https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

Physics-Based Machine Learning: Opportunities and Challenges for Mantle Convection

Marilina Valatsou[1,2], Denise Degen[3,4], Juliane Dannberg[5,6], Rene Gassmöller[5], and Florian Wellmann[1,7]

[1]RWTH Aachen University, Institute for Computational Geoscience, Geothermics and Reservoir Geophysics, Mathieustraße 30, 52074 Aachen, Germany
[2]Now at: Institute for Particle Physics and Astrophysics, ETH Zurich, Wolfgang-Pauli-Strasse 27, 8093 Zurich Switzerland
[3]Institute for Applied Geosciences, TU Darmstadt, Schnittspahnstr. 9, 64287 Darmstadt, Germany
[4]GFZ Helmholtz Centre for Geosciences, 14473 Potsdam, Germany
[5]GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany
[6]Christian-Albrecht University of Kiel, Institute of Geosciences, Kiel, Germany
[7]Fraunhofer Research Institution for Energy Infrastructures and Geothermal Systems (IEG), Am Hochschulcampus 1, 44801, Bochum, Germany

Correspondence: Denise Degen (degen@geo.tu-darmstadt.de)

Abstract. Geodynamic research revolves around understanding the Earth's subsurface, which often requires solving complex partial differential equations to produce numerical models. Using state-of-the-art solvers to tackle these high-dimensional problems can be computationally demanding, or even prohibitive. A potential solution is to use surrogate models–methods that reduce the dimensionality of the problem while maintaining its general characteristics. In this work, we focus on developing reliable and efficient surrogate models via physics-based machine learning techniques for mantle convection applications, thus mitigating the computational challenges inherent in direct forward modeling of mantle convection. For this purpose, we employ the non-intrusive reduced basis method (NI-RB), which maintains the accuracy of traditional simulations while significantly reducing computational complexity. The performance of these models is compared against high-dimensional finite element mantle convection models generated from ASPECT, an open-source geodynamical simulation software. We show that the adopted approach significantly speeds up simulations of mantle temperature distribution—enabling rapid multi-query analyses such as global sensitivity studies and uncertainty quantification—while overcoming typical model reduction issues in geodynamics. The results indicate that our surrogate models not only capture the essential dynamics of mantle convection but also offer a balance of computational efficiency and quality of the model. With the potential to transform how geodynamic modeling studies are conducted, these surrogate models hold promise for a more efficient and effective exploration of Earth's subsurface processes in future studies.

1 Introduction

The exploration of the Earth's interior, specifically the mantle, is of paramount importance due to its far-reaching implications for various surface phenomena, including the formation of georesources, orogenic processes, as a driver for the geodynamo that creates Earth's magnetic field, for maintaining temperatures at the Earth's surface in a range suitable for complex life,

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

and more (Schubert et al., 2001). The physical inaccessibility of the deeper layers of the Earth due to engineering constraints has caused an increasing use of numerical simulations for geodynamic research studies. These forward models are rigorous in their precision and offer important insight into processes occurring in the Earth's interior. However, the high computational cost associated with individual forward models poses a significant challenge, restricting the number of simulations that can be executed, especially in geodynamic contexts, as for example pointed out by van Zelst et al. (2022) and others.

In light of these challenges, there is a rising interest in the application of machine learning (ML) methodologies to develop surrogate models which, if constructed proficiently, can achieve similar levels of accuracy as the forward models, but at a fraction of the computational cost. Several studies in different application fields consider ML methods; examples include non-intrusive reduced-basis (NI-RB) methods for geotechnical engineering (Chakir and Hammond, 2018) and for modeling of urban flows (Chakir et al., 2021). Additionally, an NI-RB method, similar to the one presented in this study, is used by Wang et al. (2019) for computational fluid dynamics and by Degen et al. (2022, 2023) for geothermal applications, as well as for geodynamical and hydrological applications (Degen et al., 2023). Furthermore, neural networks (NNs) have been used by Atkins et al. (2016) to infer mantle convection parameters, and by Agarwal et al. (2020) to model the 1-D thermal convection response on Mars and later to characterize its 2-D response (Agarwal et al., 2021). Kerswell et al. (2024) utilize NNs to predict the equation of state of geomaterials, which is a necessary ingredient of geodynamic simulations.

Combining physical process models with data-driven ML techniques, referred to as physics-based machine learning (PBML), allows us to overcome limitations and caveats associated with repeated forward modeling, as stated by Degen et al. (2022). Limitations regarding physics-based forward models include computational challenges and costs that emerge when developing high-resolution models in both space and/or time, handling complex coupled partial differential equations (PDEs), and estimating uncertainties through multiple realizations as discussed by van Zelst et al. (2022). On the other hand, limitations regarding data-driven methods arise from their inability to produce explainable and interpretable models in order to understand the governing subsurface processes (Degen et al., 2023). Therefore, the appeal of PBML in this context is not merely its computational efficiency and all the aforementioned aspects, but also the scalability, interpretability, and robustness of the surrogate models, as well as their ability to provide explanatory insights.

In this study, we explore the potential of using PBML to construct surrogate models that represent the temperature distribution in the mantle as a result of mantle convection and evaluate their accuracy and efficiency. Specifically, we investigate the application of the NI-RB method as a possible implementation of PBML. NI-RB is a model order reduction technique used to approximate high-dimensional PDEs by constructing lower-dimensional models. This is achieved by extracting a reduced basis (RB) from a collection of high-fidelity solutions through a proper orthogonal decomposition (POD) and then using an ML method, particularly NNs, to weight the individual components, which are often referred to as reduced coefficients (Hesthaven and Ubbiali, 2018; Swischuk et al., 2019). This differs from other approaches where physics is informed as by design, as demonstrated, for instance, by Agarwal et al. (2025) for mantle convection applications. Meaning, that the NI-RB method maintains the relationship between the physical properties (e.g., density, viscosity) and the state outputs (e.g., temperature, pressure) and hence the physical laws of the domain (e.g., equations of motion and equation of heat transfer); this is done by combining physical process models with data-driven ML techniques. Furthermore, it also fulfills the criteria to construct

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

reliable, scalable, and interpretable surrogate models of high accuracy that require relatively few basis functions (Degen et al., 2022; Hesthaven and Ubbiali, 2018; Swischuk et al., 2019).

In our study, these high-fidelity state outputs are obtained in the form of numerical temperature distribution snapshots utilizing ASPECT, an open-source geodynamic modeling software (Kronbichler et al., 2012; Heister et al., 2017; Bangerth et al., 2022a). Our simulations consider varying material properties between models, in our case viscosity, resulting in significant differences in the predicted temperature distribution following the simulation of mantle convection over a specified period.

2 Theoretical framework and methodology

In this chapter, we provide a detailed analysis of our approach, thoroughly examining each step and explaining the methodologies employed. First, we discuss the NI-RB method, which is implemented in two stages: the offline and the online stage. In the offline stage, we construct and save the surrogate model. During the online stage, new state solutions are predicted given the previously constructed surrogate model (Benner et al., 2015; Hesthaven et al., 2016; Hesthaven and Ubbiali, 2018; Wang et al., 2019). For this study, we validate and evaluate the surrogate model against the high-dimensional forward solution generated by ASPECT, inspecting the results for inaccuracies introduced by the NI-RB method.

2.1 The Non-Intrusive Reduced Basis Method

In engineering and applied sciences, complex problems often involve parameterized PDEs, where parameters represent various physical properties (Hesthaven and Ubbiali, 2018; Willcox et al., 2021). The challenge lies in efficiently evaluating these PDEs for numerous parameter combinations. High-fidelity numerical solutions are computationally demanding, motivating the use of reduced-order modeling (ROM) methods to reduce computational expenses while preserving the general characteristics of the system (Benner et al., 2015; Degen et al., 2021, 2022, 2023; Hesthaven et al., 2016; Hesthaven and Ubbiali, 2018; Willcox et al., 2021).

RB methods are ROM techniques that aim to construct low-dimensional models of parameterized PDEs that are faster to solve than their high-dimensional equivalents. They are constructed from a set of high-fidelity snapshots (Benner et al., 2015; Degen et al., 2023; Wang et al., 2019) with the aim of obtaining reliable predictions in (near) real-time by taking advantage of the mentioned low-dimensional approximation space. In the context of our study on mantle convection, we can expect RB methods to significantly reduce the computational complexity of simulating the Earth's mantle dynamics by capturing essential features like temperature distribution and flow patterns, allowing for efficient exploration of various convection scenarios without substantial loss of accuracy.

The general idea of the RB method is to represent the solution as a combination of basis functions and coefficients, making it conceptually similar to a traditional Fourier transform approach (Degen et al., 2023). We can consider the Fourier transform process, where a signal, initially in the time domain, is disassembled into the frequency domain to ascertain its components. Within the frequency domain, various manipulations are conducted more straightforwardly, allowing the determination of its component coefficients. Subsequently, the reassembled signal, tailored to desired specifications, can be reverted to the time

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

domain (Oppenheim et al., 2009). Drawing a parallel, the RB method initially engages with the full order model, which is a collection of snapshots, and deconstructs it through, for instance, a POD approach to obtain reduced basis functions. Subsequently, the coefficients are determined through a projection method. This combination of the basis functions with their corresponding coefficients then facilitates the reconstruction of the model, thereby generating the reduced order model, or otherwise called surrogate model. The surrogate model can be used to capture features of the full order model up to a certain degree called tolerance, which is defined during the POD step (Hesthaven et al., 2016; Hesthaven and Ubbiali, 2018).

However, traditional projection-based intrusive RB methods struggle with non-linear parameter dependencies, primarily due to the challenge of efficiently computing the projection coefficients (Degen et al., 2022, 2023; Hesthaven and Ubbiali, 2018). NI-RB methods circumvent this issue by using ML techniques to determine the projection coefficients instead of relying on a Galerkin projection (Hesthaven and Ubbiali, 2018). In this work, we use an NI-RB method with a NN to determine the coefficients. This approach offers a fast and reliable solution for complex nonlinear PDEs with non-linear parametric dependencies as discussed by Hesthaven and Ubbiali (2018), and thus is considered suitable for mantle convection problems.

For the NI-RB method, we distinguish two phases: i) the offline and ii) the online stage. During the **offline stage**, the POD is implemented in the form of singular value decomposition to extract the reduced basis functions that define parts of the model architecture (Quarteroni et al., 2016). This extraction is governed by a user-defined tolerance that determines the level of detail captured by the reduced model (Degen et al., 2023). Not only may the defined tolerance enhance the accuracy of the surrogate model, but it also safeguards against spurious or nonphysical predictions that could emerge due to extrapolation beyond the parameter space that is well-represented by the snapshots. The basis functions focus on the principal characteristics of the full order model in order to meet the previously defined error tolerance, which leads to a reduction of computational costs (Chen et al., 2021). Essentially, the tolerance dictates the trade-off between computational efficiency and approximation quality, as it controls the number of basis functions that are extracted and the fidelity of the surrogate model to the full solution. The input of the POD is a set of high-fidelity solutions, analog to the training data set in ML approaches. Both the training and validation data sets are constructed prior to the offline stage. Therefore the chosen, physics-based PDEs are solved for numerous input parameter realizations, providing the solutions as snapshots. The training data set is adeptly scaled between 0 and 1, optimizing compatibility with the value range most ML algorithms predominantly operate within. Using an external high-fidelity model, we compute these snapshots by varying the input viscosities. Employing snapshots provides a concise representation of the system's behavior, allowing us to bypass the need to repeatedly solve the complex PDEs directly during the surrogate model's real-time application. In contrast, directly incorporating the underlying physics into the data would necessitate solving these PDEs multiple times, consuming more resources (Hesthaven and Ubbiali, 2018; Degen et al., 2022).

After the POD step comes a second stage within the offline phase: the projection step. Here, NNs are implemented as a tool for the search of the coefficients (Wang et al., 2019; Chen et al., 2021; Degen et al., 2022). Unlike the traditional RB methods that typically employ the Galerkin projection, NNs provide means to handle the complexity inherent in the highly non-linear PDEs of our study. Within the NN, data transitions sequentially only from the input layer to the output layer, which is characteristic of a feedforward architecture. This implies that each neuron is only linked to the subsequent layer, ensuring a one-way flow of information without feedback loops (Hesthaven and Ubbiali, 2018). In our work, ML methods are only

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

used as a projection method in order to determine the weights of the basis functions. This has the advantage that the black box problem of ML only occurs in this step. Therefore, the non-explainable approximation errors are related to this weighting step (Degen et al., 2022) only. Within the process of constructing the surrogate model, the main characteristics of the response can be preserved.

Once the model is constructed, a validation data set is used to evaluate the performance of the NN on previously unseen data.

As already established, the offline stage is computationally expensive, attributable in part to the surrogate model construction of the training dataset, since it involves solving numerous complex finite element (FE) forward simulations. Additionally, the determination of the hyperparameters of the NN contributes to the total cost of the offline stage. Efficient strategies, as presented by Hesthaven and Ubbiali (2018); Degen et al. (2022), have been implemented to mitigate these challenges. Firstly, the Latin Hypercube Sampling (LHS) method is employed, which aids in reducing the number of forward evaluations necessary to optimize the efficiency of the training and validation data set production (McKay et al., 1979). Secondly, the Bayesian Optimization with Hyperbands (BOHB) scheme is used, which employs a more systematic search to efficiently find the best hyperparameters and thus accelerates convergence during the tuning process (Falkner et al., 2018).

During the **online stage**, the trained surrogate model rapidly computes the reduced coefficients for any new set of parameters and reconstructs the approximate temperature distribution as a linear combination of the pre-computed basis functions, achieving comparable accuracy to the full-order solver at a fraction of the computational cost (Hesthaven et al., 2016). Therefore, the online phase offers the benefit of drastically faster results, which is crucial when time-sensitive decisions or real-time monitoring are required (Hesthaven et al., 2016; Quarteroni et al., 2016). In this study, we are interested in testing new input parameters, constructing efficient surrogate models, and evaluating their potential for mantle convection. Thus we provide the validation parameters as input to the surrogate model and calculate the difference between the forward ASPECT solutions generated with the validation parameter values and the surrogate model evaluation solutions.

2.2 Mantle Convection Forward modeling with ASPECT

In this study, we employ ASPECT to generate forward temperature distribution models, otherwise called true solutions, full solutions, or snapshots. These models adhere to a simple geophysical setup within a geometrical domain similar to the Earth's mantle, following the ASPECT manual chapter 5.3.1 (Bangerth et al., 2022b). The model's inner and outer radius, mirroring a spherical approximation of the Earth, are set to 3,481 km and 6,336 km, respectively. The temperature at the outer and inner boundaries is set to 973 K and 4273 K, respectively, resulting in a temperature difference of 3300 K across the mantle layer. The equations for conservation of mass, momentum, and energy are solved under the Boussinesq approximation. Rather than differentiating between upper mantle, transition zone, and lower mantle, or considering mantle discontinuities, this model assumes a constant viscosity throughout. Opting for a 2-dimensional quarter shell representation of the Earth's mantle allows for faster computations and thus more extensive experimentation.

This simplicity of the 2-D model allows for experimentation focused on input parameters, primarily the starting viscosity of the model. Various research studies have developed potential viscosity profiles for the Earth's mantle (Ricard and Wuming,

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

1991; Mitrovica and Forte, 2004; Steinberger and Calderwood, 2006; King, 2013; Nakada et al., 2015; Rudolph et al., 2015; Cathles, 2016; She and Fu, 2019). Even though the exact results are not identical between the studies, the average mantle viscosity is possibly in the order of 10^{22} $Pa \cdot s$, while there are significant viscosity differences between lithosphere, asthenosphere, transition zone, and lower mantle, as well as regionally. With this in mind, we choose to restrict our focus to viscosities in the range 10^{22}–10^{23} $Pa \cdot s$ for this study, and we do not assume any lateral or vertical variability in material properties or composition. Thus, we generate forward models using ASPECT with different input viscosities in the specified range and obtain their respective snapshots in the form of temperature distributions. The density ρ_0 is 3,500 kg/m^3, which corresponds to approximate density of mantle rocks near the surface. This value is appropriate since we consider an incompressible model and can therefore neglect the pressure dependency of the density. Furthermore, the density changes with the temperature T as $\rho(T) = \rho_0\,(1 - \alpha(T - T_{\text{ref}}))$, where the thermal expansivity is $\alpha = 4 \times 10^{-5}$ K^{-1} and the reference temperature is $T_{\text{ref}} = 293$ K. Initially, temperature varies linearly with depth, creating an unstable linear density gradient that initiates convective flow. The gravity profile was taken from PREM. All other material properties are kept constant, which results in Rayleigh numbers Ra between approximately $10^6 - 10^7$.

To simulate mantle convection over a geologically significant time-span of 1 billion years, we employ forward models using a simulation domain discretized by a static, uniform mesh of 12,288 active cells—192 along the inner and outer radii and 64 along the thickness of the model—with narrower cells near the inner radius, resulting in a total of 161,540 unknowns. This ensures a high enough spatial resolution across the entire mantle, and while an adaptive mesh would typically enhance grid density around plume features, our uniform mesh effectively captures the global thermal state. The temporal evolution of the system is captured every 10 million years to systematically record the evolution of temperature distributions. This interval strikes a balance between capturing sufficient detail of the mantle's thermal evolution and maintaining reasonable computational costs. However, despite the extensive temporal data generated, we focus exclusively on the temperature distribution snapshots taken at the final time step for the construction of the surrogate models in this study. This approach is adopted to test the boundaries of the method to predict the culmination of a billion years of convection processes, providing a snapshot of the mantle's long-term thermal state while optimizing computational resources. We note that the initial linear temperature distribution in the model in every case initiates a global overturn, which results in a stably stratified mantle. Models with the lowest Rayleigh numbers will not overcome this initial stable stratification and showcase hot and cold downwellings that dissipate quickly instead of initiating global convection, while models with higher Rayleigh numbers will create a well-mixed mantle with plumes that reach the surface and cold downwellings that reach the core-mantle boundary. While the former case is not realistic for Earth today, we include it in order to evaluate whether the created surrogate model will be able to represent and predict the change of the dynamic regime in the forward models.

Finally, training and validation datasets are created in preparation for the NI-RB method. The sampling technique used to generate input viscosity values for the training data set is the LHS method, while a random sampling is used for the validation dataset. This phase involves the generation of 100 LHS training parameter values for the viscosity within the predefined parameter range, resulting in the generation of one training dataset, which contains 100 temperature data sets/distributions, or simply, 100 realizations. We adhere to the conventionally established 10 % rule for the validation data generation (Degen

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

et al., 2023), which results in 10 values for the validation viscosities and one validation data set consisting of 10 realizations. We decided to use only 100 realizations for the construction of the training data set because this presents typical resource limitations for large-scale 3D models (van Zelst et al., 2022). Even for this relatively low number of realizations, we demonstrate that the surrogate model performs well and requires a significantly lower amount of realizations than reported for other ML studies—e.g., Agarwal et al. (2021) trained a convolutional autoencoder plus neural-network surrogate on 10,525 two-dimensional mantle convection simulations, and in another Mars-focused work, Agarwal et al. (2020) used ∼10,000 simulations to train a mixture-density-network emulator for thermal evolution. And comparable to Agarwal et al. (2025), where 94 simulations have been used with a focus on the velocity snapshots.

2.3 Data Scaling and Hyperparameter Optimization

The efficiency and accuracy of ML applications often depend on appropriate data scaling and the selection of optimal model hyperparameters. In this section, we first discuss the scaling method employed, we then provide details on the POD step of the NI-RB method, and lastly, discuss the architecture of the NN used in this study.

Scaling the data appropriately conditions it for improved accuracy and generalization in machine learning models.. To normalize, or scale, the data as part of the preprocessing, we use the training data as a common frame. Scaling the full dataset purely based on the training dataset is standard practice to avoid the creation of a bias; if the data is scaled based on the validation dataset as well, the model will be partly trained on information that it is supposed to be tested on, which will then overestimate the model's performance. We avoid giving the model this extra information about the validation data. A plethora of scaling operations were tested, including min-max normalization, normal scaling, mean normalization, and a mean scaling with a min-max normalization applied afterward. After analyzing all the listed options, we choose the mean scaling, and thus employ the Hyperbolic Tangent (Tanh) activation function for the training process as it outputs values in the range -1 to 1, which is the range we get from the mean scaling. This choice is supported by the analysis shown in Figure 1, where it is clear that the mean scaled data have similar global error ranges compared to the other tested operations, but reliably predicts the location of the plumes while maintaining low error in their vicinity, thus providing higher plume resolution. Lastly, the loss function tested in our study is the mean squared error for all presented results, as it is the default loss function to use for regression problems (Goodfellow et al., 2016).

For our mantle convection example, we introduce a tolerance for the POD step of 10^{-1}. This specific tolerance is selected after comprehensive testing, which yields 67 basis functions for the mean-scaled data set. Although this number of basis functions may seem substantial, especially for a training sample of 100, it is essential to note that we later show that this choice is the most efficient in terms of computational demand during the offline stage of the NI-RB method. The higher number of basis functions, while slightly increasing its computational cost, has facilitated a surrogate model that closely captures the complex dynamics of mantle convection with a dramatically reduced computational cost compared to the forward model. Experiments with a smaller number of basis functions, which is achieved by applying a larger tolerance, substantially reduce the quality of the surrogate model in this study.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

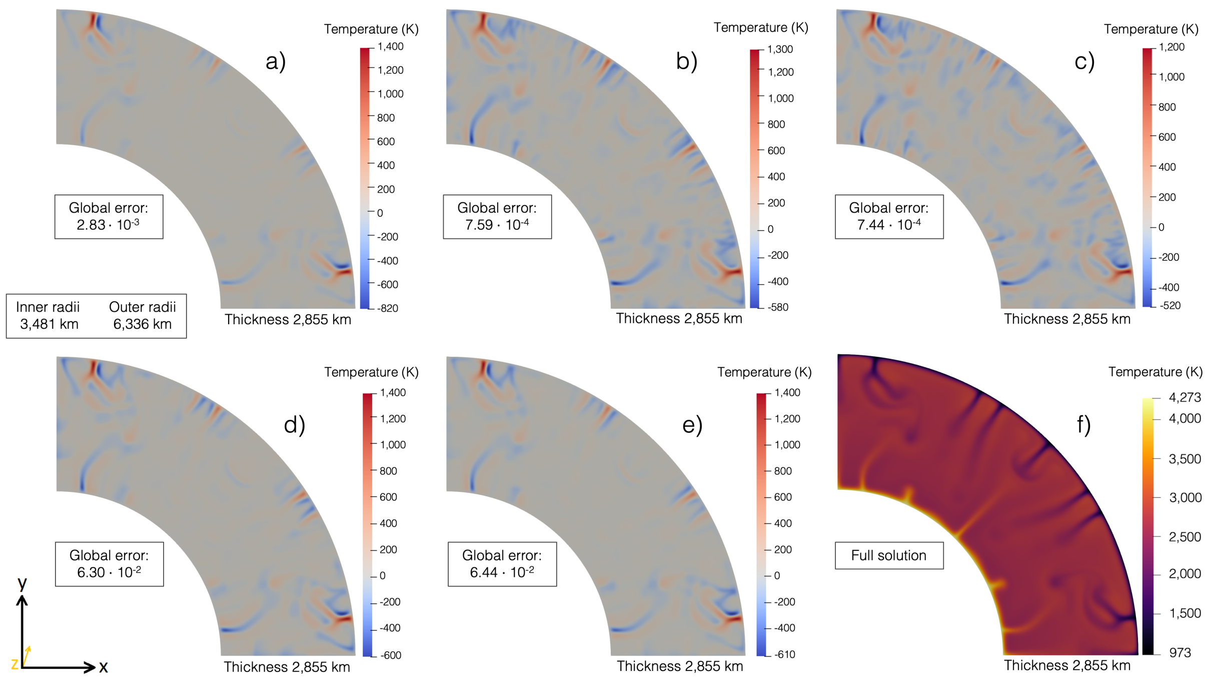

Figure 1. Comparison between different scaling operations for the same realization with viscosity of $1.71{\cdot}10^{22}\ Pa \cdot s$. Each plot is the difference between the ASPECT model and surrogate model, after scaling with a) mean scaling (tanh activation function), b) mean and minmax scaling (sigmoid activation function), c) minmax scaling (sigmoid activation function), d) normal scaling (sigmoid activation function), e) normal scaling (tanh activation function). Panel f) shows the full solution of the temperature distribution.

The hyperparameters are fine-tuned during the NN training to optimize the model's performance. Tab. 1 presents the final set of hyperparameters we adopt after a series of iterative adjustments and evaluations. By implementing these hyperparameters and the chosen POD tolerance, our surrogate model achieves a global mean squared error of $2.83{\cdot}10^{-3}$ (dimensionless, computed on the scaled validation data set), which matches the user-defined POD tolerance of 10^{-1}. The POD has to be applied to the scaled version of the data set. Hence, the provided tolerance is dimensionless. While the global error provides an almost perfect, or at least favorable view of the model, we recognize that it does not fully encapsulate the local dynamics critical to mantle convection simulations. Global error values can sometimes obscure local discrepancies that are crucial in the context of mantle convection, where the spatial distribution of the error is as important as the magnitude of the error itself. To address this, we performed a detailed examination of local errors through difference plots, which offer valuable insights into the model's spatial error distribution and its implications for accurately characterizing mantle dynamics. The difference is evaluated by subtracting the full solutions from the reduced ones. Hence, positive values in the ensuing figures indicate the surrogate solutions' overestimation, while negative values denote their underestimation.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

Table 1. Hyperparameters of the NN, with additional information on the basis functions and global error.

Hyperparameters	
Number of hidden layers	2
Number of neurons per layer	176
Number of epochs	31,281
Learning rate	$1.32{\cdot}10^{-4}$
Batch size	63
Activation function	Tanh
Optimizer	Adam
Basis functions	67
Global error (Q2)	$2.83{\cdot}10^{-3}$

3 Results and Observations

In the following, we present the results of the surrogate model, by firstly evaluating its accuracy. Then, we illustrate the performance dependency on the viscosity, and lastly we demonstrate the computational efficiency of the approach.

3.1 Accuracy of the Surrogate Model

For the discussion of the accuracy of the surrogate model, we focus primarily on the least and the most accurate model evaluations. These represent the two end member cases and illustrate the general approximation quality of the constructed surrogate.

Of the calculated differences for the ten evaluations of the surrogate model, the realization with a viscosity value of $8.27{\cdot}10^{22}$ $Pa \cdot s$, which is the second highest value tested, shows the lowest local errors. Figure 2 presents this ASPECT forward model compared against the surrogate model evaluation for the same viscosity, as well as their local differences. The model comparison has a maximum absolute local error of 54 K, which corresponds to approximately 1.64 % of the temperature difference (3300 K) across the model. Notably, this evaluation captures both the plume characteristics (shape, size, number, location, etc.) and the state of the mantle (remaining close to the initial linear temperature distribution) accurately. This is not only the case for this evaluation, but for all evaluations with higher viscosities in the range of (2.51–9.90)$\cdot10^{22}$ $Pa \cdot s$, as can be seen in Figures 4 and 5. This trend will be discussed later.

On the other hand, the maximum local errors are exhibited by the realization with viscosity value $1.51{\cdot}10^{22}$ $Pa \cdot s$, which is one of the lowest viscosities tested. This scenario is illustrated in Figure 3. The model evaluation shows a maximum absolute error of 1,400 K, which corresponds to approximately 42 % of the temperature difference across the model, or almost the full temperature difference between a plume and the background mantle. This is a consequence of mispredicting (even slightly) the plume locations. While this evaluation captures the characteristics of some plumes, specifically the two central plumes, one warm and one cold, it mispredicts the position of other hot plumes slightly, and the number and location of most cold

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

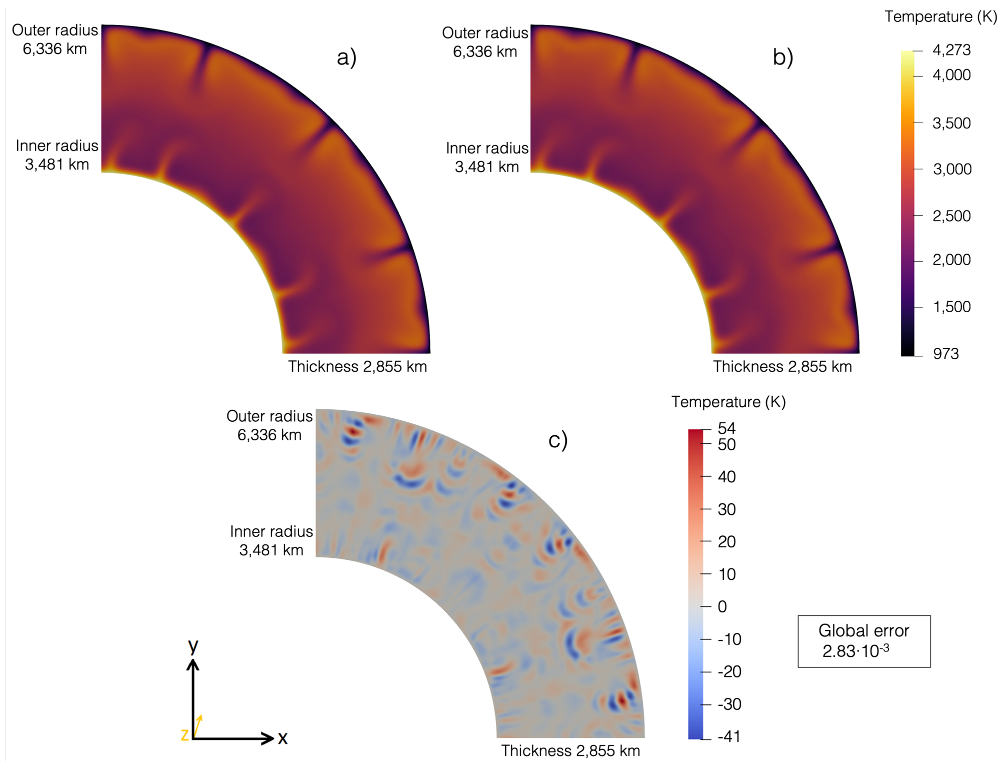

Figure 2. Demonstration of a) full ASPECT solutions, b) reduced surrogate model solutions, and c) their difference for the temperature distribution of realization 6 with a viscosity of $8.27 \cdot 10^{22}$ $Pa \cdot s$. This is considered the most accurate surrogate model evaluation.

downwellings significantly. Nevertheless, it accurately predicts the number of hot plumes (5) and while it mispredicts the number of cold downwellings (5 in the forward model, 11 in the surrogate) the predicted cold downwellings are weaker, therefore approximately maintaining the amount of heat flux in the system. Additionally, the surrogate model accurately captures the significant change of the thermal equilibration of the background mantle that results in an almost uniform temperature profile outside of upwellings and downwellings. Three more realizations exhibit similar behavior with similarly high local errors for equally low viscosities, which can be seen in Figure 6. All of the other low viscosity evaluations show similar characteristics: The most significant change from a linear temperature distribution to a constant mantle temperature is captured in all models, the number and strength of plumes is predicted approximately, some consistent features like the central plumes are predicted reliably, and the exact location and shape of other plumes is not captured very well.

Lastly, it is worth mentioning that the four realizations that showed the highest local errors all have viscosities between (1-2)$\cdot 10^{22}$ $Pa \cdot s$, and any intermediate or low errors only come from surrogate model evaluations with viscosity values higher than the aforementioned. These findings are consistent with general expectations from fluid dynamics: Models with lower viscosity have a higher Rayleigh number and therefore feature faster flow—resulting in more convective overturns over the same model time—as well as thinner and more numerous upwellings and downwellings. The resulting more complex and dynamical system

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

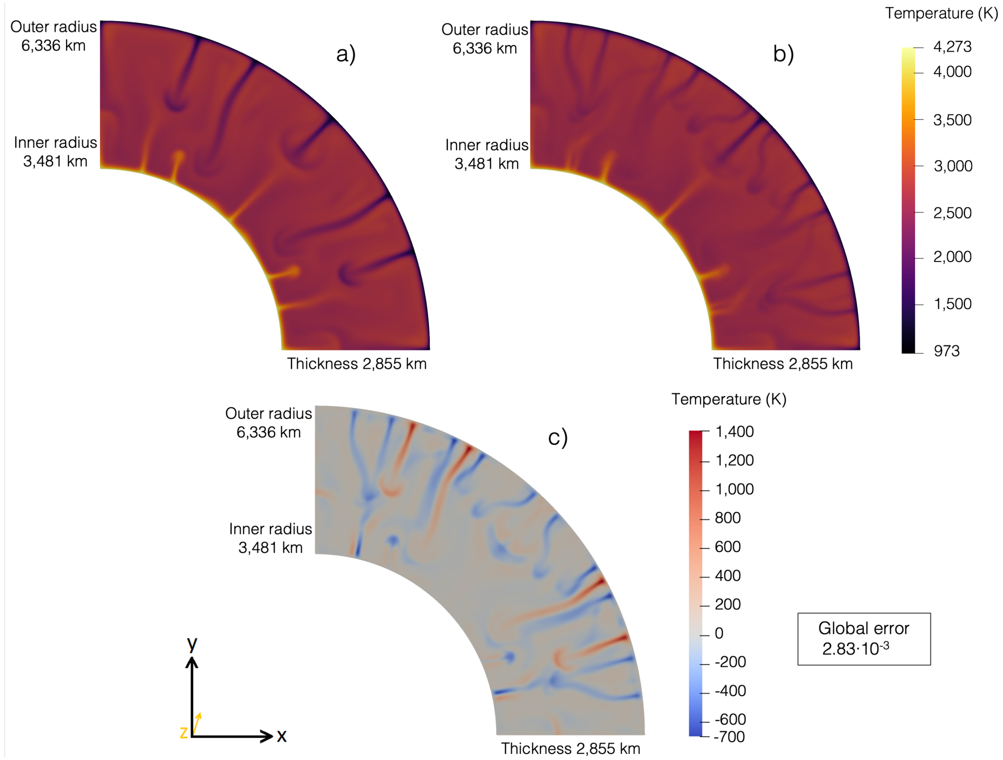

Figure 3. Demonstration of a) full ASPECT solutions, b) reduced surrogate model solutions, and c) their difference for the temperature distribution of realization 5 with a viscosity of $1.51{\cdot}10^{22}$ $Pa \cdot s$. This is considered the least accurate surrogate model evaluation.

then shows higher local errors. However, we believe that the threshold for system instability can be more precisely defined, since most surrogate model evaluations yield local errors below 500 K and appear to be more stable and accurate. To this end, we partition the 10 surrogate model evaluations into three categories, comparing each with the other evaluations as well as the entire training data set, which not only contains more samples but also has a denser sampling of viscosity values. This process is further discussed in the following section.

3.2 Surrogate Model Viscosity-Dependent Dynamics and Performance

We observe a significant trend concerning local errors in the surrogate model evaluations related to mantle convection processes. Specifically lower viscosity model evaluations are associated with higher local errors and vice versa. Furthermore, beyond a certain threshold, the local errors predominantly concern the characteristics of the plumes as opposed to concerning their position, as observed at higher viscosities. To systematically analyze this observation, we categorize the surrogate model evaluations into three distinct regimes based on the viscosity-error relationship. Note that this categorization is solely done to illustrate the different error behavior for various viscosity values. In reality, the transition between the regimes is continuous, where a lower viscosity yields a higher number of plumes and larger temperature deviations from the initial temperature pro-

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

file. This initial examination and categorization is primarily guided by the number and position of plumes and downwellings, as well as the 1D temperature profile. However, we acknowledge that a purely visual classification is subjective and not easily reproducible. Therefore, after introducing these three regimes, we will present a more quantitative measure, namely the entropy generation number, to characterize the thermodynamic behavior of each regime. Starting with the approach based on the amount and position of the plumes, we sort the forward training realizations into each of the identified three regimes. By performing this categorization, we can assess the amount of the training data contributing to each regime of viscosities. This helps to provide insights into the influence of various input viscosity values and the associated dynamic behaviors of the system and into potential weaknesses of the NI-RB method for this application.

The three regimes are as follows:

- Regime 1 represents the models with the most sluggish convection, including three out of ten surrogate model evaluations from the validation data set. This regime encompasses a spectrum of dynamic states within its three evaluations, shown in Figure 4. Scenarios a-b (and e-f) exhibit less complex dynamics compared to scenario c (and g), portraying a more advanced dynamic state and higher local errors. Nevertheless, the surrogate model accurately predicts the varying cases, demonstrating its robustness across different convective patterns within the same regime. This ability to discern and replicate multiple scenarios within a single regime is also observed in Regime 2 as is discussed later, confirming the model's reliability in capturing the nuances of the system's behavior. All three scenarios in this regime share mostly common dynamics and geometries; they are characterized by the presence of five primary warm plumes originating from the inner boundary of the mantle, and three primary cold downwellings descending from the outer boundary. With decreasing viscosity, as seen in scenario c (and g), we observe more cold downwellings and the early formation of new hot plumes, other than the primary ones that are present in all realizations. Additionally, these evaluations depict a distinct colder lower mantle and a warmer upper mantle, consistent with the inverted linear temperature profile established after the initial overturn. Among the training realizations, 34 out of 100 display substantial visual similarity with this regime, featuring viscosity values in the range of (4.51-9.90)$\cdot 10^{22}$ $Pa \cdot s$. Note that for the training realizations, the viscosity values that were tested span (1.01-9.90)$\cdot 10^{22}$ $Pa \cdot s$, while for the validation data set these values span $1.04 \cdot 10^{22} - 10^{23}$ $Pa \cdot s$; hence, the surrogate model is not actually trained on these very high viscosities of the validation data set ($9.90 \cdot 10^{22} - 10^{23}$ $Pa \cdot s$) that exist outside the training data set. Nonetheless, all evaluations from this regime demonstrate high accuracy, with local errors spanning from 54 to 200 K, as is visible in Figure 4 g-i.

- Regime 2 represents models with intermediate convection vigor, encompassing three out of ten surrogate model evaluations from the validation data set, depicted in Figure 5 a-f. Evaluations represent different stages of convective evolution, yet they share characteristic features: three to five primary cold downwellings descend from the outer boundary, and three to five warm plumes rise from the inner boundary. In these scenarios, the cold downwellings often extend deeper into the mantle when compared to Regime 1, sometimes interacting with the warm plumes, which hints at the complex interplay between thermal anomalies in ongoing convection. Despite the variations in convective patterns, the underlying consistency in plume structure across the scenarios is evident. The mantle's temperature distribution becomes more

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

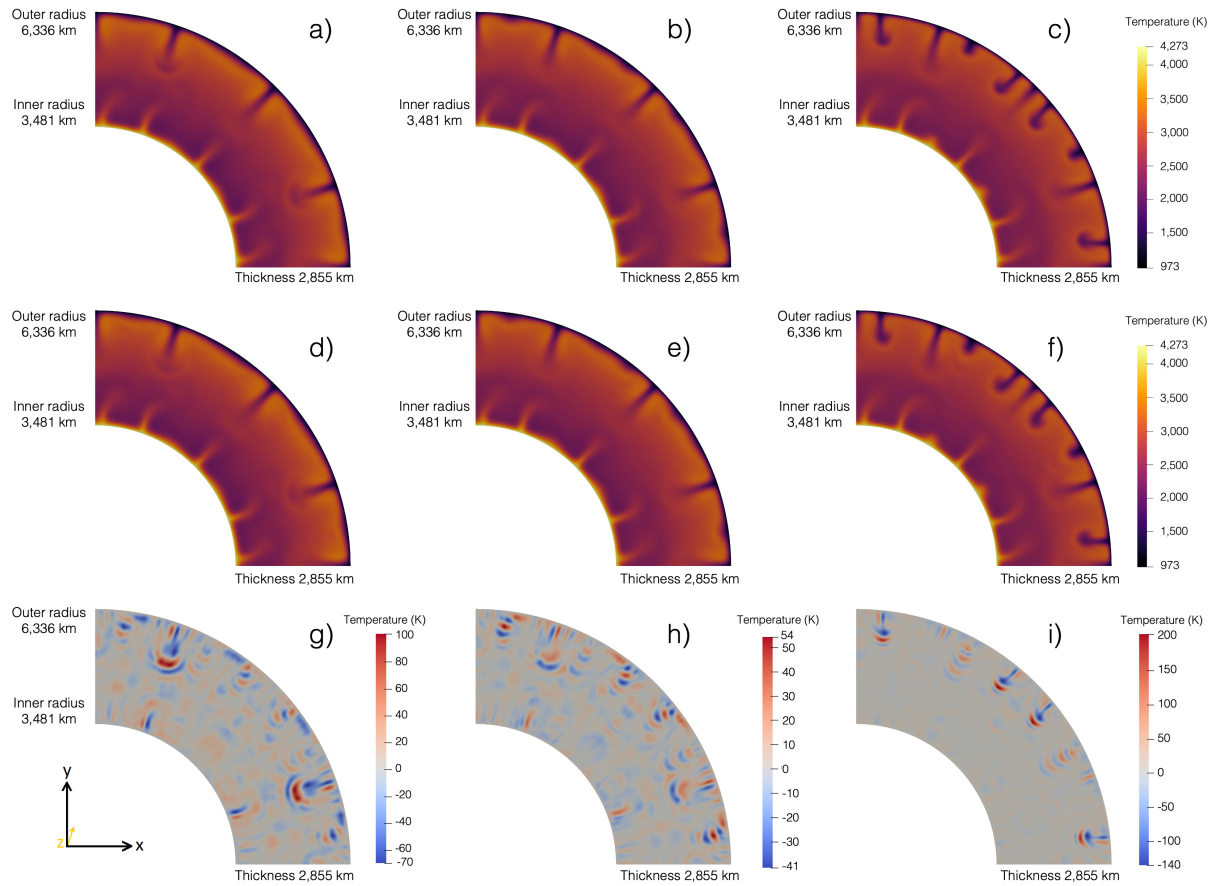

Figure 4. Regime 1, comprising of three surrogate model evaluations. First row a-c demonstrates the full ASPECT solutions, where: a) realization 7 with a viscosity value of 1.00×10^{23} $Pa\cdot s$, b) realization 6 with a viscosity value of 8.27×10^{22} $Pa\cdot s$, and c) realization 4 with a viscosity value of 6.46×10^{22} $Pa\cdot s$. Second row d-f demonstrates the respective surrogate model evaluations, and third row g-i shows their respective local errors.

homogeneous, losing its linear gradient and becoming thermally mixed, suggesting a transitional state in the convective process. From the training data set, only 24 out of 100 realizations visually align with Regime 2, a number that, while smaller than for Regime 1, still contributes to a surrogate model with robust predictive capabilities for this range of dynamics. The local errors for this regime remain comparably low, ranging from 150 to 460 K as seen in Figure 5 g-i, which confirms that the NN is sufficiently trained for the associated viscosity values, spanning (2.51-4.50)$\cdot10^{22}$ $Pa\cdot s$. The diversity within Regime 2's scenarios underscores the surrogate model's versatility in capturing different stages of evolution without compromising the distinctive features of mantle convection behavior.

– Regime 3 represents the models with the most vigorous convection, comprising four out of ten surrogate model evaluations from the validation data set, as illustrated in Figure 6 a-h. These evaluations pose a greater challenge due to their

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

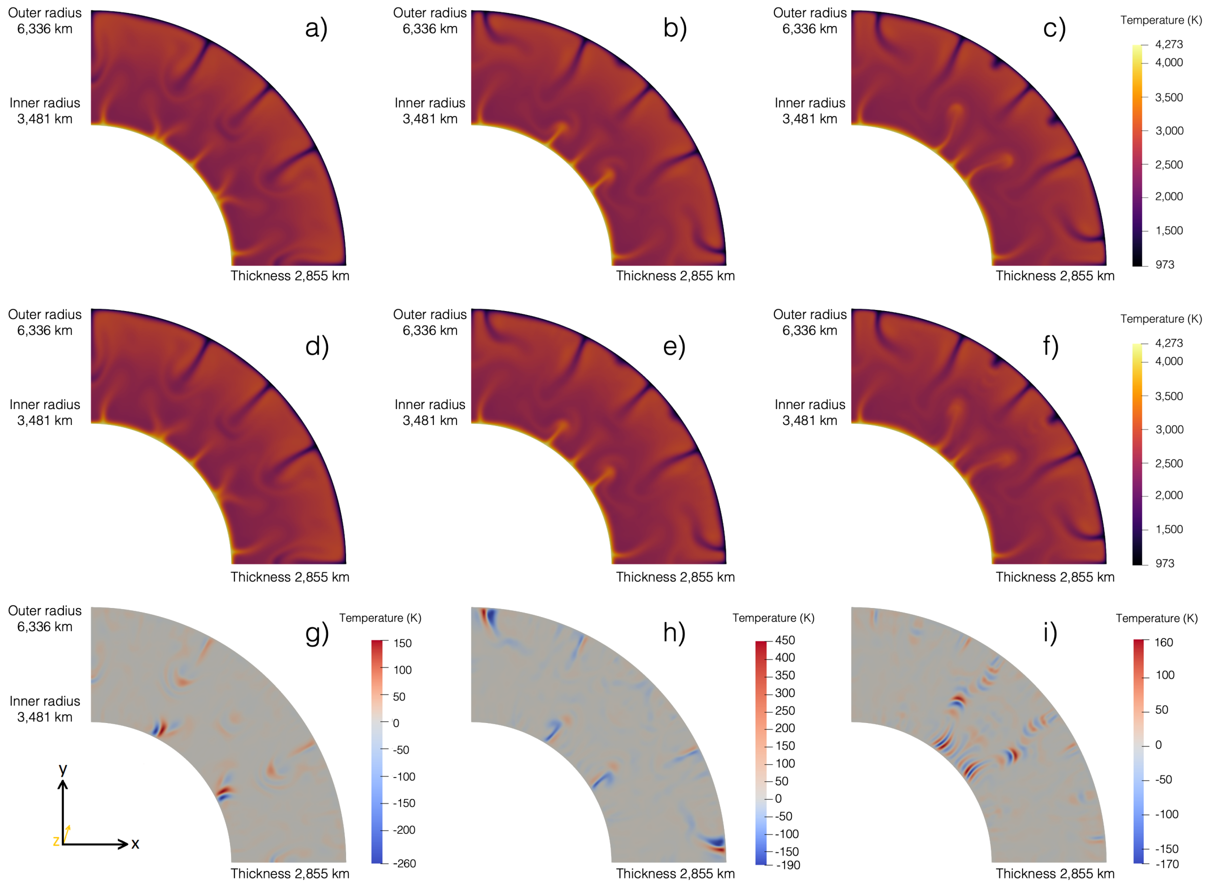

Figure 5. Regime 2, comprising of three surrogate model evaluations. First row a-c demonstrates the full ASPECT solutions, where: a) realization 8 with a viscosity value of $3.74{\cdot}10^{22}\ Pa \cdot s$, b) realization 0 with a viscosity value of $3.21{\cdot}10^{22}\ Pa \cdot s$, and c) realization 1 with a viscosity value of $2.96{\cdot}10^{22}\ Pa \cdot s$. Second row d-f demonstrates the respective surrogate model evaluations, and third row g-i shows their respective local errors.

complexity, particularly in terms of distinct plume characteristics. Despite this, a consistent feature emerges: a robust warm plume ascends from the center of the inner boundary, with a corresponding cold downwelling from the outer boundary directly above it. This interplay of plumes in the surrogate model aligns well with the full ASPECT solutions, exhibiting minimal local errors in these key regions. The evaluations further depict a mantle that is thermally well-mixed, lacking any clear separation into distinct warm or cool regions, mirroring the behavior observed in the full-order models. The absence of other distinct plume formations and the presence of larger local errors, which span between 1,200-1,400 K as is visible in Figure 6 i-l, categorize these evaluations as the least accurate within our surrogate model. Notably, 40 out of 100 training realizations correspond visually to this regime, indicating a substantial contribution to the model's NN learning phase and suggesting a more extensive training coverage for this lower viscosity range, between (1.00-

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

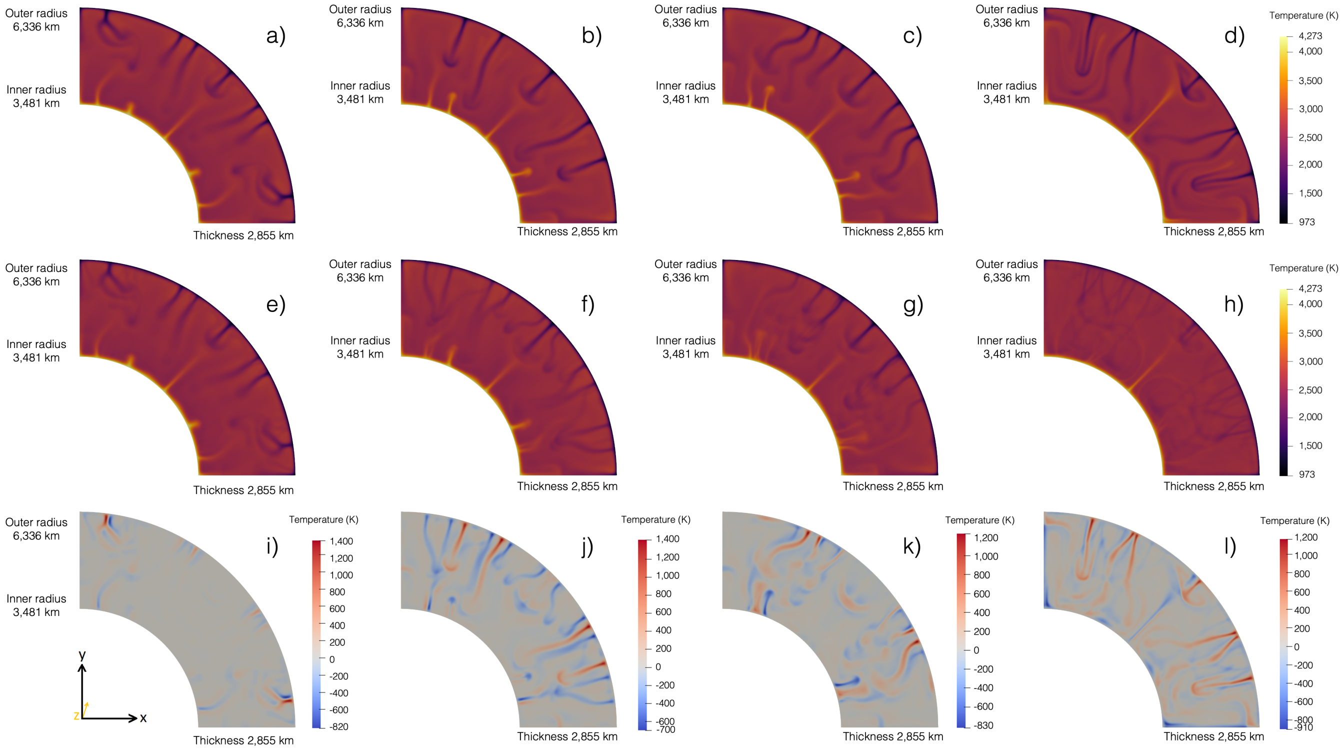

Figure 6. Regime 3, comprising of four surrogate model evaluations. First row a-d demonstrates the full ASPECT solutions, where: a) realization 3 with a viscosity value of $1.71^{22}\ Pa \cdot s$, b) realization 5 with a viscosity value of $1.51^{22}\ Pa \cdot s$, c) realization 9 with a viscosity value of $1.39^{22}\ Pa \cdot s$, and d) realization 2 with a viscosity value of $1.04^{22}\ Pa \cdot s$. Second row e-h demonstrates the respective surrogate model evaluations, and third row i-l shows their respective local errors.

2.50)$\cdot 10^{22}\ Pa \cdot s$. Therefore, it is safe to assume that the lower performance of the surrogate model within Regime 3 stems solely from the higher dynamics and complexity of the system, which is associated with lower viscosities, rather than a shortfall in training or model design.

3.3 Computational Efficiency of the NI-RB Method

The NI-RB's great advantage is its capacity for rapid evaluations of the surrogate model in the online stage, presuming a sufficient surrogate model quality and hence accuracy. Thus, the computational cost for the final methodology is expected to be low. A typical measure in the NI-RB community for evaluating computational time gain is the 'speed-up' metric (Benner et al., 2015; Hesthaven et al., 2016; Degen et al., 2023), which contrasts the computational time of the forward simulation against that of the online stage while disregarding any offline stage costs. The speed-ups for this application are indicated as orders of magnitude but a few clarifications must first be made concerning their estimations.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

A total of 110 ASPECT forward simulations were executed to generate the results presented in this work: a training data set consisting of 100 simulations and a validation data set consisting of 10 simulations. Due to significant variations in the input viscosity parameter values, ASPECT computational times for the forward simulations varied significantly, ranging from 6–115 core hours (1.5-23 hours, $5.4{\cdot}10^6$ ms - $8.3{\cdot}10^7$ ms) for a simulation time of 1 billion years. Consistent initial and boundary conditions were maintained across all simulations, and all models were conducted on a university cluster powered by Rocky Linux 8 OS with an Intel Skylake Platinum 8160 CPU clocked at 2.1 GHz, with 48 cores per node, 2 sockets per node with 24 cores each, and approximately 4 GB main memory per core. By default, each simulation employed 4 cores and 4 GB of memory; however, under certain circumstances, very few simulations employed up to 5 cores and/or 10 GB of memory per core, specifically the ones with wallclock times of 23 hours.

For the PBML aspect of the study, Python version 3.9.16 was utilized on a machine running MacOS Ventura. The machine utilized an Apple M1 chip with an 8-core CPU (4 performance cores at 3.2 GHz), along with 8 GB of memory. The ML framework was implemented using TensorFlow and Keras, at versions 2.19.0 and 3.10.0, respectively.

When discussing the cost of the offline stage of the NI-RB method, it is essential to consider the hyperparameter tuning aspect, which can greatly affect the performance of algorithms, such as the NN, and can introduce significant variability in computational demands. In this study, the hyperparameter optimization requires less computational time than a single forward evaluation of the full-order model, which is also observed in other application fields (Degen et al., 2023). This implies that the cost of the offline stage is mainly determined by the construction of the training and validation data sets, which corresponds to about 110 simulations. We provide the computational cost of the offline stage as the amount of forward evaluations since the generation of the data sets is fully parallelizable, yielding significant variations in the computational cost depending on the specific computational resources and parallelization strategies employed.

The computational time gain has been calculated by comparing one forward simulation to one single evaluation of the surrogate model. However, we account for a range of values due to big variations in forward model computational times, as we judge that calculating an average time from all 110 simulations is not representative or even insightful. It is noteworthy that since the forward modeling process permits parallelization, up to 50 simultaneous simulations were active, expediting the process significantly (of course not influencing the total cost of the forward models).

The final performance measurements yield, as expected, a drastic computational time advantage for the surrogate model online stage, which only requires 223 ms CPU time to generate reduced solutions for different evaluations. This online stage time includes the model loading time of 192 ms, which refers to the duration needed to load the surrogate model into memory before it can start making predictions. Of course, the surrogate model needs to be loaded only once, even for a large number of realizations. Therefore, for a large parameter study, the influence of the loading time becomes negligible and we can exclude the model loading time from the online stage cost, thus only needing to consider the time for the model to execute once loaded, which is 31 ms. Therefore, for the first model prediction, we obtain a speed-up of four to five orders of magnitude compared to the forward model, and for all following model predictions the speed-up increases by another order of magnitude. Since we typically need many model predictions, we achieve overall speed-ups of five to six orders of magnitude compared to traditional model evaluations.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

4 Discussion

In this study, we have detailed the application of the NI-RB method, a PBML technique for developing accurate and efficient surrogate models tailored to the complex demands of geodynamic modeling of mantle convection. The challenges inherent to this high-dimensional and nonlinear domain necessitate innovative approaches, which our implementation of the NI-RB method addresses.

Central to our method is the POD step, which effectively extracts the physical behavior of the system. In our case, constructing 67 basis functions for a training data set consisting of 100 realizations leads to an increased computational load during the offline stage; this number of basis functions is notably higher than what is typically required for simpler or more regular NI-RB applications: for instance, Chakir and Hammond (2018) report approximately 10-35 basis functions for an elastoplasticity problem in geotechnics. However, this investment yields a surrogate model that is cost-effective during the online stage, providing rapid predictions—at only 31 ms per evaluation once loaded—that are accurate under a wide range of viscosities, corresponding to speed-ups of five to six orders of magnitude compared to the full-order solver.

Subsequently, ML is employed in the form of NNs to assign weights to the extracted characteristics from the POD. Here, we acknowledge that the true measure of a model's performance extends beyond the NN global error metrics. Our surrogate model exhibits global mean squared errors in the order of 10^{-3}, well within our set POD tolerance of 10^{-1}. However, to comprehensively assess the model's accuracy, especially given the dynamic nature of mantle convection, we relied on the analysis of local errors. Difference plots were employed to analyze the local errors between the surrogate and the full-order model solutions across varying viscosities. These plots highlight a decreasing accuracy of the surrogate model with decreasing viscosities; this indicates that the global error metric alone might not be sufficient for the surrogate's fidelity, especially when characterizing the dynamic processes involved in mantle convection. Therefore, while our surrogate model achieves global errors well within our target range, indicating high accuracy from a broad perspective, the local error analysis paints a more complex picture. In order to better understand the local errors, we analyze the statistical entropy in the models, which is discussed in the next section.

To summarize, the advantages of the NI-RB compared to other PBML methods boil down to two points: Firstly, the NI-RB method requires considerably less data as input to the POD step than data-driven methods, because it leverages the underlying physics rather than treating it as mere data (Degen et al., 2022). Secondly, it localizes the approximation errors to the weighting process only, thereby preserving the main characteristics of the response (Degen et al., 2022). The advantage of data efficiency cannot be overstated, given that generating data through forward simulations is extremely resource intensive—a characteristic common to geodynamic computations in particular (van Zelst et al., 2022), and geoscientific applications in general—even when parallelizing the generation of the forward simulations.

The observed trade-off underscores the surrogate model's high accuracy with 100 samples in viscosity regimes 1 and 2, while also suggesting the potential for greater detail and precision at lower viscosities with an increased sample size. It is important to recognize that the performance variations, particularly at these lower viscosities, may also be influenced by the linear solver of ASPECT. Even though we consider only the final time step for the construction of the surrogate model, the mantle

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

convection problems need to be solved for many time steps to retrieve the desired result after one billion years. Consequently, the numerical errors produced by the solver for each time step accumulate over time. This has a potentially higher impact on the lower viscosity realizations since they require more time steps. For example, for realization 7 with viscosity $1.00^{23}\ Pa \cdot s$ (highest viscosity tested for the validation data set), ASPECT performed 673 time steps, whereas for realization 2 with viscosity $1.04^{22}\ Pa \cdot s$ (lowest viscosity tested), ASPECT performed 1799 time steps; all realizations that we have categorized in Regime 3 consist of more than 1350 time steps, while for Regimes 1 and 2 we stay within 673-1055 time steps for viscosities down to $2.96 \times 10^{22}\ Pa \cdot s$. In addition, small differences at an early stage can lead to very different model evolutions (Colli et al., 2015, and references therein), which could further increase the problem. This variability points to the inherent numerical challenges that solvers face when tackling complex geodynamical simulations, an aspect that should be considered when interpreting the surrogate model's fidelity.

It is also important to highlight the distinct relationship between viscosity levels and the corresponding temperature distributions that was already observed in Section 3. The observed transitions between different dynamic regimes at varying viscosity values are consistent with fundamental fluid dynamics principles. Lower viscosities naturally lead to more vigorous convection and accelerated thermal mixing, while higher viscosities yield more stable, less dynamically complex configurations. Remarkably, even amid the heightened complexity and dynamic conditions present in Regime 3, our surrogate model still accurately reflects key system characteristics, such as the mantle mixing and the primary plumes and downwellings. In certain instances, the model also captures the position, shape, and number of secondary plumes with a reasonable degree of accuracy. By accurately capturing this expected behavior, the surrogate model demonstrates its reliability and predictive capability. Moreover, these results highlight that the model is well-suited not only for reproducing known convection patterns but also for providing insights into the sensitivity of mantle dynamics to viscosity variations.

4.1 Entropy Generation: Classification and Sampling

The categorization of the training data set into three distinct regimes is pivotal for understanding the different behaviors exhibited by the surrogate model across these segments. Such a differentiation allows for insights into why the surrogate behaves differently under each regime and sets the stage for potential enhancements in the sampling strategy, such as focusing on the regimes and sub-regimes of higher local errors for future studies. Following the visual segmentation, the emphasis is placed on providing a mathematical framework to give substance to this subdivision, which is anticipated to offer a more objective and reliable classification of the regimes. This is particularly important, as it will enable us to later demonstrate that different levels of entropy production are present within each regime, aligning with the varying plume behaviors observed previously.

Traditionally, the Rayleigh number Ra has been a well-established metric for quantifying the convective stability within the Earth's mantle, influenced by variables like fluid density and viscosity. Critical thresholds of Ra dictate the onset and nature of convective behaviors, ranging from mild to vigorous turbulence, in response to underlying thermal gradients (Condie, 2016; Turcotte and Schubert, 2014). However, Ra focuses on the input parameters and would, in more complex scenarios, not account for non-constant material properties. To focus on the resulting temperature and pressure distributions, our study opts to focus on entropy as a more comprehensive thermodynamic indicator. Entropy encompasses not only the heat transfer

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

within the system but also the fluid flow, offering a dual perspective that the Rayleigh number does not provide. This dual perspective is advantageous, particularly when considering the geometry of the system and its ability to offer a distribution of the entropy in the entire domain, thereby allowing for a more nuanced understanding of the convection process. Essentially, entropy serves as a thermodynamic metric that quantifies the irreversibility of a system, providing insights on the efficiency of heat and fluid flow processes (Börsing et al., 2017). The utility of entropy production as a diagnostic tool for surrogate modeling in convection-dominated systems—including the delineation of different dynamical regimes—has similarly been demonstrated in geothermal fault systems by Simader et al. (2026). With this study, we create a mathematical basis to identify areas necessitating enhanced sampling, e.g., in regions of higher entropy, which is instrumental for refining the accuracy of our surrogate model.

To mathematically quantify the entropy production in a 2-D mantle convection system as a first step, we consider both the irreversible component of heat transfer and the dissipative consequences of viscous fluid dynamics, as suggested by Börsing et al. (2017), where the viscous component is adapted in accordance with Dannberg et al. (2022). These contributions can be expressed mathematically as follows in equations 1 and 2 respectively:

$$\dot{S}''_{therm} = \frac{\lambda_m}{T_0^2}(\nabla T)^2, \text{ and} \tag{1}$$

$$\dot{S}''_{visc} = \frac{\eta}{T_0}\dot{\varepsilon}^2, \tag{2}$$

where λ_m $(Wm^{-1}K^{-1})$ is the thermal conductivity of the medium, η $(Pa \cdot s)$ the dynamic viscosity, $\dot{\varepsilon}$ $(1/s)$ the effective strain rate, and T_0 (K) the reference temperature. The reference temperature is calculated as $T_0 = (T_c + T_h)/2$, where T_c is the temperature of the cold boundary and T_h the temperature of the warmer boundary. Assuming that both components are significant for our study, the overall entropy production for the system can be expressed after Börsing et al. (2017) as

$$\dot{S}''_{total} = \dot{S}''_{therm} + \dot{S}''_{visc}. \tag{3}$$

Instead of merely computing the average of the entropy production of a model, a more insightful approach is to determine the entropy generation number. This metric does not only provide a dimensionless representation of the entropy, but also incorporates normalization based on thermal properties. Again, referring to the work by Börsing et al. (2017), the entropy generation number can be formulated as

$$N_{\dot{S}} = \frac{\dot{S}_{total}}{\left(\frac{q^2}{\lambda_m \cdot T_0^2} \cdot A\right)}, \quad \text{with } \dot{S}_{total} = \int_A \dot{S}''_{total}\, dA \tag{4}$$

where A is the area of the system. The bulk heat flux between the model's upper and lower boundaries is denoted as q. Moreover, for the purpose of this paper, the term 'entropy generation' will be synonymous with the entropy generation number.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

Building upon these concepts, we systematically compute the reduced entropy generation $N_{\dot{S}_{red}}$ for each realization in the training and validation data set, which is defined as:

$$N_{\dot{S}_{red}} = \frac{N_{\dot{S}} - N_{\dot{S}_{min}}}{N_{\dot{S}_{max}} - N_{\dot{S}_{min}}}. \quad (5)$$

Here $N_{\dot{S}_{min}}$ is the minimum entropy generation number of the entire data set (both training and validation), and $N_{\dot{S}_{max}}$ the maximum value. The computed reduced entropy generation numbers are illustrated in Figure 7, demonstrating the variability across the different training scenarios.

Observing the results, the analysis of entropy generation reveals three notable patterns within the data, as captured in Figure 7, which align with the visual established regimes. These patterns, or entropy regimes provide a framework for understanding the system's behavior:

1. Regime 1, depicted in yellow, corresponds to lower entropy generation associated with higher viscosity values, signifying a more stable system state that has not yet overcome the initial condition.

2. Regime 2, represented by orange dots, denotes an intermediate state of entropy and viscosity, suggesting a moderately dynamic system.

3. Regime 3, marked in red, features a rapid increase in entropy generation correlating with low viscosity values, indicating a high and rapidly changing level of system dynamics. The red points show a distinct cluster formation contrary to the more continuous distribution of the illustrations of the other two regimes. Notably, the entropy-based analysis further illustrates that Regime 2 may serve as a transitional range between the stable, low-entropy states of Regime 1 and the highly dynamic, high-entropy states of this regime.

The entropy regime boundaries were determined by visually inspecting the complete set of training data plots and matching the observed ranges of reduced entropy with the regime descriptions, as was discussed in section 3.2—no strict quantitative thresholds were applied. Although these regimes offer a broad characterization of the system, they also underscore the presence of finer variations within each regime. For instance, within Regime 2, we observe sub-regions with higher entropy generation, showing states of increased dynamics in the system. This is in accordance to the expected higher efficiency of convective heat transfer away from boundaries between different regimes and was previously also described in Börsing et al. (2017) and Huang and Wellmann (2021).

The observation of variations of entropy production within a regime also suggests that focusing solely on the regimes does not provide the full picture of the convective state of a system. Instead, it may be more beneficial to concentrate on the thermodynamic states separately and as a continuum. Such an approach could enhance our strategy for generating additional training samples, ensuring that we capture the nuanced behavior of the system rather than its segmented representation. Hence, future sampling efforts may benefit from targeting specific thermodynamic states that correspond to specific viscosity values, which are indicative of the system's intricate behavior, rather than broad regime categories. Ultimately, the results in Figure 7

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

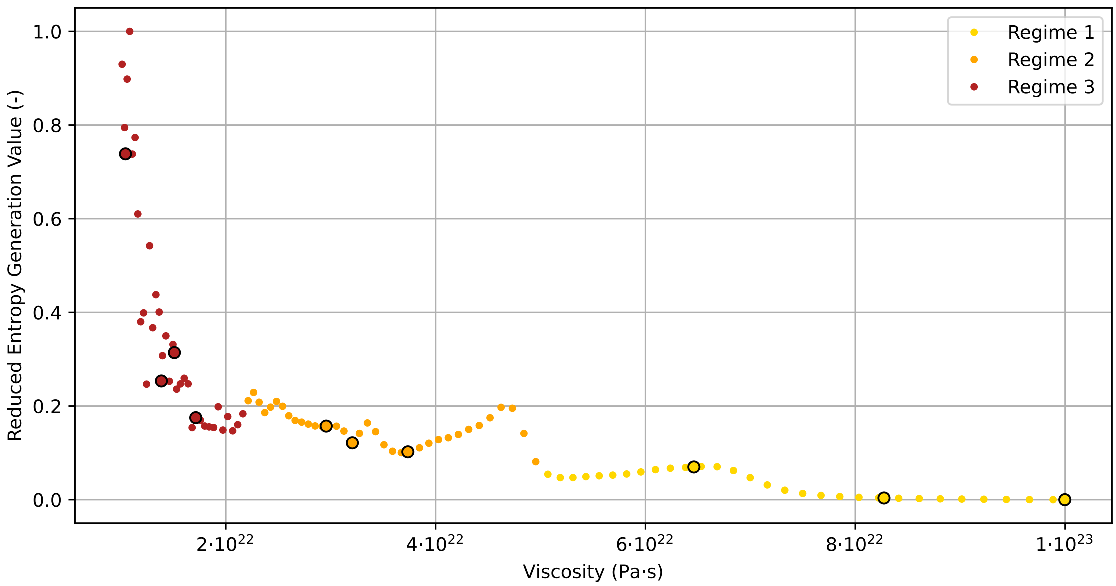

Figure 7. Reduced entropy generation values for all 100 training realizations with different input viscosity values, colored by the identified convection regimes. In addition, the entropy generation values for all 10 validation realizations are plotted in larger markers with black borders.

accentuate the potential superiority of entropy as a metric over the initial visual method used to split the realizations into the primary three regimes.

It is also interesting to note that the viscosity ranges within each regime seem mostly in agreement when comparing the visual classification and the entropy-based one. This observation is further supported by Tab. 2, where the viscosity range of each regime is shown for both classifications.

Moreover, particularly in Regime 3, which is characterized by lower viscosities, there seems to be a diverse spectrum of entropy values, which suggests that prioritizing additional sampling for these highest entropy values would be crucial in order to achieve greater accuracy in the surrogate model for the corresponding viscosities.

Overall, we believe that the entropy generation serves as a valuable indicator, which can be used to enhance sample diversity and aid in determining how many training samples are required. The high entropy generation values observed are attributable to low viscosities, a correlation previously observable both visually and logically, and now reaffirmed through entropy generation considerations.

While the entropy analysis simply confirms visually observed trends, it establishes a quantitative methodology that could be applicable in interpreting complex, multi-parameter systems; it is the approach's generalizability that stands out as particularly valuable. We believe an entropy analysis could explain the relationship between the surrogate model's performance and the

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

Table 2. Viscosity values of the 3 regimes for the visual and entropy-based subdivision methods.

Regime	Viscosity from visual subdivision ($\cdot 10^{22}\ Pa \cdot s$)	Viscosity from entropy subdivision ($\cdot 10^{22}\ Pa \cdot s$)
1	4.51 - 9.90	5.01 - 9.90
2	2.51 - 4.50	2.21 - 5.00
3	1.00 - 2.50	1.00 - 2.20

entropy generation not only for our application but also when dealing with non-linear parameter interactions, for example, simultaneous variations in both initial viscosity and density. As such, applying entropy analysis to a multi-parameter challenge represents a compelling next step, offering a potential key to unlocking the intricacies of the surrogate model's performance.

4.2 Benefits of Surrogate Models

So far we have discussed how to efficiently construct a surrogate model for mantle convection studies. However, the key discussion point is when a surrogate model is advantageous and when the high dimensional model evaluations are preferable. Surrogate models demonstrate their highest benefits in many-query or real-time applications, i.e., in cases where either the forward model needs to be evaluated many times or a response is required in a short amount of time. More specifically, the aspect of requiring numerous model evaluations is interesting to enable the usage of extensive parameter studies such as global sensitivity analysis, to improve the understanding of the properties that drive mantle convection, and how these two are correlated with each other.

However, we require 100 simulations to construct the presented surrogate model. Consequently, a question might be whether these 100 simulations are not sufficient for the exploration of the parameter behavior. Thus, as a first step, we evaluate the accuracy of the surrogate model's predictions on the validation data set by comparing them to an alternative estimation method. Specifically, we compare the surrogate model's prediction quality for the validation data against the difference of the validation data and the closest-matching realization from the training data set (where the "closest" match is determined based on viscosity value similarity). The results are presented in Tab. 3 and evaluated according to their mean, minimum, and maximum temperature differences. As before, we see behavioral differences for the different regimes. For all realizations of Regime 3, we obtain more accurate predictions through the surrogate model. For the average deviations, we get 2 % to 43 % more precise results for the surrogate model. This is underlying the previous observations, since this is the regime with the highest entropy generation. At this stage, we want to highlight the ninth realization from the validation data set. So, similar viscosity values yield very different temperature distributions, highlighting the challenges of this regime. Still, the surrogate model shows a better performance for this realization. In Regime 2, we obtain also slightly better prediction with the surrogate model for nearly all realizations. However, the difference between the two approaches remains small. Therefore, for Regime 2, both ap-

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

Table 3. Comparison of the difference between the validation data set and the surrogate model predictions, and with the differences between the validation data set and the closet training data set realization. Note μ_{val} denotes the viscosity value of the validation data, $\mu_{train,closest}$ the viscosity value of the training data set that is closest to the validation data, and ΔT the temperature difference.

ID	Regime	μ_{val} $(\cdot 10^{22}\ Pa \cdot s)$	$\mu_{\text{train,closest}}$ $(\cdot 10^{22}\ Pa \cdot s)$	ΔT_{mean} (K)		ΔT_{min} (K)		ΔT_{max} (K)	
				FE	NI-RB	FE	NI-RB	FE	NI-RB
0	2	3.206	3.273	43.63	26.13	-1309.56	-394.36	957.67	783.68
1	2	2.957	2.986	11.25	22.55	-297.47	-481.01	293.57	215.77
2	3	1.044	1.035	115.68	105.89	-1271.76	-948.15	1505.23	1137.83
3	3	1.713	1.679	86.00	48.84	-1453.57	-547.49	1366.88	1366.38
4	1	6.462	6.531	11.01	6.18	-363.47	-127.15	228.18	157.88
5	3	1.510	1.496	80.21	78.73	-1382.86	-650.76	1335.86	1380.53
6	1	8.274	8.222	1.50	5.10	-16.79	-45.70	28.44	55.12
7	1	10.00	9.886	2.89	10.17	-30.08	-91.92	29.38	103.12
8	2	3.735	3.673	9.98	8.97	-214.68	-270.42	274.59	120.65
9	3	1.386	1.396	85.57	74.93	-1352.81	-817.51	1337.96	1210.17

proaches perform equally. For Regime 1, the FE models overall yield slightly better results than the surrogate model but also the surrogate well approximates the solution, which corresponds nicely to the lower entropy generation in this regime.

Nevertheless, it is important to keep in mind the purpose of constructing surrogate models. In this study, we focus on the construction of the surrogate model and therefore perform the evaluation only for ten validation samples. The main advantage of surrogate models, however, is their capacity for extensive parameter studies involving potentially hundreds of thousands to millions of forward evaluations. Such studies typically involve multiple parameters beyond viscosity alone. For the viscosity-only scenario presented here, the parameter space is likely already well sampled by our 100 realizations, which explains the comparatively minor differences observed between surrogate model predictions and closely matching training simulations. When additional parameters are introduced, however, it becomes significantly less probable to find matching realizations within the training set, highlighting the necessity and strength of surrogate modeling approaches.

In realistic geodynamic applications, surrogate models become particularly valuable and the results from this study are promising. Earth's mantle convection, for instance, corresponds most closely to Regime 3, characterized by low viscosity values and a Rayleigh number typically exceeding 10^7. Importantly, Regime 3 also involves the most computationally expensive finite element (FE) simulations, making our surrogate model especially beneficial in this critical domain. Thus, it remains both feasible and advantageous to invest additional computational resources in the offline phase to specifically enhance the surrogate model's accuracy for this challenging but highly relevant regime.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

As a final note, surrogate models may perform even better in scenarios focusing on global model quantities rather than detailed regional states. In this study, we focused on the surrogate model construction to represent the entire temperature state, but some geodynamic applications may prioritize global metrics such as the number of plumes, the thickness of boundary layers, and/or mantle mixing rates, which are less sensitive to local discrepancies. The results presented here support this perspective, demonstrating the surrogate model's ability to reliably capture these global characteristics even within the dynamically complex Regime 3.

4.3 Future Potential and Applications

Considering the results of the previous sections we see several valuable opportunities for the application of Physics-based Machine Learning methods in future geodynamic modeling studies:

Geologically constrained models: While the match between the surrogate models and the real forward models decreases in accuracy with increasing Rayleigh number—which makes it likely that complex mantle convection models may differ substantially from surrogate models—the surrogate models reproduce features that are consistent across models even for high Rayleigh numbers as seen in Fig. 6. This implies that models that are partially or mostly constrained by data assimilation will make it much easier for surrogate models to predict realistic model outcomes. Studies that use past surface plate motions, or other prescribed velocity or traction boundary conditions could therefore benefit from generating significantly more models in their parameter searches using the technique presented here.

Inverse problems: Many applications in geodynamics require to find the numerical model that best matches some geological or geophysical observables (plate velocities, surface topography, surface heat flux, travel times of seismic waves, etc.). Finding the best-fit model generally requires a large number of forward models at often prohibitive computational cost (see for example Baumann and Kaus (2015), who computed 1.9×10^6 forward models). The NI-RB method opens up the possibility to reduce the number of required forward simulations, by filling in the parameter space with surrogate models, making these problems more tractable.

Instantaneous models: In this study, we have shown that the Non-Intrusive Reduced Basis method using Neural Networks for the projections allows it to capture the characteristics of the temperature distribution after 1 billion years of convection and thermal evolution. This is a substantial challenge for the surrogate models, since small differences early on due to small changes in model parameters typically grow over time and lead to large differences at the final state. However, there is a category of problems in geodynamic modeling that has a more direct relationship between input parameters and solution: Instantaneous models assume a given distribution of temperature and chemical composition (and resulting density) and use it to compute the velocity and pressure by solving the instantaneous Stokes equations. This solution can then be compared to observed surface deformation, plate velocities, seismic activity, etc. to inform what model parameters best fit the Earth's behavior. Because the model does not evolve over time, parameters like the viscosity more directly affect the resulting flow field, which should facilitate the use of the NI-RB method even in models with low or strongly varying viscosity. Additionally, these models can have extreme computational cost (for example, the global models of Stadler et al. (2010) had up to 1.2 billion unknowns and

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

required a minimum of 5000–7000 processor cores for their solution), making the use of surrogate models attractive when conducting a parameter study.

Code and data availability. ASPECT version 2.4.0, Heister et al. (2017); Kronbichler et al. (2012); Bangerth et al. (2022a)) used in these computations is freely available under the GPL v2.0 or later license through its software landing page https://geodynamics.org/resources/aspect or https://aspect.geodynamics.org and is being actively developed on GitHub and can be accessed via https://github.com/geodynamics/aspect. All generated ASPECT data required for the reproduction of this project can be found on Zenodo (https://doi.org/10.5281/zenodo.16751301; Valatsou et al., 2026).

Author contributions. All authors discussed and interpreted the presented work. Marilina Valatsou carried out the simulations and drafted the manuscript. All authors read and approved the final manuscript.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Solid Earth. The authors declare that there are no further conflict of interests.

Acknowledgements. The authors gratefully acknowledge funding provided by the Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz through the project SQuaRe (project number: 02E12062C). D.D. was supported through funding provided by the Bundesministerium für Forschung, Technologie und Raumfahrt through the project PBML-HM (project number: 16|S24062). R.G. and J.D. were partially supported by US National Science Foundation awards EAR-1925677, EAR-2054605, and EAR-2149126. J.D. was also supported by HELMHOLTZ funding of first-time professorial appointments of excellent women scientists (EBP-01-08). We thank the Computational Infrastructure for Geodynamics (geodynamics.org) which is funded by the US National Science Foundation under award EAR-0949446, EAR-1550901, and EAR-2149126 for supporting the development of ASPECT.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

References

Agarwal, S., Tosi, N., Breuer, D., Padovan, S., Kessel, P., and Montavon, G.: A machine-learning-based surrogate model of Mars' thermal evolution, Geophysical Journal International, 222, 1656–1670, https://doi.org/10.1093/gji/ggaa234, 2020.

Agarwal, S., Tosi, N., Kessel, P., Breuer, D., and Montavon, G.: Deep learning for surrogate modeling of two-dimensional mantle convection, Physical Review Fluids, 6, https://doi.org/10.1103/physrevfluids.6.113801, 2021.

Agarwal, S., Bekar, A. C., Hüttig, C., Greenberg, D. S., and Tosi, N.: Physics-based machine learning for mantle convection simulations, Physics of Fluids, 37, 2025.

Atkins, S., Valentine, A. P., Tackley, P. J., and Trampert, J.: Using pattern recognition to infer parameters governing mantle convection, Physics of the Earth and Planetary Interiors, 257, 171–186, 2016.

Bangerth, W., Dannberg, J., Fraters, M., Gassmoeller, R., Glerum, A., Heister, T., Myhill, R., and Naliboff, J.: ASPECT v2.4.0, https://doi.org/10.5281/zenodo.6903424, 2022a.

Bangerth, W., Dannberg, J., Fraters, M., Gassmoeller, R., Glerum, A., Heister, T., Myhill, R., and Naliboff, J.: ASPECT: Advanced Solver for Problems in Earth's ConvecTion, User Manual, https://doi.org/10.6084/m9.figshare.4865333, doi:10.6084/m9.figshare.4865333, 2022b.

Baumann, T. and Kaus, B. J.: Geodynamic inversion to constrain the non-linear rheology of the lithosphere, Geophysical Journal International, 202, 1289–1316, 2015.

Benner, P., Gugercin, S., and Willcox, K.: A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems, SIAM review, 57, 483–531, 2015.

Börsing, N., Wellmann, J. F., Niederau, J., and Regenauer-Lieb, K.: Entropy production in a box: Analysis of instabilities in confined hydrothermal systems, Water Resources Research, 53, 7716–7739, https://doi.org/10.1002/2017wr020427, 2017.

Cathles, L. M.: Viscosity of the earth's mantle, Princeton Legacy Library, Princeton University Press, Princeton, NJ, 2016.

Chakir, R. and Hammond, J.: A non-intrusive reduced basis method for elastoplasticity problems in geotechnics, Journal of Computational and Applied Mathematics, 337, 1–17, https://doi.org/10.1016/j.cam.2017.12.044, 2018.

Chakir, R., Streichenberger, B., and Chatellier, P.: A Non-Intrusive Reduced Basis Method for Urban Flows Simulation, in: 14th WCCM-ECCOMAS Congress, CIMNE, https://doi.org/10.23967/wccm-eccomas.2020.029, 2021.

Chen, W., Wang, Q., Hesthaven, J. S., and Zhang, C.: Physics-informed machine learning for reduced-order modeling of nonlinear problems, Journal of Computational Physics, 446, 110 666, https://doi.org/https://doi.org/10.1016/j.jcp.2021.110666, 2021.

Colli, L., Bunge, H.-P., and Schuberth, B. S.: On retrodictions of global mantle flow with assimilated surface velocities, Geophysical Research Letters, 42, 8341–8348, 2015.

Condie, K. C.: Earth as an Evolving Planetary System, Elsevier, https://doi.org/10.1016/c2015-0-00179-4, 2016.

Dannberg, J., Gassmöller, R., Li, R., Lithgow-Bertelloni, C., and Stixrude, L.: An entropy method for geodynamic modelling of phase transitions: capturing sharp and broad transitions in a multiphase assemblage, Geophysical Journal International, 231, 1833–1849, 2022.

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, 102 143, 2021.

Degen, D., Cacace, M., and Wellmann, F.: 3D multi-physics uncertainty quantification using physics-based machine learning, Scientific Reports, 12, https://doi.org/10.1038/s41598-022-21739-7, 2022.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

Degen, D., Caviedes Voullième, D., Buiter, S., Hendricks Franssen, H.-J., Vereecken, H., González-Nicolás, A., and Wellmann, F.: Perspectives of physics-based machine learning strategies for geoscientific applications governed by partial differential equations, Geoscientific Model Development, 16, 7375–7409, https://doi.org/10.5194/gmd-16-7375-2023, 2023.

Falkner, S., Klein, A., and Hutter, F.: BOHB: Robust and Efficient Hyperparameter Optimization at Scale, https://doi.org/10.48550/ARXIV.1807.01774, 2018.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Heister, T., Dannberg, J., Gassmöller, R., and Bangerth, W.: High Accuracy Mantle Convection Simulation through Modern Numerical Methods. II: Realistic Models and Problems, Geophysical Journal International, 210, 833–851, https://doi.org/10.1093/gji/ggx195, 2017.

Hesthaven, J. and Ubbiali, S.: Non-intrusive reduced order modeling of nonlinear problems using neural networks, Journal of Computational Physics, 363, 55–78, https://doi.org/10.1016/j.jcp.2018.02.037, 2018.

Hesthaven, J. S., Rozza, G., and Stamm, B.: Introduction, in: Certified reduced basis methods for parametric partial differential equations, SpringerBriefs in mathematics, pp. 1–13, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-22470-1, 2016.

Huang, P.-W. and Wellmann, F.: An explanation to the Nusselt–Rayleigh discrepancy in naturally convected porous media, Transport in Porous Media, 137, 195–214, 2021.

Kerswell, B., Cerpa, N. G., Tommasi, A., Godard, M., and Padrón-Navarta, J. A.: RocMLMs: Predicting Rock Properties Through Machine Learning Models, Journal of Geophysical Research: Machine Learning and Computation, 1, e2024JH000 264, https://doi.org/https://doi.org/10.1029/2024JH000264, e2024JH000264 2024JH000264, 2024.

King, S. D.: Models of Mantle Viscosity, in: AGU Reference Shelf, pp. 227–236, American Geophysical Union, https://doi.org/10.1029/rf002p0227, 2013.

Kronbichler, M., Heister, T., and Bangerth, W.: High Accuracy Mantle Convection Simulation through Modern Numerical Methods, Geophysical Journal International, 191, 12–29, https://doi.org/10.1111/j.1365-246X.2012.05609.x, 2012.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, Technometrics, 21, 239, https://doi.org/10.2307/1268522, 1979.

Mitrovica, J. and Forte, A.: A new inference of mantle viscosity based upon joint inversion of convection and glacial isostatic adjustment data, Earth and Planetary Science Letters, 225, 177–189, 2004.

Nakada, M., Okuno, J., Lambeck, K., and Purcell, A.: Viscosity structure of Earth's mantle inferred from rotational variations due to GIA process and recent melting events, Geophysical Journal International, 202, 976–992, https://doi.org/10.1093/gji/ggv198, 2015.

Oppenheim, A. V., Schafer, R. W., Yoder, M. A., and Padgett, W. T.: Discrete-time signal processing, Pearson, Upper Saddle River, NJ, 3 edn., 2009.

Quarteroni, A., Manzoni, A., and Negri, F.: Construction of RB Spaces by SVD-POD, pp. 115–140, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-15431-2_6, 2016.

Ricard, Y. and Wuming, B.: Inferring the viscosity and the 3-D density structure of the mantle from geoid, topography and plate velocities, Geophysical Journal International, 105, 561–571, https://doi.org/10.1111/j.1365-246x.1991.tb00796.x, 1991.

Rudolph, M. L., Lekić, V., and Lithgow-Bertelloni, C.: Viscosity jump in Earth's mid-mantle, Science, 350, 1349–1352, 2015.

Schubert, G., Turcotte, D. L., and Olson, P.: Mantle Convection in the Earth and Planets, Cambridge University Press, https://doi.org/10.1017/cbo9780511612879, 2001.

She, Y. and Fu, G.: Viscosities of the crust and upper mantle constrained by three-dimensional GPS rates in the Sichuan–Yunnan fragment of China, Earth, Planets and Space, 71, https://doi.org/10.1186/s40623-019-1014-x, 2019.

https://doi.org/10.5194/egusphere-2026-4171
Preprint. Discussion started: 22 July 2026

Simader, T., Degen, D., Meier, T., Rüter, H., and Wellmann, F.: Surrogate models in convection-dominated fault systems: considerations for efficient and reliable realizations, Comput. Geosci., 30, 2026.

Stadler, G., Gurnis, M., Burstedde, C., Wilcox, L. C., Alisic, L., and Ghattas, O.: The dynamics of plate tectonics and mantle flow: From local to global scales, science, 329, 1033–1038, 2010.

Steinberger, B. and Calderwood, A. R.: Models of large-scale viscous flow in the Earth's mantle with constraints from mineral physics and surface observations, Geophysical Journal International, 167, 1461–1481, 2006.

Swischuk, R., Mainini, L., Peherstorfer, B., and Willcox, K.: Projection-based model reduction: Formulations for physics-based machine learning, Computers & Fluids, 179, 704–717, https://doi.org/10.1016/j.compfluid.2018.07.021, 2019.

Turcotte, D. and Schubert, G.: Geodynamics, Cambridge University Press, https://doi.org/10.1017/cbo9780511843877, 2014.

Valatsou, M., Degen, D., Dannberg, J., Gassmöller, R., and Wellmann, F.: Physics-Based Machine Learning: Opportunities and Challenges for Mantle Convection, Zenodo [data set] and [code], https://doi.org/10.5281/zenodo.16751301, 2026.

van Zelst, I., Crameri, 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.

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, Journal of Computational Physics, 384, 289–307, https://doi.org/https://doi.org/10.1016/j.jcp.2019.01.031, 2019.

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.