulator-free solution of high-dimensional stochastic elliptic partial differential equations using deep neural networks, *J. Comput. Phys.*, 404, 109120, <https://doi.org/10.1016/j.jcp.2019.109120>, 2020.
- Kevlahan, N. K.-R.: Adaptive Wavelet Methods for Earth Systems Modelling, *Fluids*, 6, 236, <https://doi.org/10.3390/fluids6070236>, 2021.
- Khuri, A. I. and Mukhopadhyay, S.: *Response surface methodology*, Wiley Interdisciplinary Reviews: Computational Statistics, 2, 128–149, 2010.
- Kohl, T., Evans, K., Hopkirk, R., and Rybach, L.: Coupled hydrologic, thermal and mechanical considerations for the simulation of hot dry rock reservoirs, *Geothermics*, 24, 345–359, 1995.
- Kolditz, O., Görke, U.-J., Shao, H., and Wang, W.: *Thermo-hydro-mechanical-chemical processes in porous media: benchmarks and examples*, vol. 86, Springer Science & Business Media, ISBN 978-3-642-27176-2, 2012.
- Koltzer, N., Scheck-Wenderoth, M., Bott, J., Cacace, M., Frick, M., Sass, I., Fritsche, J.-G., and Bär, K.: *The Effects of Regional*

- Fluid Flow on Deep Temperatures (Hesse, Germany), *Energies*, 12, 2081, <https://doi.org/10.3390/en12112081>, 2019.
- Kotsiantis, S. B., Zaharakis, I., and Pintelas, P.: Supervised machine learning: A review of classification techniques, *Emerging artificial intelligence applications in computer engineering*, 160, 3–24, 2007.
- Kuffour, B. N. O., Engdahl, N. B., Woodward, C. S., Condon, L. E., Kollet, S., and Maxwell, R. M.: Simulating coupled surface–subsurface flows with ParFlow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model, *Geosci. Model Dev.*, 13, 1373–1397, <https://doi.org/10.5194/gmd-13-1373-2020>, 2020.
- Kumar, P., Debele, S. E., Sahani, J., Rawat, N., Marti-Cardona, B., Alfieri, S. M., Basu, B., Basu, A. S., Bowyer, P., Charizopoulos, N., Jaakko, J., Loupis, M., Menenti, M., Mickovski, S. B., Pfeiffer, J., Pilla, F., Pröll, J., Pulvirenti, B., Rutzinger, M., Sannigrahi, S., Spyrou, C., Tuomenvirta, H., Vojinovic, Z., and Zieher, T.: An overview of monitoring methods for assessing the performance of nature-based solutions against natural hazards, *Earth-Sci. Rev.*, 217, 103603, <https://doi.org/10.1016/j.earscirev.2021.103603>, 2021.
- Kärnä, T., Kramer, S. C., Mitchell, L., Ham, D. A., Piggott, M. D., and Baptista, A. M.: Thetis coastal ocean model: discontinuous Galerkin discretization for the three-dimensional hydrostatic equations, *Geosci. Model Dev.*, 11, 4359–4382, <https://doi.org/10.5194/gmd-11-4359-2018>, 2018.
- Käser, M. and Dumbser, M.: An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes – I. The two-dimensional isotropic case with external source terms, *Geophys. J. Int.*, 166, 855–877, <https://doi.org/10.1111/j.1365-246x.2006.03051.x>, 2006.
- Lawrence, B. N., Rezný, M., Budich, R., Bauer, P., Behrens, J., Carter, M., Deconinck, W., Ford, R., Maynard, C., Mullerworth, S., Osuna, C., Porter, A., Serradell, K., Valcke, S., Wedi, N., and Wilson, S.: Crossing the chasm: how to develop weather and climate models for next generation computers?, *Geosci. Model Dev.*, 11, 1799–1821, <https://doi.org/10.5194/gmd-11-1799-2018>, 2018.
- Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A.: Fourier neural operator for parametric partial differential equations, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2010.08895>, 2020.
- Liang, J., Li, W., Bradford, S., and Šimůnek, J.: Physics-Informed Data-Driven Models to Predict Surface Runoff Water Quantity and Quality in Agricultural Fields, *Water*, 11, 200, <https://doi.org/10.3390/w11020200>, 2019.
- Lions, J.-L., Maday, Y., and Turinici, G.: Résolution d'EDP par un schéma en temps pararéel, *CR. Acad. Sci. I-Math.*, 332, 661–668, 2001.
- Liu, W. and Ramirez, A.: State of the art review of the environmental assessment and risks of underground geo-energy resources exploitation, *Renew. Sust. Energ. Rev.*, 76, 628–644, 2017.
- Lu, L., Meng, X., Mao, Z., and Karniadakis, G. E.: DeepXDE: A deep learning library for solving differential equations, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1907.04502>, 2019.
- Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G. E.: Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nature Machine Intelligence*, 3, 218–229, 2021.
- Lu, L., Meng, X., Cai, S., Mao, Z., Goswami, S., Zhang, Z., and Karniadakis, G. E.: A comprehensive and fair comparison of two neural operators (with practical extensions) based on fair data, *Comput. Method. Appl. M.*, 393, 114778, <https://doi.org/10.1016/j.cma.2022.114778>, 2022.
- Maday, Y.: The parareal in time algorithm, Citeseer, <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=2d51619eb5d9a281cb26d48243a3ee5c7036e795> (last access: 16 September 2023), 2008.
- Mahesh, B.: Machine learning algorithms – a review, *International Journal of Science and Research (IJSR)*, 9, 381–386, 2020.
- Mainini, L. and Willcox, K.: Surrogate modeling approach to support real-time structural assessment and decision making, *AIAA Journal*, 53, 1612–1626, 2015.
- Malek, A. and Beidokhti, R. S.: Numerical solution for high order differential equations using a hybrid neural network–optimization method, *Appl. Math. Comput.*, 183, 260–271, 2006.
- Manassero, M. C., Afonso, J. C., Zyserman, F., Zlotnik, S., and Fomin, I.: A reduced order approach for probabilistic inversions of 3-D magnetotelluric data I: general formulation, *Geophys. J. Int.*, 223, 1837–1863, 2020.
- Maxwell, R. M., Condon, L. E., and Melchior, P.: A Physics-Informed, Machine Learning Emulator of a 2D Surface Water Model: What Temporal Networks and Simulation-Based Inference Can Help Us Learn about Hydrologic Processes, *Water*, 13, 3633, <https://doi.org/10.3390/w13243633>, 2021.
- Meng, X. and Karniadakis, G. E.: A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems, *J. Comput. Phys.*, 401, 109020, <https://doi.org/10.1016/j.jcp.2019.109020>, 2020.
- Meng, X., Li, Z., Zhang, D., and Karniadakis, G. E.: PPINN: Parareal physics-informed neural network for time-dependent PDEs, *Comput. Method. Appl. M.*, 370, 113250, <https://doi.org/10.1016/j.cma.2020.113250>, 2020.
- Miao, T., Lu, W., Lin, J., Guo, J., and Liu, T.: Modeling and uncertainty analysis of seawater intrusion in coastal aquifers using a surrogate model: a case study in Longkou, China, *Arab. J. Geosci.*, 12, 1, <https://doi.org/10.1007/s12517-018-4128-8>, 2019.
- Mo, S., Shi, X., Lu, D., Ye, M., and Wu, J.: An adaptive Kriging surrogate method for efficient uncertainty quantification with an application to geological carbon sequestration modeling, *Comput. Geosci.*, 125, 69–77, 2019.
- Mohammadi, B.: A review on the applications of machine learning for runoff modeling, *Sustainable Water Resources Management*, 7, 98, <https://doi.org/10.1007/s40899-021-00584-y>, 2021.
- Morales-Hernández, M., Sharif, M. B., Gangrade, S., Dullo, T. T., Kao, S.-C., Kalyanapu, A., Ghafoor, S. K., Evans, K. J., Madadi-Kandjani, E., and Hodges, B. R.: High-performance computing in water resources hydrodynamics, *J. Hydroinform.*, 22, 1217–1235, <https://doi.org/10.2166/hydro.2020.163>, 2020.
- Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M.: *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, ISBN 978-1-118-91601-8, 2016.
- Naliboff, J. B., Buitter, S. J., Péron-Pinvidic, G., Osmundsen, P. T., and Tetreault, J.: Complex fault interaction controls continental rifting, *Nat. Commun.*, 8, 1179, <https://doi.org/10.1038/s41467-017-00904-x>, 2017.

- Navarro, M., Le Maître, O. P., Hoteit, I., George, D. L., Mandli, K. T., and Knio, O. M.: Surrogate-based parameter inference in debris flow model, *Comput. Geosci.*, 22, 1447–1463, 2018.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, *Water Resour. Res.*, 57, e2020WR028091, <https://doi.org/10.1029/2020wr028091>, 2021.
- Nield, D. A. and Bejan, A.: Heat Transfer Through a Porous Medium, in: *Convection in Porous Media*, 37–55, Springer, <https://doi.org/10.1007/978-1-4614-5541-7>, 2017.
- O’Sullivan, M. J., Pruess, K., and Lippmann, M. J.: State of the art of geothermal reservoir simulation, *Geothermics*, 30, 395–429, 2001.
- Ortega-Gelabert, O., Zlotnik, S., Afonso, J. C., and Díez, P.: Fast Stokes flow simulations for geophysical-geodynamic inverse problems and sensitivity analyses based on reduced order modeling, *J. Geophys. Res.-Sol. Ea.*, 125, e2019JB018314, <https://doi.org/10.1029/2019JB018314>, 2020.
- Özgen-Xian, I., Kesserwani, G., Caviedes-Voullième, D., Molins, S., Xu, Z., Dwivedi, D., Moulton, J. D., and Steefel, C. I.: Wavelet-based local mesh refinement for rainfall–runoff simulations, *J. Hydroinform.*, 22, 1059–1077, <https://doi.org/10.2166/hydro.2020.198>, 2020.
- Pang, G., Lu, L., and Karniadakis, G. E.: fPINNs: Fractional physics-informed neural networks, *SIAM J. Sci. Comput.*, 41, A2603–A2626, 2019.
- Pang, G., D’Elia, M., Parks, M., and Karniadakis, G. E.: nPINNs: nonlocal Physics-Informed Neural Networks for a parametrized nonlocal universal Laplacian operator. Algorithms and Applications, *J. Comput. Phys.*, 422, 109760, <https://doi.org/10.1016/j.jcp.2020.109760>, 2020.
- Paniconi, C. and Putti, M.: Physically based modeling in catchment hydrology at 50: Survey and outlook, *Water Resour. Res.*, 51, 7090–7129, <https://doi.org/10.1002/2015WR017780>, 2015.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, in: *Advances in Neural Information Processing Systems 32*, edited by: Wallach, H., Larochelle, H., Beygelzimer, A., d’Alché-Buc, F., Fox, E., and Garnett, R., Curran Associates, Inc., 8024–8035, <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf> (last access: 24 September 2023), 2019.
- Patil, A., Huard, D., and Fonnesbeck, C. J.: PyMC: Bayesian Stochastic Modelling in Python, *J. Stat. Softw.*, 35, 1–81, <https://doi.org/10.18637/jss.v035.i04>, 2010.
- Peherstorfer, B. and Willcox, K.: Data-driven operator inference for nonintrusive projection-based model reduction, *Comput. Method. Appl. M.*, 306, 196–215, <https://doi.org/10.1016/j.cma.2016.03.025>, 2016.
- Peherstorfer, B., Willcox, K., and Gunzburger, M.: Survey of multifidelity methods in uncertainty propagation, inference, and optimization, *SIAM Rev.*, 60, 550–591, 2018.
- Pelletier, D., Fortin, A., and Camarero, R.: Are FEM solutions of incompressible flows really incompressible?(or how simple flows can cause headaches!), *Int. J. Numer. Meth. Fl.*, 9, 99–112, 1989.
- Peng, W., Zhou, W., Zhang, J., and Yao, W.: Accelerating physics-informed neural network training with prior dictionaries, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2004.08151>, 2020.
- Permann, C. J., Gaston, D. R., Andriš, D., Carlsen, R. W., Kong, F., Lindsay, A. D., Miller, J. M., Peterson, J. W., Slaughter, A. E., Stogner, R. H., and Martineau, R. C.: MOOSE: Enabling massively parallel multiphysics simulation, *SoftwareX*, 11, 100430, <https://doi.org/10.1016/j.softx.2020.100430>, 2020.
- Piggott, M., Pain, C., Gorman, G., Power, P., and Goddard, A.: h, r, and hr adaptivity with applications in numerical ocean modelling, *Ocean Model.*, 10, 95–113, <https://doi.org/10.1016/j.ocemod.2004.07.007>, 2005.
- Poulet, T., Veveakis, M., Paesold, M., and Regenauer-Lieb, K.: REDBACK: An open-source highly scalable simulation tool for rock mechanics with dissipative feedbacks, in: *AGU Fall Meeting Abstracts*, San Francisco, USA, 13–15 December 2014, <https://ui.adsabs.harvard.edu/abs/2014AGUFM.H21M..02P/abstract> (last access: 17 September 2023), 2014.
- Prud’homme, C., Rovas, D. V., Veroy, K., Machiels, L., Maday, Y., Patera, A. T., and Turinici, G.: Reliable Real-Time Solution of Parametrized Partial Differential Equations: Reduced-Basis Output Bound Methods, *Journal of Fluids Engineering*, 124, 70–80, 2002.
- Qian, E., Grepl, M., Veroy, K., and Willcox, K.: A certified trust region reduced basis approach to PDE-constrained optimization, *SIAM J. Sci. Comput.*, 39, S434–S460, 2017.
- Quarteroni, A., Manzoni, A., and Negri, F.: *Reduced Basis Methods for Partial Differential Equations: An Introduction*, UNITEXT, Springer International Publishing, ISBN 9783319154305, 2015.
- Rabiti, C., Alfonsi, A., Cogliati, J., Mandelli, D., Kinoshita, R., Sen, S., Wang, C., Talbot, P. W., Maljovec, D. P., and Chen, J.: RAVEN user manual, Tech. Rep., Idaho National Lab.(INL), Idaho Falls, ID (United States), <https://doi.org/10.2172/1784874>, 2020.
- Raissi, M. and Karniadakis, G. E.: Hidden physics models: Machine learning of nonlinear partial differential equations, *J. Comput. Phys.*, 357, 125–141, 2018.
- Raissi, M., Perdikaris, P., and Karniadakis, G. E.: Numerical Gaussian processes for time-dependent and nonlinear partial differential equations, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1703.10230>, 2017.
- Raissi, M., Perdikaris, P., and Karniadakis, G. E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *J. Comput. Phys.*, 378, 686–707, 2019.
- Rannabauer, L., Dumbser, M., and Bader, M.: ADER-DG with a-posteriori finite-volume limiting to simulate tsunamis in a parallel adaptive mesh refinement framework, *Comput. Fluids*, 173, 299–306, <https://doi.org/10.1016/j.compfluid.2018.01.031>, 2018.
- Rao, C., Sun, H., and Liu, Y.: Hard encoding of physics for learning spatiotemporal dynamics, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2105.00557>, 2021.
- Rathgeber, F., Ham, D. A., Mitchell, L., Lange, M., Luporini, F., McRae, A. T., Bercea, G.-T., Markall, G. R., and Kelly, P. H.: Firedrake: automating the finite element method by composing abstractions, *ACM T. Math. Software*, 43, 1–27, <https://doi.org/10.1145/2998441>, 2016.

- Razavi, S., Tolson, B. A., and Burn, D. H.: Review of surrogate modeling in water resources, *Water Resour. Res.*, 48, 7, <https://doi.org/10.1029/2011WR011527>, 2012.
- Regenauer-Lieb, K., Veveakis, M., Poulet, T., Wellmann, F., Karrech, A., Liu, J., Hauser, J., Schrank, C., Gaede, O., and Trefry, M.: Multiscale coupling and multiphysics approaches in earth sciences: Theory, *Journal of Coupled Systems and Multiscale Dynamics*, 1, 49–73, 2013.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, *Nature*, 566, 195–204, 2019.
- Richards, L. A.: Capillary conduction of liquids through porous mediums, *Physics*, 1, 318–333, 1931.
- Rizzo, C. B., de Barros, F. P., Perotto, S., Oldani, L., and Guadagnini, A.: Adaptive POD model reduction for solute transport in heterogeneous porous media, *Comput. Geosci.*, 22, 297–308, <https://doi.org/10.1007/s10596-017-9693-5>, 2017.
- Rochlitz, R., Skibbe, N., and Günther, T.: custEM: Customizable finite-element simulation of complex controlled-source electromagnetic data, *Geophysics*, 84, F17–F33, <https://doi.org/10.1190/geo2018-0208.1>, 2019.
- Rogelj, J. and Knutti, R.: Geosciences after Paris, *Nat. Geosci.*, 9, 187–189, 2016.
- Rosas-Carbajal, M., Linde, N., Kalscheuer, T., and Vrugt, J. A.: Two-dimensional probabilistic inversion of plane-wave electromagnetic data: methodology, model constraints and joint inversion with electrical resistivity data, *Geophys. J. Int.*, 196, 1508–1524, 2014.
- Rousset, M. A., Huang, C. K., Klie, H., and Durlafsky, L. J.: Reduced-order modeling for thermal recovery processes, *Comput. Geosci.*, 18, 401–415, 2014.
- Sabeena, J. and Reddy, P. V. S.: A Review of Weather forecasting schemes, *i-manager's Journal on Pattern Recognition*, 4, 27–30, <https://doi.org/10.26634/jpr.4.2.13731>, 2017.
- Santoso, R., Degen, D., Cacace, M., and Wellmann, F.: State-of-the-art physics-based machine learning for hydro-mechanical simulation in geothermal applications, *European Geothermal Congress, Berlin, Germany, 17–21 October 2022*, 1–10, <https://www.researchgate.net/profile/Ryan-Santoso-2/publication/> (last access: 10 September 2023), 2022.
- Schaa, R., Gross, L., and du Plessis, J.: PDE-based geophysical modelling using finite elements: examples from 3D resistivity and 2D magnetotellurics, *J. Geophys. Eng.*, 13, S59–S73, <https://doi.org/10.1088/1742-2132/13/2/S59>, 2016.
- Schenk, O. and Gärtner, K.: Solving unsymmetric sparse systems of linear equations with PARDISO, *Future Gener. Comp. Sy.*, 20, 475–487, 2004.
- Shah, V., Joshi, A., Ghosal, S., Pokuri, B., Sarkar, S., Ganapathysubramanian, B., and Hegde, C.: Encoding invariances in deep generative models, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1906.01626>, 2019.
- Sharma, R., Farimani, A. B., Gomes, J., Eastman, P., and Pande, V.: Weakly-supervised deep learning of heat transport via physics informed loss, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1807.11374>, 2018.
- Shen, C., Chen, X., and Laloy, E.: Editorial: Broadening the Use of Machine Learning in Hydrology, *Frontiers in Water*, 3, 681023, <https://doi.org/10.3389/frwa.2021.681023>, 2021.
- Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., and Demir, I.: A comprehensive review of deep learning applications in hydrology and water resources, *Water Sci. Technol.*, 82, 2635–2670, <https://doi.org/10.2166/wst.2020.369>, 2020.
- Smith, S., reedmaxwell, i-ferguson, Engdahl, N., FabianGasper, Avery, P., Chennault, C., Jourdain, S., grapp1, Condon, L., Kulkarni, K., Bansal, V., xy124, Bennett, A., basileh, Thompson, D., DrewLazzeriKitware, Swilley, J., Beisman, J., Beisman, J., alanquits, Coon, E., Bertolacci, I. M. S., Lührs, L., arezaii, aureliayang and cswoodward: parflow/parflow: ParFlow Version 3.10.0 (v3.10.0), Zenodo [code], <https://doi.org/10.5281/zenodo.6413322>, 2022.
- Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulat.*, 55, 271–280, 2001.
- Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, *J. Hydrol.*, 523, 739–757, 2015.
- Stefansson, I., Berre, I., and Keilegavlen, E.: A fully coupled numerical model of thermo-hydro-mechanical processes and fracture contact mechanics in porous media, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2008.06289>, 2020.
- Stewart, I. S. and Lewis, D.: Communicating contested geoscience to the public: Moving from “matters of fact” to “matters of concern”, *Earth-Sci. Rev.*, 174, 122–133, 2017.
- Swischuk, R., Mainini, L., Peherstorfer, B., and Willcox, K.: Projection-based model reduction: Formulations for physics-based machine learning, *Comput. Fluids*, 179, 704–717, <https://doi.org/10.1016/j.compfluid.2018.07.021>, 2019.
- Taron, J. and Elsworth, D.: Thermal–hydrologic–mechanical–chemical processes in the evolution of engineered geothermal reservoirs, *Int. J. Rock Mech. Min.*, 46, 855–864, 2009.
- Taron, J., Elsworth, D., and Min, K.-B.: Numerical simulation of thermal-hydrologic-mechanical-chemical processes in deformable, fractured porous media, *Int. J. Rock Mech. Min.*, 46, 842–854, 2009.
- Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., and Barajas-Solano, D.: Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems, *Water Resour. Res.*, 56, e2019WR026731, <https://doi.org/10.1029/2019wr026731>, 2020.
- Tran, H., Leonarduzzi, E., la Fuente, L. D., Hull, R. B., Bansal, V., Chennault, C., Gentine, P., Melchior, P., Condon, L. E., and Maxwell, R. M.: Development of a Deep Learning Emulator for a Distributed Groundwater–Surface Water Model: ParFlow-ML, *Water*, 13, 3393, <https://doi.org/10.3390/w13233393>, 2021.
- Trott, C., Berger-Vergiat, L., Poliakoff, D., Rajamanickam, S., Lebrun-Grandie, D., Madsen, J., Awar, N. A., Gligoric, M., Shipman, G., and Womeldorff, G.: The Kokkos EcoSystem: Comprehensive Performance Portability for High Performance Computing, *Comput. Sci. Eng.*, 23, 10–18, <https://doi.org/10.1109/mcse.2021.3098509>, 2021.
- Vallier, B., Magnenet, V., Schmittbuhl, J., and Fond, C.: THM modeling of hydrothermal circulation at Rittershofen geothermal site, France, *Geothermal Energy*, 6, 22, <https://doi.org/10.1186/s40517-018-0108-1>, 2018.

- van Genuchten, M.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, 44, 892–898, 1980.
- van Zelst, I., Cramer, F., Pusok, A. E., Glerum, A., Dannberg, J., and Thieulot, C.: 101 geodynamic modelling: how to design, interpret, and communicate numerical studies of the solid Earth, *Solid Earth*, 13, 583–637, <https://doi.org/10.5194/se-13-583-2022>, 2022.
- Vermeulen, P. and Heemink, A.: Model-reduced variational data assimilation, *Mon. Weather Rev.*, 134, 2888–2899, 2006.
- Vermeulen, P., Heemink, A. W., and Te Stroet, C. B.: Reduced models for linear groundwater flow models using empirical orthogonal functions, *Adv. Water Resour.*, 27, 57–69, 2004.
- Veroy, K., Prud'homme, C., Rovas, D. V., and Patera, A. T.: A Posteriori Error Bounds for Reduced-Basis Approximation of Parametrized Noncoercive and Nonlinear Elliptic Partial Differential Equations, in: *Proceedings of the 16th AIAA computational fluid dynamics conference*, Orlando, FL, 23–26 June 2003, vol. 3847, <https://doi.org/10.2514/6.2003-3847>, 2003.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nat. Methods*, 17, 261–272, <https://doi.org/10.1038/s41592-019-0686-2>, 2020.
- Volkwein, S.: Model reduction using proper orthogonal decomposition, *Lecture Notes*, Institute of Mathematics and Scientific Computing, University of Graz, <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c8022ab24391f734cba5b41539f651e327c87b48> (last access: 22 September 2023), 2011.
- Wainwright, H. M., Finsterle, S., Jung, Y., Zhou, Q., and Birkholzer, J. T.: Making sense of global sensitivity analyses, *Comput. Geosci.*, 65, 84–94, 2014.
- Wang, Q., Hesthaven, J. S., and Ray, D.: Non-intrusive reduced order modeling of unsteady flows using artificial neural networks with application to a combustion problem, *J. Comput. Phys.*, 384, 289–307, <https://doi.org/10.1016/j.jcp.2019.01.031>, 2019.
- Wang, S., Yu, X., and Perdikaris, P.: When and why PINNs fail to train: A neural tangent kernel perspective, *J. Comput. Phys.*, 449, 110768, <https://doi.org/10.1016/j.jcp.2021.110768>, 2022.
- Wellmann, J. F. and Reid, L. B.: Basin-scale geothermal model calibration: Experience from the Perth Basin, Australia, *Energy Proced.*, 59, 382–389, 2014.
- Willard, J., Jia, X., Xu, S., Steinbach, M., and Kumar, V.: Integrating physics-based modeling with machine learning: A survey, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2003.04919>, 2020.
- Willcox, K. E., Ghattas, O., and Heimbach, P.: The imperative of physics-based modeling and inverse theory in computational science, *Nature Computational Science*, 1, 166–168, 2021.
- Wirtz, D., Karajan, N., and Haasdonk, B.: Surrogate modeling of multiscale models using kernel methods, *Int. J. Numer. Meth. Eng.*, 101, 1–28, <https://doi.org/10.1002/nme.4767>, 2015.
- Wu, J.-L., Kashinath, K., Albert, A., Chirila, D., Xiao, H., and Prabhat: Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems, *J. Comput. Phys.*, 406, 109209, <https://doi.org/10.1016/j.jcp.2019.109209>, 2020.
- Yang, L., Zhang, D., and Karniadakis, G. E.: Physics-informed generative adversarial networks for stochastic differential equations, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1811.02033>, 2018.
- Yang, L., Meng, X., and Karniadakis, G. E.: B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data, *J. Comput. Phys.*, 425, 109913, <https://doi.org/10.1016/j.jcp.2020.109913>, 2020.
- Yang, Y. and Perdikaris, P.: Physics-informed deep generative models, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1812.03511>, 2018.
- Young, C.-C., Liu, W.-C., and Wu, M.-C.: A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events, *Appl. Soft Comput.*, 53, 205–216, <https://doi.org/10.1016/j.asoc.2016.12.052>, 2017.
- Yu, M., Yang, C., and Li, Y.: Big data in natural disaster management: a review, *Geosciences*, 8, 165, <https://doi.org/10.3390/geosciences8050165>, 2018.
- Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., and Behl, M.: Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community, *Water Resour. Res.*, 56, e2019WR027038, <https://doi.org/10.1029/2019wr027038>, 2020.
- Zanchetta, A. D. L. and Coulibaly, P.: Hybrid Surrogate Model for Timely Prediction of Flash Flood Inundation Maps Caused by Rapid River Overflow, *Forecasting*, 4, 126–148, <https://doi.org/10.3390/forecast4010007>, 2022.
- Zenker, E., Worpitz, B., Widera, R., Huebl, A., Juckeland, G., Knüpfer, A., Nagel, W. E., and Bussmann, M.: Alpaka – An Abstraction Library for Parallel Kernel Acceleration, *IEEE Computer Society*, Chicago, USA, 23–27 May 2016, <https://doi.org/10.1109/IPDPSW.2016.50>, 2016.
- Zhang, E., Dao, M., Karniadakis, G. E., and Suresh, S.: Analyses of internal structures and defects in materials using physics-informed neural networks, *Science Advances*, 8, eabk0644, <https://doi.org/10.1126/sciadv.abk0644>, 2022.
- Zhu, Y., Zabar, N., Koutsourelakis, P.-S., and Perdikaris, P.: Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data, *J. Comput. Phys.*, 394, 56–81, 2019.
- Zounemat-Kermani, M., Batelaan, O., Fadaee, M., and Hinkelmann, R.: Ensemble Machine Learning Paradigms in Hydrology: A Review, *J. Hydrol.*, 598, 126266, <https://doi.org/10.1016/j.jhydrol.2021.126266>, 2021.