



# How biased are our models? – a case study of the alpine region

Denise Degen<sup>1</sup>, Cameron Spooner<sup>2,3</sup>, Magdalena Scheck-Wenderoth<sup>2,4</sup>, and Mauro Cacace<sup>2</sup>

<sup>1</sup>Computational Geoscience and Reservoir Engineering, RWTH Aachen University, Willnerstr. 2, 52062 Aachen, Germany

<sup>2</sup>GFZ German Research Centre for Geosciences, Helmholtz Centre Potsdam, Telegrafenberg, 14473 Potsdam, Germany

<sup>3</sup>Institute of Earth and Environmental Science, Potsdam University, 14476 Potsdam, Germany

<sup>4</sup>Department of Geology, Geochemistry of Petroleum and Coal, RWTH Aachen University, 52056 Aachen, Germany

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

Received: 18 February 2021 – Discussion started: 1 June 2021

Revised: 20 October 2021 – Accepted: 24 October 2021 – Published: 24 November 2021

**Abstract.** Geophysical process simulations play a crucial role in the understanding of the subsurface. This understanding is required to provide, for instance, clean energy sources such as geothermal energy. However, the calibration and validation of the physical models heavily rely on state measurements such as temperature. In this work, we demonstrate that focusing analyses purely on measurements introduces a high bias. This is illustrated through global sensitivity studies. The extensive exploration of the parameter space becomes feasible through the construction of suitable surrogate models via the reduced basis method, where the bias is found to result from very unequal data distribution. We propose schemes to compensate for parts of this bias. However, the bias cannot be entirely compensated. Therefore, we demonstrate the consequences of this bias with the example of a model calibration.

## 1 Introduction

Understanding the subsurface is as important in the field of geosciences as understanding climatic processes. In this paper, we focus on the understanding of the subsurface temperature field, which is of major importance for geothermal applications. Here, we focus on numerical process simulations to improve our understanding of the subsurface. These simulations are based on both geological and physical models; however, in this paper we will primarily further investigate the latter. The physical model has two major sources of uncertainties arising from the physical processes itself (i.e., neglected processes, generalizations) (i.e., Houghton et al., 2001; Murphy et al., 2004; Refsgaard et al., 2007) and from the physical parameters (i.e., thermal conductivity, ra-

diogenic heat production) in terms of ranges (i.e., Freyemark et al., 2017; Lehmann et al., 1998; Vogt et al., 2010; Wagner and Clauser, 2005) and their distribution (i.e., Feyen and Caers, 2006; Floris et al., 2001).

To compensate for both sources of uncertainties, one commonly performs model calibrations, either deterministically (i.e., Doherty and Hunt, 2010; Fuchs and Balling, 2016; Hill and Tiedeman, 2006; Wellmann and Reid, 2014) or stochastically (i.e., Elison et al., 2019; Linde et al., 2017). Model calibrations aim to compensate for existing model error by adjusting the model parameters to a given data set. Naturally, the data set itself is subject to uncertainties. However, if we perform, for instance, stochastic model calibrations such as Markov chain Monte Carlo (Iglesias and Stuart, 2014), we are able to take these uncertainties into account. Nonetheless, there is another problem related to the data set, i.e., data distribution. Note that in the following we introduce the problems arising from data distribution through the example of temperature measurements. Still, many of the presented problems are generalizable for other geophysical data sources.

The first problem related to the data distribution is the depth location of the individual measurements. Our geothermal models have a depth in the magnitude of 100 km. In contrast, our deepest thermal measurements are commonly at a depth of 5 to 7 km. The second problem is also related to data density. Focusing on the horizontal data distribution, we face the problem of data sparsity and unequal data distribution. In certain model areas, we have very few temperature measurements and in other areas, we have a much larger data density. This inequality can be compensated by using data-weighting schemes (i.e., Degen et al., 2021; Lerch, 1991). However,

we also have areas where no temperature measurements exist. Data weighting cannot compensate for these nonexistent measurements. The problem is further enlarged by the data source. Most of our temperature measurements come from the hydrocarbon industry; however, their targets and those of the geothermal industry are not the same in every region. This means that we can face the problem of lower data resolution in areas of interest whilst possessing higher data resolution in areas that are not of primary interest.

The problem of data sparsity is widely recognized (i.e., Cherpeau and Caumon, 2015; Zehner et al., 2010). However, there are no studies systematically investigating the bias we introduce due to temperature measurements in a geothermal setting. Studies for the measurement bias are common in the field of remote sensing (i.e., Feng et al., 2016; Schwarz et al., 2020); however, their focus is entirely different. In remote sensing, the location of the measurements is subjected to uncertainties. In contrast, our problems do not arise from imprecise measurement locations but their distribution. Naturally, our locations are also associated with uncertainties; however, in basin-scale applications they are of minor importance.

In this paper, we aim to provide a systematic investigation of the bias induced by measurement distribution. Therefore, we perform global sensitivity analyses to determine the influence of the model parameters (i.e., thermal conductivity, radiogenic heat production) on the model response (i.e., temperature) within the spatial extent of the Alpine models. Sensitivity analyses can be subdivided into local and global analyses. We choose a global sensitivity analysis to investigate not only the influence of the parameters themselves but also the parameter correlations. Note that a local sensitivity analysis assumes that all parameters are independent of each other (Degen et al., 2021; Saltelli, 2002; Saltelli et al., 2010; Sobol, 2001; Wainwright et al., 2014). Furthermore, we want to avoid a possible overestimation of the influences. A previous model study showed that the local sensitivity analysis can overestimate the influences (Degen et al., 2021). Global sensitivity analyses have been performed before in, for example, Baroni and Tarantola (2014); Cannavó (2012); Cloke et al. (2008); Degen et al. (2021); Fernández et al. (2017); van Griensven et al. (2006); Song et al. (2015); Tang et al. (2007); Wainwright et al. (2014); Zhan et al. (2013); however, they are either in a different geophysical setting and or with a different focus of interest.

Global sensitivity analyses have the disadvantage of being computationally very demanding since they require several thousand to several hundred thousand forward simulations. This makes these analyses infeasible even for state-of-the-art finite element problems. To compensate for the expensive nature of the method, we employ the reduced basis method to construct suitable surrogate models. The principle idea is to replace the original high dimensional model with a low dimensional model while keeping the key characteristic of the problem (Benner et al., 2015; Hesthaven et al., 2016; Prud'homme et al., 2002; Quarteroni et al., 2015). In

this paper, we do not focus on the observation space alone but also investigate the entire temperature state. Hence, we need a surrogate model for the entire state. The reduced basis method is able to provide us with this, in contrast to many other surrogate model techniques (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). The reduced basis method is widely known in mathematical applications (i.e., Benner et al., 2015; Grepl, 2005; Hesthaven et al., 2016; Aretz-Nellesen et al., 2019; Kärcher et al., 2018; Prud'homme et al., 2002; Quarteroni et al., 2015; Rozza et al., 2007); however, only few geoscientific applications exist (Degen et al., 2020a). Nevertheless, some studies do use comparable approaches (Ghasemi and Gildin, 2016; Gosses et al., 2018; Rizzo et al., 2017; Rousset et al., 2014; Zlotnik et al., 2015).

In this paper, we investigate the problems related to data distribution for the case study of the Alpine region. The geological model, covering the Alpine orogen and its forelands, is taken from a previous study (Spooner et al., 2020b). Thermal studies of the Alpine region are of interest to understand how the present-day deformation is linked to the thermal field. Therefore, we want to illustrate how the interpretation of the temperature field might be biased.

## 2 Materials and methods

In the following, we briefly introduce the concepts of global sensitivity analyses and the reduced basis method. Furthermore, we introduce the physical model and the temperature data used throughout this study.

### 2.1 Global sensitivity analysis

In this study, we investigate the measurement bias and therefore require knowledge of which parameters the temperature distribution is sensitive to. Therefore, we employ a sensitivity analysis (SA). We distinguish two types of sensitivity analyses: local and global. The local sensitivity analysis investigates the influence of the model parameters with respect to a user-defined reference parameter set. All parameter variations are considered independently of each other and only the vicinity of the input parameters is explored, e.g., in a variation range of  $\pm 1\%$  of the input parameters (Sobol, 2001; Wainwright et al., 2014). In contrast, the global sensitivity analysis explores the entire parameter space and also investigates the parameter correlations (Sobol, 2001). In this paper, we use a global sensitivity analysis with the Saltelli sampler (Saltelli, 2002; Saltelli et al., 2010), and we investigate two types of sensitivity indices: the first- and total-order indices. First-order indices describe the influence arising from the model parameter itself. Total-order indices additionally contain information about the parameter correlation (Sobol,

2001). We perform the SA with the Python library SALib (Herman and Usher, 2017) and 100 000 realizations per parameter to reduce the statistical error. For further information regarding the global sensitivity analysis, refer to Sobol (2001); Saltelli (2002); Saltelli et al. (2010), and for a comparison between local and global sensitivity analysis, refer to Wainwright et al. (2014) and Degen et al. (2021).

## 2.2 Forward problem

For this case study, we are using a conductive heat transfer problem (Turcotte and Schubert, 2002). To ensure that we investigate the relative importance of the parameters and for better efficiency, we use the following non-dimensional form:

$$\frac{\lambda}{\lambda_{\text{ref}} S_{\text{ref}} l_{\text{ref}}^2} \nabla^2 \left( \frac{T - T_{\text{ref}}}{T_{\text{ref}}} \right) + \frac{S}{S_{\text{ref}} T_{\text{ref}} \lambda_{\text{ref}}} = 0, \quad (1)$$

where  $\lambda$  is the thermal conductivity,  $S$  is the radiogenic heat production, and  $T$  is the temperature. The subscript “ref” denotes the respective reference parameters, and  $l_{\text{ref}}$  is the reference length. Note that the Laplace operator acts on the normalized space.

## 2.3 Reduced-order modeling

In this work, we require a surrogate model that is representative of the entire temperature state to ensure the feasibility of the study. Therefore, we use the reduced-basis (RB) method for the surrogate model construction, a projection-based model order reduction technique. It aims to replace the original high-dimensional model with a low-dimensional representation while keeping the input–output relationship the same. Hence, the method preserves the underlying physics. One limitation of the RB method is that it is restricted to underlying low-dimensional parameter spaces. With higher-dimensional parameter spaces the complexity of the parameter space tends to increase, leading to longer construction times and surrogate model dimensions that are too large. The RB method destroys the sparsity pattern of the system, meaning that a large surrogate model will require a longer execution time than the original finite element model due to its dense nature. To overcome this issue, we use a hierarchical sensitivity study, as we will discuss in Sect. 3.1.

The RB method is comprised of the following two parts: the offline and online stages. During the offline stage, we construct our surrogate model. This stage is computationally expensive but needs to be performed only once. In the online stage, we use the low-dimensional surrogate model. This stage is computationally fast and therefore ideal for expensive outer-loop processes such as the global sensitivity analysis. In previous studies, we showed that the RB method yields a speed-up of several orders of magnitude for the here-described physical problem (Degen et al., 2020a, 2021).

All reduced models are generated with the software package DwarfElephant (Degen et al., 2020a). Degen et al. (2020a) also contains a detailed description of the reduced-order model construction, which is omitted here for the sake of brevity. For further information regarding the RB method, refer to Hesthaven et al. (2016); Prud’homme et al. (2002); Quarteroni et al. (2015), and for a detailed overview of various model order reduction techniques, refer to Benner et al. (2015). Further information regarding the RB method in the field of geosciences is presented by Degen et al. (2020a) and specifically for basin-scale thermal applications by Degen et al. (2021).

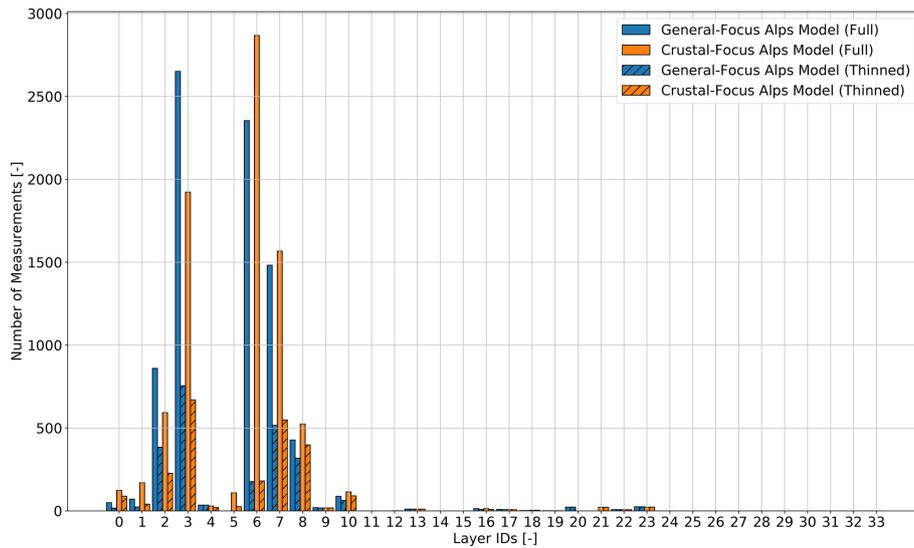
## 2.4 Temperature data

We present the temperature data set in the form of a histogram in Fig. 1 and illustrate the spatial distribution in Fig. 2. These temperature data are identical to those presented by Spooner et al. (2020b). The entire data set is comprised of 8120 measurements with a maximum depth of 7.3 km and a mean depth of 1.8 km. The Italian National Geothermal Database (Trumpy and Manzella, 2017) provides the data for the southern foreland of the Alps. For the northern foreland, the data are derived from the Upper Rhine Graben (URG) database provided in Freymark et al. (2017) and references therein. The data of the Molasse Basin are retrieved from Przybycin et al. (2015) and references therein, whereas the data from the Alps are compiled from Luijendijk et al. (2020).

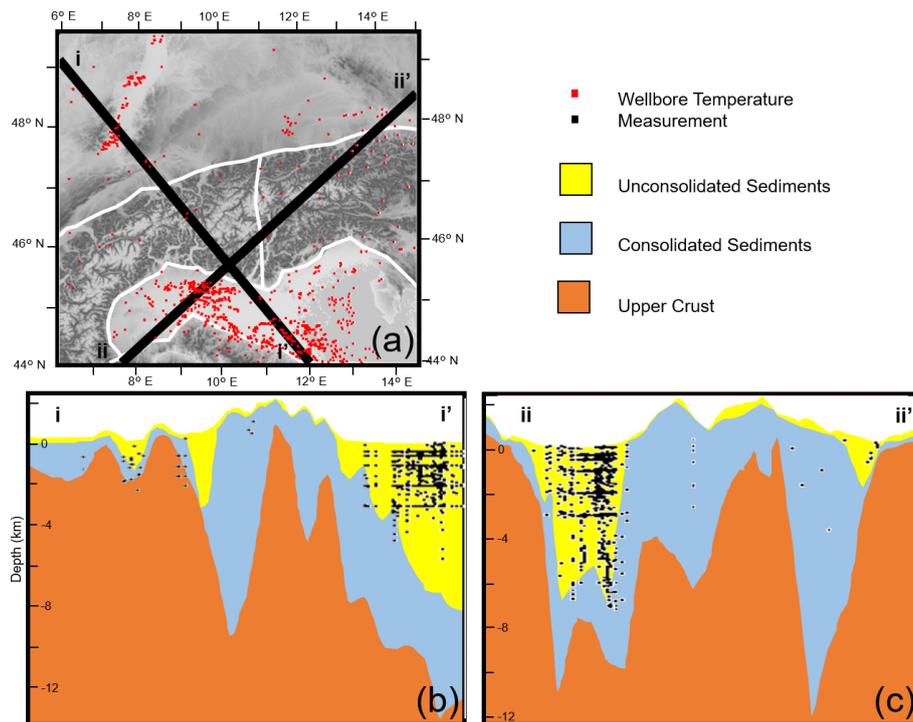
The spatial distribution of measurements varies widely across the region, it is sparse in the Molasse Basin (103) and Alps (83) and ranges to being dense in the Po Basin (7619). In an effort to alleviate a significant bias and to improve the efficiency of the presented methods, the data set was filtered to give a more uniform measurement density across the region, with a significant reduction in the Po Basin (2028) whilst retaining those in the Molasse Basin (103) and Alps (83). Deeper measurements ( $> 2$  km) were preferentially maintained throughout the region as they better indicate crustal temperatures, a particular focus of the work undertaken here. This procedure resulted in a filtered data set of 2388 wellbore temperature measurements with a mean depth of 2.3 km.

### 2.4.1 Weighting

A common issue of the temperature data for the calibration of thermal models is their unequal distribution. To compensate for this inequality, we introduce a weighting scheme in this paper. There are different possibilities to weight the measurement data. In this paper, we use a regional weighting scheme that combines quantitative measures and our knowledge about the geophysical setting and the data quality. As previously mentioned, the data set was reduced to 2388 data



**Figure 1.** Distribution of the measurements according to the geological layers. For the layer IDs, please refer to Table A1 in the Appendix.



**Figure 2.** Spatial distribution of the temperature measurements (a) projected on the surface, (b) along the cross section (i), and (c) along the cross section (ii).

points in total. We subdivide the model into the following four regions:

- the Alps with 83 measurements,
- the URG with 177 measurements,
- the Molasse with 103 measurements,
- and the Po Basin with 2025 measurements.

As we can see, the Po Basin contains many more temperature measurements than the other regions. Additionally, we need to take into account that the temperature measurements of the Alps are non-robust since they are minimum temperature values. In addition, the data from the Upper Rhine Graben need to be treated carefully since we do not account for convective processes in this paper. These aspects yield the following weighting scheme:

- the Po Basin is not weighted,
- the Molasse is weighted by a factor of 20 since the Po Basin contains 20 times more data points,
- and the Upper Rhine Graben and the Alps are weighted by a factor of 0.5.

The weight of 0.5 of the data from the Upper Rhine Graben and the Alps is based on the extensive experience of the authors from previous studies in the regions. This value is subjective and can be updated once quantitative measures of the data quality are available. Keep in mind that this study aims to demonstrate the effects of the data bias rather than to provide an optimal weighting scheme for the Alpine region.

The weighting scheme is applied to the quantity of interest of the global sensitivity analyses. Here, we consider the L2 norm of the difference between the measured ( $T_{\text{mes}}$ ) and simulated temperatures ( $T_{\text{sim}}$ ). We apply the weighting to this temperature misfit:

$$\sum_i \sqrt{((T_{i,\text{mes}} - T_{i,\text{sim}}) \cdot \omega_i)^2}, \quad (2)$$

where  $\omega$  is the weighting factor. The procedure for the model calibration (see Sect. 4.3) is analogous, with the difference that we apply it to L1 instead of the L2 norm in the cost function.

### 3 Alpine region

In this paper, we study two versions of the Alps model.

1. The first one focuses on the sediments and the lithospheric mantle. This model has been presented in Spooner et al. (2020b) and is from here on denoted as the “General-Focus Alps” model. It consists of 31 geological layers. Each layer has a homogeneous and isotropic thermal conductivity and radiogenic heat production.
2. The second model concentrates on the upper crust and is denoted as the “Crustal-Focus Alps” model. This model contains 34 geological layers, and again each layer has a homogeneous and isotropic thermal conductivity and radiogenic heat production. For this second model, we have a higher number of geological layers because several layers of the “General-Focus Alps” model have been further subdivided, as demonstrated in Table A1.

Both models have an extent of 640 km in the  $x$  direction and 600 km in the  $y$  direction. In the vertical direction both models extend down to the lithosphere–asthenosphere boundary (LAB). The models are discretized using hexahedrons with a horizontal resolution of about 21.33 km  $\times$  19.35 km.

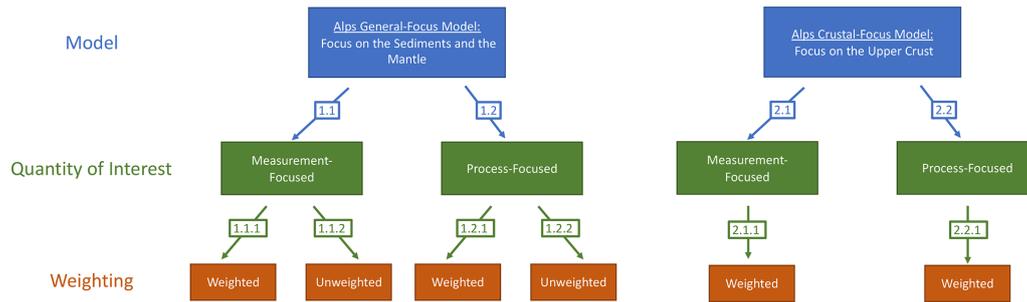
At the top of both models we apply a Dirichlet boundary condition representing the annual average surface temperatures (Böhm et al., 2009; Fan and Van den Dool, 2008; Locarnini et al., 2013) varying from  $-10^\circ\text{C}$  (Alps) to  $16^\circ\text{C}$  (Adriatic Sea). Additionally, at the base of the model, we assign a Dirichlet boundary condition varying between  $1250^\circ\text{C}$  below the Vosges massif and  $1400^\circ\text{C}$  below the Bohemian massif (Schaeffer and Lebedev, 2013). For further information regarding the physical and geological setting of the General-Focus Alps model, refer to Spooner et al. (2020b).

For the reference thermal conductivity, we use a value of  $3.0 \text{ W m}^{-1} \text{ K}^{-1}$  (corresponding to the largest thermal conductivity). Analogously, the reference length is 640 000 m (corresponding to the maximum model extent) and the reference radiogenic heat production  $2.6 \mu\text{W m}^{-3}$  (corresponding to the largest radiogenic heat production). The reference parameters are the same for both models.

In this paper, in addition to the General-Focus Alps model already presented in Spooner et al. (2020b), we use the Crustal-Focus Alps model, where the upper crust below the Po Basin was thinned in order to better fit temperature observations from the previous thermal modeling work (Spooner et al., 2020b), with requisite thickening of the lower crust carried out in order to compensate. Inconsistencies in the original classification of unconsolidated sediments and consolidated sediments were also rectified, specifically in the region of the southern Alps. Small alterations to the depth of the Moho were also made as a result of more recent observations (Magrin and Rossi, 2020). The gravity residual of the newly generated structural model was then re-minimized using the same methodology described in Spooner et al. (2019a), achieving a misfit as good as the original model. An overview of all models and analyses presented in this paper is given in Fig. 3.

#### 3.1 Thermal model

To avoid the problem of the parameter space dimension becoming too large, we perform a hierarchical global sensitivity analysis. The setup of the hierarchical sensitivity analysis is shown in Figs. 4 and 5. The setup for both the General-Focus and Crustal-Focus Alps model is the same. Therefore, we explain the hierarchical sensitivity analysis using the General-Focus Alps model. For the top-level sensitivity analysis, we separately combine layers with equal thermal conductivities and radiogenic heat productions, reducing the number of thermal parameters from 62 to 19. This top-level sensitivity analysis investigates the influences of the thermal properties in the entire model region. However, the investigated properties combine several entities, so in order to isolate the thermal properties that are influencing the temperature distribution, we perform additional sensitivity analysis for those properties that exceed our threshold value of  $1 \times 10^{-2}$  for the total-order sensitivity indices. This threshold was chosen at a level where we observed a significant decrease in the sensi-



**Figure 3.** Schematic overview of the models used in this paper.

tivity indices. In total, we perform three additional sensitivity analysis for the following indices:

1. unconsolidated sediments and the lower crust (red rectangle of Fig. 4 and peak 1 of Fig. 5),
2. unconsolidated and consolidated sediments (gray rectangle of Fig. 4 and peak 2 of Fig. 5),
3. and the upper crust (blue rectangles of Fig. 4 and peak 3 of Fig. 5).

Each of these additional sensitivity analyses also contains a thermal parameter from the top-level sensitivity analysis to enable a comparison between all analyses. We investigate all thermal properties of the upper crust instead of only those that are above the threshold since the upper crust has been the primary interest in previous studies (Spooner et al., 2020b). Note that in this section we only present the setup of the hierarchical sensitivity analysis. A detailed presentation of the individual analyses follows in the next sections.

### 3.2 Influence of the quantity of interest

In this paper, we want to investigate how much our analyses are influenced by focusing on measurements. This is important since we calibrate and validate our analyses with, for instance, temperature measurements. The sensitivity analysis investigates the relative changes that are induced by changes in the model parameters (i.e., thermal conductivity and radiogenic heat production). For the sensitivity analysis, we need to define a quantity of interest, which allows us to define with respect to what measure the changes are investigated. To investigate the influence of the measurements, we perform the hierarchical sensitivity analyses with two different quantities of interest for the General-Focus Alps model (branch 1.1 and 1.2 of Fig. 3).

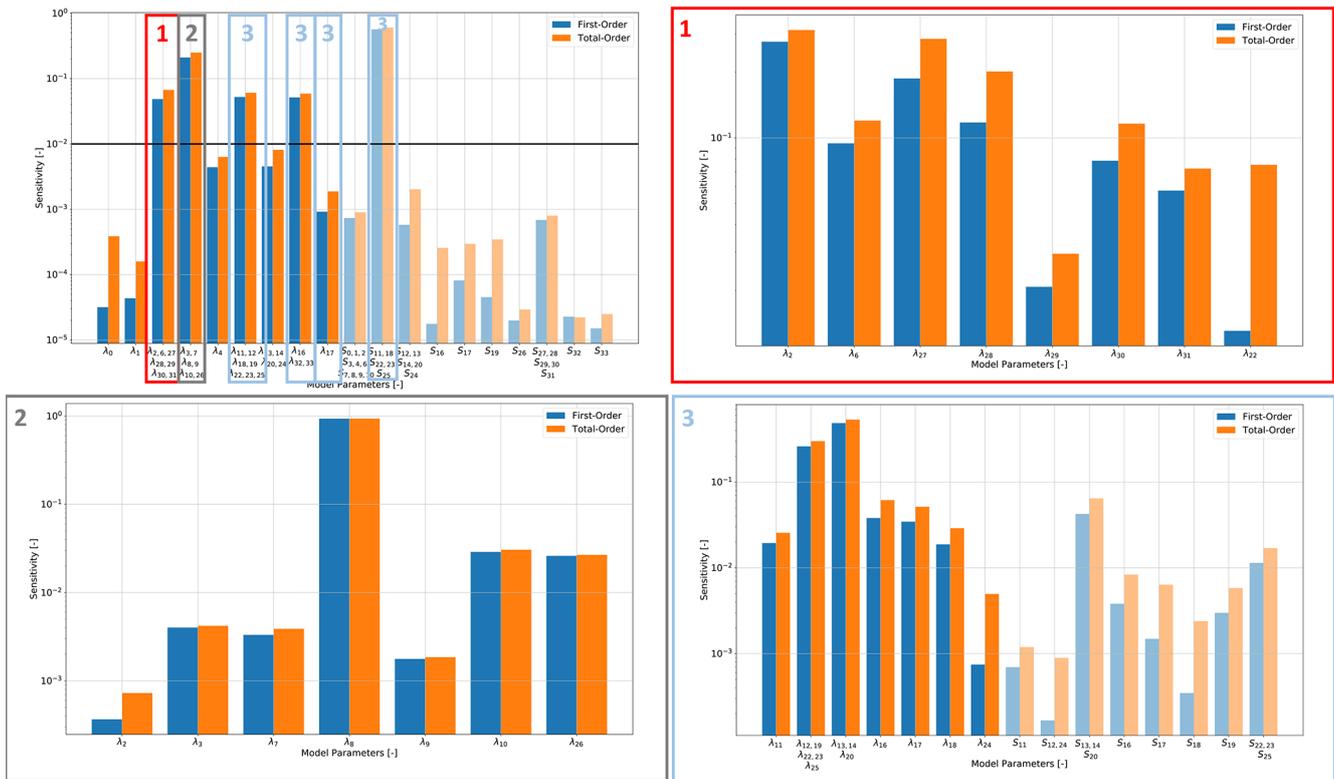
1. The first quantity of interest is defined as the sum of the absolute temperature values of the entire model. This results in a sensitivity analysis that is representative of the physical processes since all regions in the model are treated equally.

2. The second quantity of interest is defined as the absolute misfit between the simulated and measured temperature values. Hence, the resulting sensitivity analysis is focused on the temperature measurements.

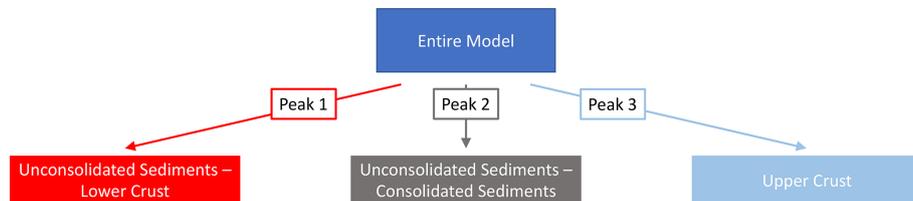
In the following, we focus on the difference in the total order sensitivity indices between those two hierarchical sensitivity analyses (branch 1.1 and 1.2 of Fig. 3) to present the bias introduced by the measurements and the consequences of using temperature data from the hydrocarbon industry for the calibration of geothermal models. In this study, we use only the General-Focus Alps model to avoid any influence from factors other than the measurements. The results of the hierarchical global sensitivity analysis are presented in Figs. 6 to 9. We again follow the procedure illustrated in Fig. 5. This means that the hierarchical analysis consists of the following four global sensitivity analyses: (i) the entire model (Fig. 6), (ii) the unconsolidated sediments and lower crust (Fig. 7), (iii) the unconsolidated sediments and consolidated sediments (Fig. 8), and (iv) the upper crust (Fig. 9).

Focusing on the difference between the hierarchical sensitivity analyses, we make two key observations.

1. We observe tendentially higher difference for the thermal conductivities of deeper geological layers. This is highlighted in Fig. 6 with gray rectangles. Here, we observe the highest differences for the following indices:
  - $\lambda_1$ , i.e., the thermal conductivity of the unconsolidated sediments of the Upper Rhine Graben below 1 km;
  - $\lambda_4$ , i.e., the thermal conductivity of the unconsolidated sediments of the Molasse Basin;
  - $\lambda_{11}, \lambda_{12}, \lambda_{18}, \lambda_{19}, \lambda_{22}, \lambda_{23}$ , and  $\lambda_{25}$ , which denote the thermal conductivities of the Appennine, Istrea, Molasse, eastern Alps, Po, northeastern Adria, and southeastern Adria upper crust, respectively;
  - $\lambda_{13}, \lambda_{14}, \lambda_{20}$ , and  $\lambda_{24}$ , which denote the thermal conductivities of the Moldanubian, Bohemian, western Alps, and Ivrea upper crust, respectively;
  - $\lambda_{17}$ , i.e., the thermal conductivity of the Vosges upper crust.



**Figure 4.** Representation of the hierarchical process-focused sensitivity analysis of the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1.



**Figure 5.** Schematic representation of the hierarchical global sensitivity analysis.

Furthermore, this can be confirmed by looking at the additional sensitivity analysis of the unconsolidated sediments of the lower crust (Fig. 7), where we observe higher differences for the lower crust thermal conductivities.

- The difference in the sensitivity indices tend to be larger for the radiogenic heat production than for the thermal conductivity. This is highlighted in Figs. 6 and 9 with red rectangles.

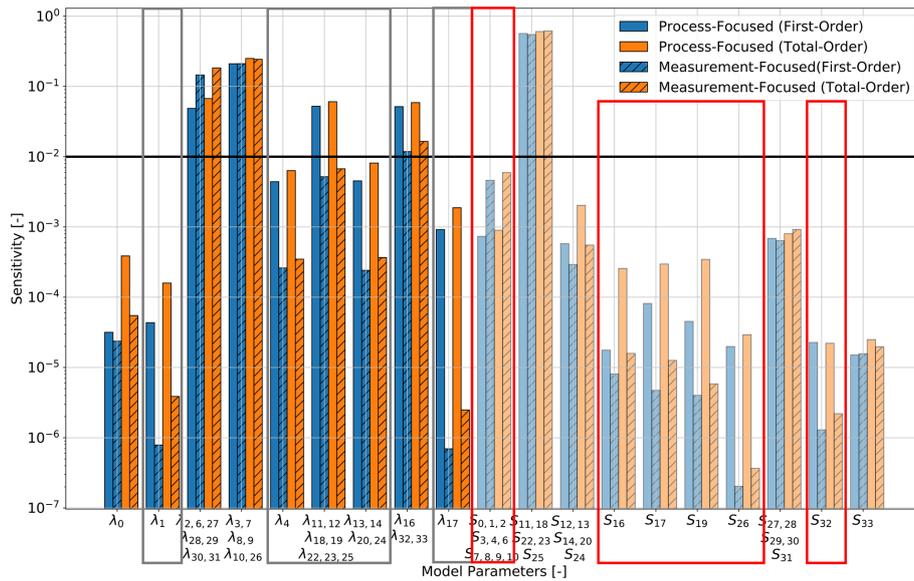
Furthermore, in the case of the process-focused analyses, the model is sensitive to more parameters and we obtain a slightly higher parameter correlation.

Here we focus on the difference observable for the analysis of the unconsolidated and consolidated sediments. We obtain huge differences in the sensitivities for both sediment types.

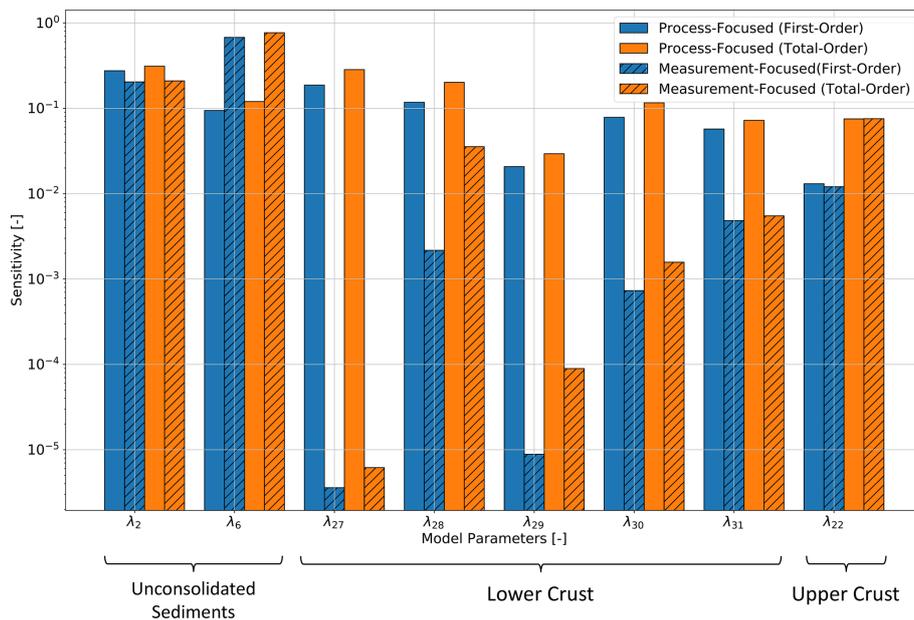
For the thermal conductivities of the unconsolidated sediments, the measurement-focused analysis returns tendentially higher influences, whereas for the consolidated sediments the process-focused analysis results in tendentially higher influences of the thermal conductivities.

Finally, we switch our focus to the analysis of the upper crust. For the upper crust, we observe six thermal conductivities with a significant difference in the sensitivity indices:

- $\lambda_{13}$ ,  $\lambda_{14}$ , and  $\lambda_{20}$ , which denote the thermal conductivities of the Moldanubian, Bohemian, and western Alps upper crust, respectively;
- $\lambda_{16}$ , i.e., the thermal conductivity of the Saxothuringian upper crust;
- $\lambda_{17}$ , i.e., the thermal conductivity of the Vosges upper crust;



**Figure 6.** Top-level sensitivity analysis (focusing on the entire Alps model) with different quantities of interest of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1. The solid black line denotes the threshold value for determining if the parameters are influencing the model response.



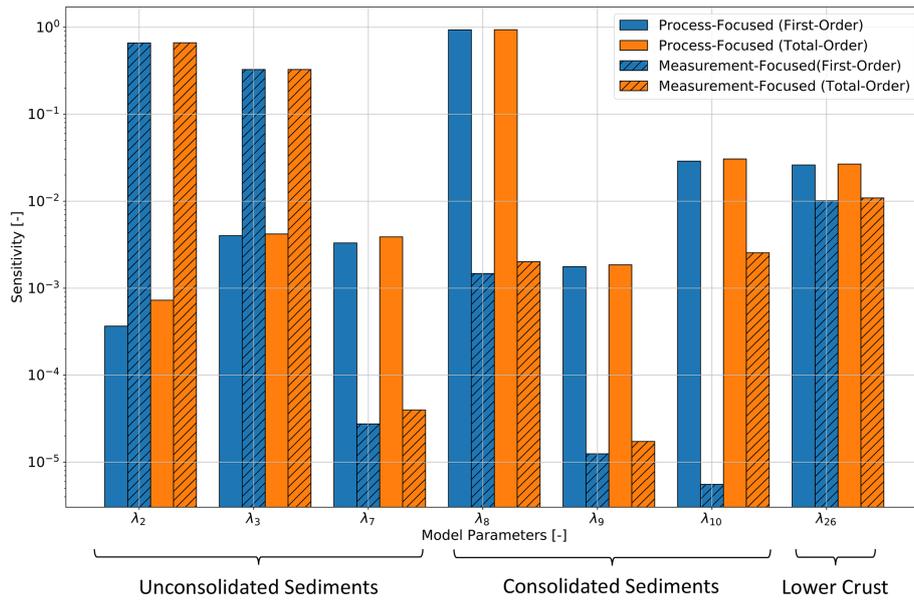
**Figure 7.** Sensitivity analysis of the unconsolidated sediments and lower crust with different quantities of interest of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1.

- $\lambda_{18}$ , i.e., the thermal conductivity of the Molasse upper crust;
- $\lambda_{24}$ , i.e., the thermal conductivity of the Ivrea upper crust.

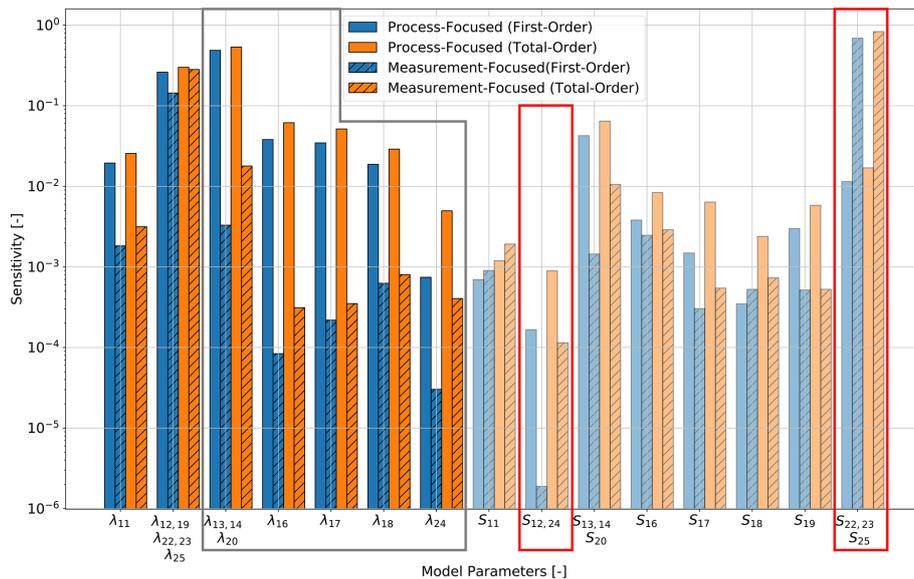
The differences for the radiogenic heat production are the highest for the following indices:

- $S_{12}$  and  $S_{24}$ , which denote the radiogenic heat production of the Istrea and Ivrea upper crust, respectively;
- $S_{22}$ ,  $S_{23}$ , and  $S_{25}$ , which denote the radiogenic heat production of the Po, northeastern Adria, and southeastern Adria upper crust, respectively.

Note that the layers of the upper crust ( $\lambda_{22}$  in Fig. 7) and lower crust ( $\lambda_{26}$  in Fig. 8) do not add any further informa-



**Figure 8.** Sensitivity analysis of the unconsolidated and consolidated sediments with different quantities of interest of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1.



**Figure 9.** Sensitivity analysis of the upper crust with different quantities of interest of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1.

tion to this section. Both are properties directly taken from the top-level sensitivity analysis and are required to enable a comparison between the top-level and lower-level sensitivity analyses. However, they represent only one property from their respective lithological unit. Therefore, they are not representative of any kind of trend analysis.

### 3.3 Influence of the weighting

The consequences of introducing a weighting scheme have been already partly addressed in Degen et al. (2021). However, there the authors focused on the consequences for the process of model calibrations. Here, we want to investigate how we can compensate for the measurement bias by applying weights.

Analogous to the previous section, we focus on the differences in the total order sensitivity indices. Again, we per-

form the hierarchical analyses (Figs. 10 to 13) as presented in Fig. 5. Thus, analogously to the previous section, we present the results of the global sensitivity analysis, focusing on (i) the entire model (Fig. 10), (ii) the unconsolidated sediments and lower crust (Fig. 11), (iii) the unconsolidated and consolidated sediments (Fig. 12), and (iv) the upper crust (Fig. 13). For all analyses, we can observe that the weighted scenario tends to be closer to the process-focused analysis than the non-weighted scenario for the thermal conductivities. This is highlighted by the gray rectangles in Figs. 10 and 13. The behavior is very prominent for the thermal conductivity of the Moldanubian lower crust (gray rectangle of Fig. 11).

In contrast, we observe for the thermal conductivities of the Upper Rhine Graben layers a closer resemblance of the non-weighted scenario to the process-focused analysis (blue rectangle of Fig. 10).

We also observe for the radiogenic heat production that for most layers the indices of the weighted case are closer to the process-focused analysis than the non-weighted case (red rectangles of Fig. 10). Differing from this trend is the radiogenic heat production of the Istrea and Ivrea upper crust. Furthermore, we observe that the weighted analysis overestimates the influence of the Molasse upper crust (Fig. 13).

### 3.4 Reduced basis method

Before discussing the results of this paper, we briefly present the surrogate models obtained through the RB method in terms of cost and accuracy. In total, we consider five different surrogate models, as listed in Table 1. Here, the first four models are based on the General-Focus Alps model (for the setup, please refer to Fig. 5), and the fifth model is based on the Crustal-Focus Alps model.

We observe from Table 1 that the model dimension for the surrogate model varies between 54 to 360. Since we require one finite element (FE) simulation for each basis function, this corresponds to 54 to 360 FE simulations depending on the surrogate model. Consequently, the cost for the most expensive surrogate model (fifth RB model) equals the cost we require to calculate 360 FE simulations, which is several orders of magnitude lower than the total number of forward simulations performed in this paper. We do not provide the actual execution times since they vary vastly between various hardware structures. Furthermore, note that this stage is fully parallelizable in contrast to some inversion methods. For all surrogate models, we reach the pre-defined maximum relative error tolerance of  $5 \times 10^{-4}$ .

## 4 Discussion

In the following, we discuss the consequences of focusing a study on measurements. Therefore, we discuss the changes in the sensitivities for the different quantities of interest and

weighting schemes. Furthermore, we demonstrate the consequences through a deterministic model calibration example.

### 4.1 Influence of the quantity of interest

The different quantities of interest represent the bias introduced by the unequal distribution of the measurement locations. Hence, we can use the difference in the sensitivity analysis to discuss the bias that is induced by the temperature measurements. So far, we had two key observations for the study of the different quantities of interest:

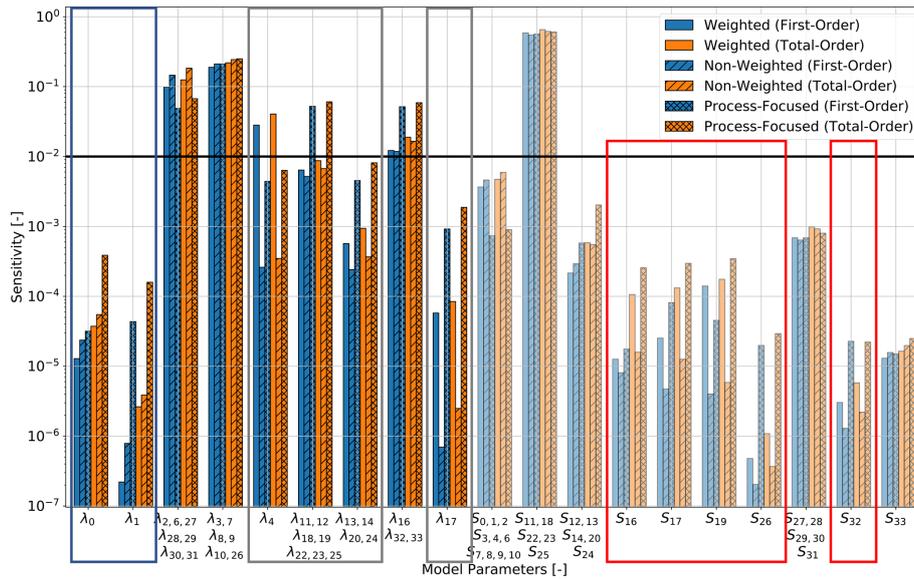
1. the difference in the indices for the thermal conductivities is higher for deeper layers,
2. the differences are higher for the radiogenic heat productions than for the thermal conductivities.

Both of these observations can be explained by having a closer look at the depth distribution of the temperature measurements (Fig. 14). We can see that most measurements are located at a depth of up to 2 km. The deepest measurement is at depth of about 7.3 km, whereas the model extends to a maximum depth of about 140.5 km. Hence, most measurements are located in shallower geological layers, and in the deepest layers we find no measurements at all (Fig. 1). Therefore, the measurement-focused analysis tends to underestimate the influences of the deeper geological layers and overestimates the influences of shallower layers. This is true for both thermal conductivity and radiogenic heat production.

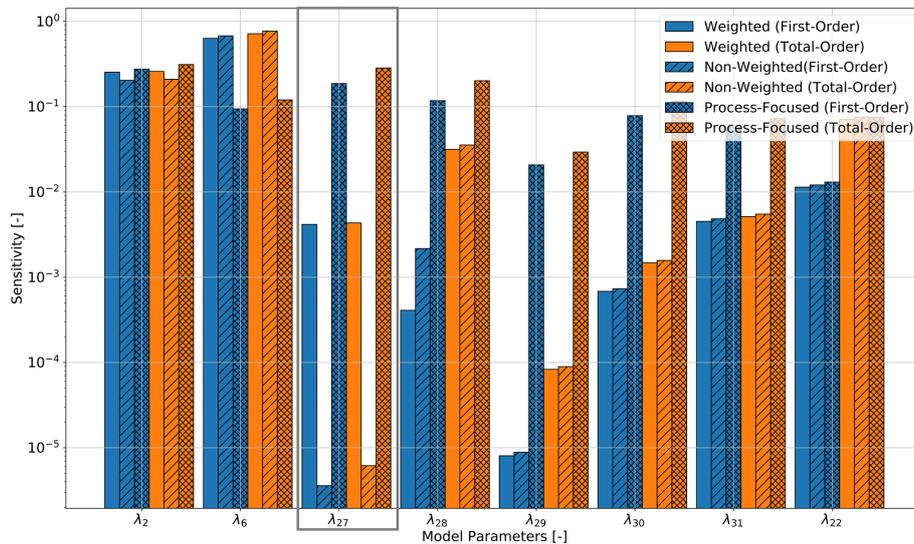
We investigate the phenomenon more closely for the analysis of the unconsolidated and consolidated sediments. Here, we have a prominent overestimation of the influences of the unconsolidated sediments and an underestimation of the consolidated sediments, delineated as follows:

- 384 data points in the unconsolidated sediments of the Upper Rhine Graben above 1 km ( $\lambda_0$  in Fig. 6),
- 755 data points in the unconsolidated sediments of the Upper Rhine Graben below 1 km ( $\lambda_1$  in Fig. 6),
- 516 data points in the unconsolidated sediments of the Po Basin below 2 km ( $\lambda_7$  in Fig. 6),
- 318 data points in the Consolidated sediments outside of sedimentary basins ( $\lambda_8$  in Fig. 6),
- 18 data points in the consolidated sediments of the Molasse Basin ( $\lambda_9$  in Fig. 6),
- 63 data points in the consolidated sediments of the Po Basin ( $\lambda_{10}$  in Fig. 6).

The much higher data density in the unconsolidated sediments explains the high influence of the thermal conductivities of the unconsolidated sediments for the measurement-focused analysis. The only remaining question is why the



**Figure 10.** Top-level sensitivity analysis (focusing on the entire Alps model) with different weighting schemes of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1. The solid black line denotes the threshold value for determining if the parameters are influencing the model response.

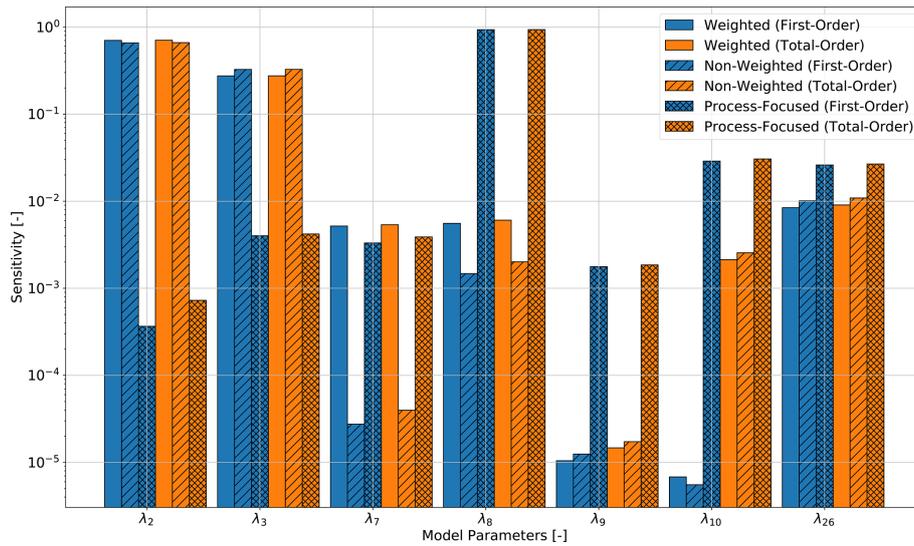


**Figure 11.** Sensitivity analysis of the unconsolidated sediments and lower crust with different weighting schemes of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1.

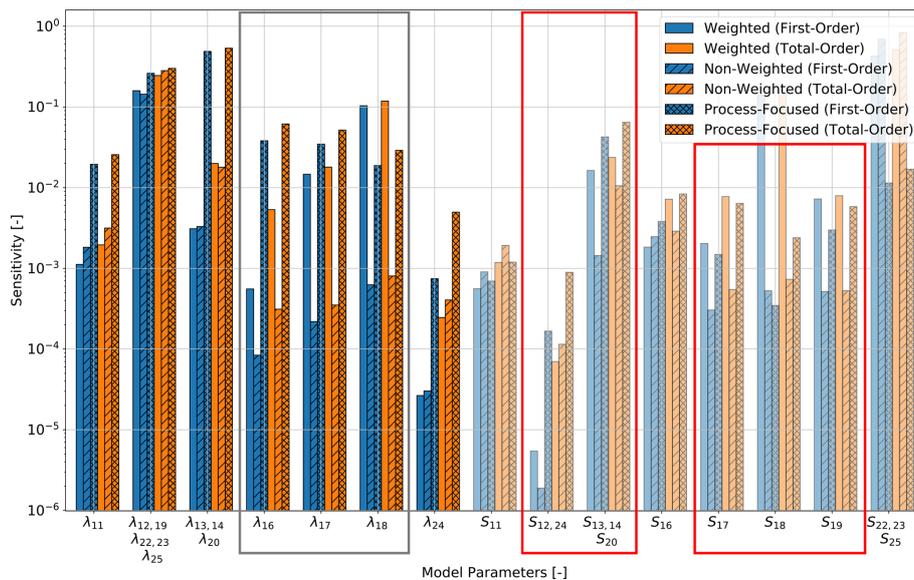
influence of the thermal conductivity of the unconsolidated sediments in the Po Basin below 2 km is underestimated despite it containing 516 data points. This might be a bias introduced by the high data density of 755 data points in the unconsolidated sediments of the Upper Rhine Graben below 1 km ( $\lambda_1$ ).

The behavior is more pronounced for the radiogenic heat production for lithological reasons. The highest influences of the radiogenic heat productions arise from the upper crust (Fig. 6), meaning that the radiogenic heat production is more

prominent in deeper parts of the model. However, these parts of the model are further away from our measurement locations. Hence, the measurement-focused analysis highly underestimates the influence of the radiogenic heat production. The same effect can be observed for the thermal conductivity of the upper crust ( $\lambda_5$  in Fig. 6). For the measurement-focused analysis, the influence of the thermal conductivity is below the threshold, whereas for the process-focused analysis it is above.



**Figure 12.** Sensitivity analysis of the unconsolidated and consolidated sediments with different weighting schemes of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the layer IDs and symbols, please refer to Table A1.



**Figure 13.** Sensitivity analysis of the upper crust with different weighting schemes of the hierarchical global sensitivity analysis for the General-Focus Alps model. For the Layer IDs and symbols, please refer to Table A1.

The consequence of the data distribution becomes obvious once we look at the analysis of the unconsolidated sediments and lower crust (Fig. 7). For all lower crustal layers, the influence is significantly underestimated in the measurement-focused scenario. Consequently, by focusing on the measurement in the further analysis we would lose all information related to the lower crust, despite the layer possibly being important for the physical understanding of the subsurface.

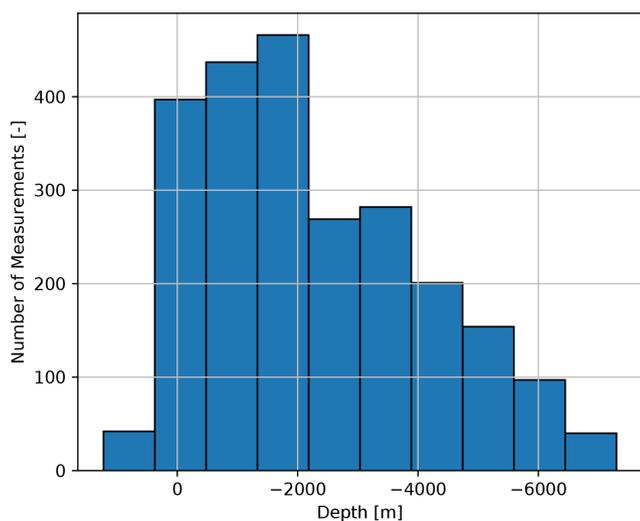
In addition, for the analysis of the upper crust (Fig. 9) we are confronted with the consequences of the unequal data distribution. The huge difference in the influences of the thermal

conductivities of the Saxothuringian, Vosges, Molasse, and Ivrea upper crust is caused by a very low or zero data density. In addition to this, the influence of the Moldanubian, Bohemian, and western Alps upper crust is underestimated. We have data for the Moldanubian and western Alps upper crust but no data for the Bohemian upper crust, yielding this discrepancy.

The influence of the radiogenic heat production of the Istrea and Ivrea upper crust is underestimated in the measurement-focused study due to the lack of data, whereas the influence of the radiogenic heat production of the Po,

**Table 1.** Overview of the various RB models. Here we present the focus of the different surrogate models and their dimensions.

	Geological model	Area of interest	Number basis functions
1. RB model	General-Focus Alps model	Entire model	265
2. RB model	General-Focus Alps model	Unconsolidated sediments and lower crust	70
3. RB model	General-Focus Alps model	Unconsolidated sediments and consolidated sediments	54
4. RB model	General-Focus Alps model	Upper crust	104
5. RB model	Crustal-Focus Alps model	Entire model	360

**Figure 14.** Distribution of the measurements according to depth.

northeastern Adria, and southeastern Adria upper crust is overestimated. This is likely caused by the measurements available for both the Po and northeastern Adria upper crust layers.

We also observed slightly higher parameter correlations for the process-focused analysis. This is probably related to the fact that the model is sensitive to more parameters.

#### 4.2 Influence of the weighting

We observed that the weighted measurement-focused analysis tends to be closer to the process-focused analysis. This becomes understandable by looking at the applied weighting scheme. We applied a regional weighting scheme to compensate for the unequal data distribution in the four regions of our model. Hence, we can compensate partly for the measurement bias. However, we are not able to fully compensate for the data sparsity. The main reason for this is that we can compensate for fewer data points but not for regions without data points since no measurements are available to which we could apply a higher weight. This can be observed, for instance, in the properties related to the layers of the Molasse.

We observed that the sensitivity indices of the thermal properties related to the layers inside the Upper Rhine Graben are further apart for the weighted and process-

focused comparison than for the non-weighted process-focused one. This is related to the choice of the weighting scheme. We chose to put less weight on the temperature data from the Upper Rhine Graben since we do not account for convective effects in this paper. Analogously, the properties of the Apennine upper crust layers also have a too small influence for the weighted scenario. As a reminder, we downgraded the importance of the temperature data in this region since the data consists of minimum temperature data.

Through the weighting we are able to compensate for the underestimation of the unconsolidated sediments of the Po Basin. Hence, the bias most likely induced by the high data density of the other layers can be reduced.

For the thermal conductivities of the Saxothuringian, Vosges, and Molasse upper crust (gray rectangle of Fig. 13), we are again able to remove parts of the data bias caused by the data sparsity of these layers. The same phenomenon is observable for the radiogenic heat production of the upper crust (red rectangles of Fig. 13).

Note that the weighting scheme is case study and aim specific. Depending on our knowledge about data quality, regions of interest, and other aspects the weighting scheme can be designed in a case specific manner. In this paper, we do not aim to provide “the ideal” weighting scheme for the Alpine region. Instead, we demonstrate the impact of a weighting scheme for thermal modeling. In addition, note that due to the high impact of the weighting this also means that we need to carefully consider the weighting scheme. An incorrect weighting scheme will increase the bias.

#### 4.3 Calibration example

So far, we have presented that we obtain significantly differing sensitivities for the process-focused and measurement-focused study. In the following, we demonstrate the consequences of this difference through a deterministic model calibration. We choose the example of a model calibration because this is a typical inverse process that relies on observation data.

Model calibration aims to compensate for existing model errors by adjusting the model parameters in accordance with our temperature measurements. Analogous to Degen et al. (2021), we use a sensitivity-driven model calibration for more robust results. In this study, we performed various sensitivity analyses. For the model calibration, we require the

measurement-focused sensitivity analyses (branch 1.1 and 2.1 of Fig. 3). We need these sensitivity analyses because they represent the information content that can be derived from the temperature data. In the case of the General-Focus model, five thermal parameters that can be calibrated are yielded (Table 2). The data are insensitive to the remaining parameters. Hence, we cannot calibrate these values. We are left with mostly shallow layers to calibrate. The exception is the lithospheric mantle, which is influential due to its large volume.

In the following, we discuss the results of the automated model calibration and its consequences. Note that in this work we use the model calibration in a slightly different way. Usually, it is used to compensate for model errors. That means of course that it also identifies the problematic model areas. In this work, we employ the model calibration as an identification tool for model errors. Therefore, we use the calibrated values by Spooner et al. (2020b) as initial values, which have been obtained through a “trial-and-error” model calibration. As a result, large discrepancies between our initial values and calibrated values identify model problems.

The first model problem that we can identify is the measurement bias through an unequal data distribution (General-Focus – unweighted). This can be at least partly removed through data weighting (General-Focus – weighted), yielding smaller differences between initial and calibrated values. Nonetheless, we observe a low radiogenic heat production in the upper crust, meaning that our model is non-ideal in the description of the upper crust. This also leads to thermal conductivities that are too low in the sediments and too high in the lithospheric mantle.

Therefore, we introduce a second model, the Crustal-Focus model. For this model, we obtain a good agreement for the upper crust but greater discrepancies in unconsolidated sediments (below 1 km) and the lithospheric mantle. Hence, we can remove the error in the upper crust but at the same time introduce new error sources.

For the calibration of the unweighted General-Focus model, we achieve a  $R^2$  value of 0.87, and for the weighted case we achieve a value of 0.86. Also, for the Crustal-Focus model, we obtain a  $R^2$  value of 0.86. This shows that we are able to fit any temperature distribution at the cost of obtaining partly unphysical thermal conductivities. To illustrate this, we focus in this paper on the thermal properties and not on the temperatures.

Note that we do not aim to present the “optimal” model in this paper. Instead, we want to demonstrate various components that influence the model. Generating an optimal model is not possible since all models are per definition wrong (Box, 1979). We present here two models that fulfill different purposes. The General-Focus model is better if we are interested in the entire model domain. In the case that our area of interest is only the upper crust, the Crustal-Focus model is preferable.

#### 4.4 Influence of the model

We have discussed the consequences of the model change for the calibrated thermal conductivities. Now we want to briefly discuss the consequences for the sensitivities. Therefore, we repeat the process-focused and measurement-focused sensitivity analysis for the Crustal-Focused model. Note that we consider only the weighted scenario (branch 2.1.1 and 2.2.1 of Fig. 3).

For the Crustal-Focus model, we thinned the upper crust. This can be clearly observed in the decreased sensitivities of the model to the upper crust layers (red box of Fig. 15). However, this change is only visible in the process-focused analysis. The measurement-focused analysis mostly fails to resolve these changes due to the data sparsity in the upper crust (red box of Fig. 16). Underestimated changes are observable for the Saxothuringian upper crust. This again highlights the information loss of measurement-focused studies and the dangers associated with calibrations.

The radiogenic heat production of most of the lower crust is more influential for the Crustal-Focused model as the upper crust was thinned by thickening the lower crust. The only exception is the Saxothuringian lower crust ( $\lambda_{26}$ ). For the process-focused analysis (Fig. 15) it loses importance, and for the measurement-focused analysis (Fig. 16) it gains importance. For both models, we apply a Dirichlet boundary condition at the top and the bottom of the model. Hence, the temperature distribution is determined by the ratio of the thermal properties. Therefore, the difference in the Saxothuringian lower crust likely arises from the changes of other geological layers. The same is likely for the changes of the thermal conductivity of the unconsolidated sediments in the Molasse Basin. In addition, the changes in the influences arising from the radiogenic heat production of the lithospheric mantle are caused by other layers, especially considering the very low values of these layers.

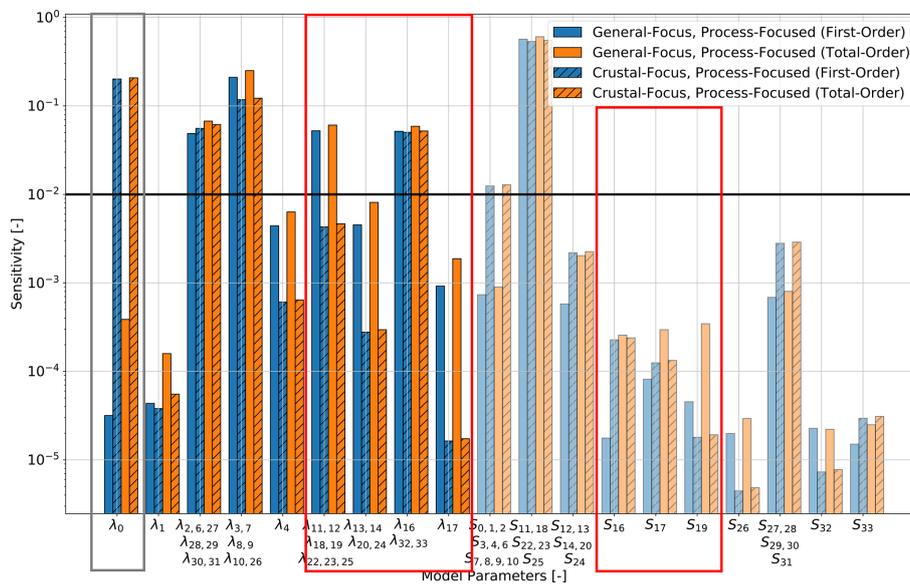
Furthermore, we observe a higher influence of the unconsolidated sediments in the Upper Rhine Graben (gray box of Fig. 15) although the model has not been changed around the Upper Rhine Graben. However, this might be an effect of the reclassification in the unconsolidated and consolidated sediments. These changes are more pronounced for the measurement-focused analysis (gray box of Fig. 16) than for the process-focused analysis. This is again caused by the data distribution since we have more measurements at a shallower depth.

#### 4.5 Gravity model

The model change is observable in both the model calibration for the thermal properties and the corresponding sensitivities. However, if we look at the gravity residuals (Fig. 17), we do not observe any significant changes. This highlights a general point for the construction of geological models. We have different data sources available for the construction of

**Table 2.** Comparison of the initial thermal properties and the calibrated thermal properties for different geological models and different weighting schemes. The parameter that is not considered in the model calibration due to sensitivities that are too low is denoted with n/a.

Parameter	Initial value	Calibrated value		
		General-Focus – unweighted	General-Focus – weighted	Crustal-Focus – weighted
$\lambda_2$ [W m <sup>-1</sup> K <sup>-1</sup> ]	2.0	1.53	1.70	2.02
$\lambda_3$ [W m <sup>-1</sup> K <sup>-1</sup> ]	2.3	1.33	2.04	3.45
$\lambda_{4,5}$ [W m <sup>-1</sup> K <sup>-1</sup> ]	1.8	n/a	1.62	1.53
$\lambda_6$ [W m <sup>-1</sup> K <sup>-1</sup> ]	2.0	1.86	2.02	2.03
$\lambda_{32,33}$ [W m <sup>-1</sup> K <sup>-1</sup> ]	3.0	3.71	3.18	2.5
$S_{22,23,25}$ [ $\mu$ W m <sup>3</sup> ]	1.3	0.2	0.8	1.3



**Figure 15.** Comparison of the sensitivities of the process-focused study for both the General-Focus and Crustal-Focus Alps model. The solid black line denotes the threshold value for determining if the parameters are influencing the model response.

a geological model. It is crucial to incorporate multiple data sources and not rely on a single data source. If we would have constructed a model of the Alps purely based on gravity, we would not have been able to identify the problem of the thickness of the upper crust.

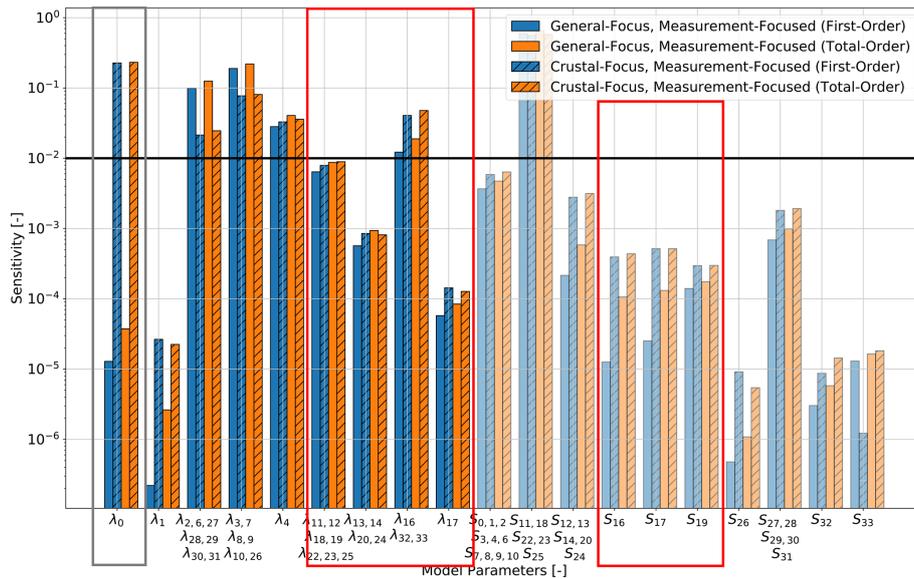
**4.6 Outlook**

In this paper, we have seen that the measurements induced a significant bias. This opens the discussion of subsequent projects. Therefore, we would like to investigate how we can decrease this bias by incorporating further data sources that give us only an indirect measure of the temperature. Furthermore, it would be interesting to further explore the field of joint inversion to incorporate various geophysical data sources already used during model construction.

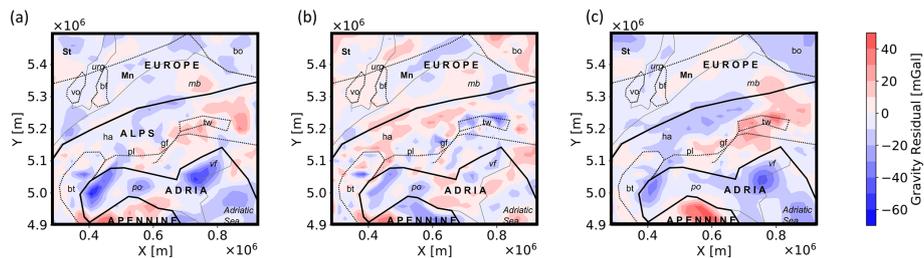
**5 Conclusions**

In this paper, we have demonstrated the bias that a measurement-focused study can cause. This bias can be partly removed through automated and customized data-weighting schemes. However, as is typical for geoscientific applications, many areas of the model do not have any associated data. Unfortunately, it is not possible to compensate for the bias arising from these areas. This shows the importance of focusing on regions where data are present whenever possible.

However, many inverse processes such as deterministic and stochastic model calibrations are dependent on measurement data. In this case the bias is unavoidable. Nonetheless, we need to be aware of which kind of bias we are introducing through this procedure to take the effects for all further analyses into account. We need to be aware that the data are often only informative towards the shallower layers. Hence,



**Figure 16.** Comparison of the sensitivities of the measurement-focused study for both the General-Focus and Crustal-Focus Alps model. The solid black line denotes the threshold value for determining if the parameters are influencing the model response.



**Figure 17.** Gravity residual of (a) the General-Focus model, (b) the Crustal-Focus model, and (c) the difference between the General-Focus model and Crustal-Focus model. Acronyms are as follows: St stands for Saxothuringian Zone, Mn stands for Moldanubian Zone, Ha stands for Helvetic Alps, bo stands for Bohemian Massif, vo stands for Vosges Massif, bf stands for Black Forest Massif, tw stands for Tauern Window, bt stands for Briançonnais Terrane, pl stands for Periadriatic Lineament, gf stands for Guidicarie Fault, urg stands for Upper Rhine Graben, mb stands for Molasse Basin, po stands for Po Basin, and vf stands for Veneto–Friuli plain.

we lose the information about deeper layers and at the same time overestimate the influence of the shallower layers. This also means that we are unable to calibrate and validate the lower parts of our geological models. Nonetheless, these regions are important to avoid influences from, for instance, the lower boundary condition.

We have also seen the importance of considering various data sources. The changes from the General-Focus to the Crustal-Focus model were only visible in the thermal studies but not in the gravity residuals.

Note that although we performed the analyses for the case study of the Alps, these aspects hold in general since the data distribution shown here is typical for geoscientific applications.

## Appendix A: Appendix symbols and Layer IDs

**Table A1.** Symbols and layer IDs for both the General-Focus model and Crustal-Focus model.

Layer	Layer ID	Property	Symbol
Unconsolidated sediments URG (top 1 km)	0	thermal conductivity radiogenic heat production	$\lambda_0$ $S_0$
Unconsolidated sediments URG (below 1 km)	1	thermal conductivity radiogenic heat production	$\lambda_1$ $S_1$
Unconsolidated sediments rest (top 1 km)	2	thermal conductivity radiogenic heat production	$\lambda_2$ $S_2$
Unconsolidated sediments rest (below 1 km)	3	thermal conductivity radiogenic heat production	$\lambda_3$ $S_3$
<b>General-Focus model:</b>			
Unconsolidated sediments Molasse	4	thermal conductivity radiogenic heat production	$\lambda_4$ $S_4$
<b>Crustal-Focused model:</b>			
Unconsolidated sediments Molasse (top 1 km)	4	thermal conductivity radiogenic heat production	$\lambda_4$ $S_4$
Unconsolidated sediments Molasse (below 1 km)	5	thermal conductivity radiogenic heat production	$\lambda_4$ $S_4$
Unconsolidated sediments Po Basin (top 2 km)	6	thermal conductivity radiogenic heat production	$\lambda_6$ $S_6$
Unconsolidated sediments Po Basin (below 2 km)	7	thermal conductivity radiogenic heat production	$\lambda_7$ $S_7$
Consolidated sediments	8	thermal conductivity radiogenic heat production	$\lambda_8$ $S_8$
Consolidated sediments Molasse	9	thermal conductivity radiogenic heat production	$\lambda_9$ $S_9$
Consolidated sediments Po Basin	10	thermal conductivity radiogenic heat production	$\lambda_{10}$ $S_{10}$
Upper crust Apennine	11	thermal conductivity radiogenic heat production	$\lambda_{11}$ $S_{11}$
Upper crust Istria	12	thermal conductivity radiogenic heat production	$\lambda_{12}$ $S_{12}$
Upper crust Moldanubian	13	thermal conductivity radiogenic heat production	$\lambda_{13}$ $S_{13}$
<b>General-Focus model:</b>			
Upper crust Bohemian	14	thermal conductivity radiogenic heat production	$\lambda_{14}$ $S_{14}$
<b>Crustal-Focused model:</b>			
Upper crust Bohemian	14	thermal conductivity radiogenic heat production	$\lambda_{14}$ $S_{14}$
Upper crust Bohemian Volcanics	15	thermal conductivity radiogenic heat production	$\lambda_{14}$ $S_{14}$

Table A1. Continued.

Layer	Layer ID	Property	Symbol
Upper crust Saxothuringia	16	thermal conductivity radiogenic heat production	$\lambda_{16}$ $S_{16}$
Upper crust Vosges	17	thermal conductivity radiogenic heat production	$\lambda_{17}$ $S_{17}$
Upper crust Molasse	18	thermal conductivity radiogenic heat production	$\lambda_{18}$ $S_{18}$
Upper crust eastern Alps	19	thermal conductivity radiogenic heat production	$\lambda_{19}$ $S_{19}$
<b>General-Focus model:</b>			
Upper crust western Alps	20	thermal conductivity radiogenic heat production	$\lambda_{20}$ $S_{20}$
<b>Crustal-Focused model:</b>			
Upper crust western Jura	20	thermal conductivity radiogenic heat production	$\lambda_{20}$ $S_{20}$
Upper crust western Alps	21	thermal conductivity radiogenic heat production	$\lambda_{20}$ $S_{20}$
Upper crust Po Basin	22	thermal conductivity radiogenic heat production	$\lambda_{22}$ $S_{22}$
Upper crust northeastern Adria	23	thermal conductivity radiogenic heat production	$\lambda_{23}$ $S_{23}$
Upper crust Ivrea	24	thermal conductivity radiogenic heat production	$\lambda_{24}$ $S_{24}$
Upper crust southeastern Adria	25	thermal conductivity radiogenic heat production	$\lambda_{25}$ $S_{25}$
Lower crust Saxothuringian	26	thermal conductivity radiogenic heat production	$\lambda_{26}$ $S_{26}$
Lower crust Moldanubian	27	thermal conductivity radiogenic heat production	$\lambda_{27}$ $S_{27}$
Lower crust Alps	28	thermal conductivity radiogenic heat production	$\lambda_{28}$ $S_{28}$
Lower crust Ivrea	29	thermal conductivity radiogenic heat production	$\lambda_{29}$ $S_{29}$
Lower crust Liguria and Apennine	30	thermal conductivity radiogenic heat production	$\lambda_{30}$ $S_{30}$
Lower crust Adria	31	thermal conductivity radiogenic heat production	$\lambda_{31}$ $S_{31}$
Lithospheric mantle northwest	32	thermal conductivity radiogenic heat production	$\lambda_{32}$ $S_{32}$
Lithospheric mantle southeast	33	thermal conductivity radiogenic heat production	$\lambda_{33}$ $S_{33}$

*Code availability.* For the construction of the reduced models, we used the software package DwarfElephant (<https://doi.org/10.5281/zenodo.4074777>, Degen et al., 2020b; Degen et al., 2020a). The software, which is based on the finite element solver MOOSE (Permann et al., 2020), is freely available at <https://doi.org/10.5281/zenodo.4074777> (Degen et al., 2020b). The sensitivity analyses are performed with the Python library SALib <https://salib.readthedocs.io/en/latest/>, last access: 18 November 2021 (Herman and Usher, 2017), and the model calibrations are performed with the Python library scipy <https://scipy.org>, last access: 18 November 2021 (Virtanen et al., 2020).

*Data availability.* The structural model 3D-ALPS constrained in Spooner et al. (2019a) and used in the General-Focus model is freely available as DOI and online material via the following link <https://doi.org/10.5880/GFZ.4.5.2019.004> (Spooner et al., 2019b). The thermal field (3D-ALPS\_TR) generated in Spooner et al. (2020b) and used in the General-Focus model is freely available as DOI and online material via the following link <https://doi.org/10.5880/GFZ.4.5.2020.007> (Spooner et al., 2020a).

*Author contributions.* DD was responsible for the conceptualization, methodology, and software and writing the original draft. CS was responsible for the data curation and for writing, reviewing, and editing the manuscript. MSW and MC were responsible for the reviewing of the concepts and the reviewing and editing of the manuscript.

*Competing interests.* The authors declare that they have no conflict of interest.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

*Acknowledgements.* The authors gratefully acknowledge the Earth System Modelling Project (ESM) for funding this work by providing computing time on the ESM partition of the supercomputer JUWELS (Jülich Supercomputing Centre, 2019) at the Jülich Supercomputing Centre (JSC) under the application 16050 entitled "Quantitative HPC Modelling of Sedimentary Basin System". Furthermore, we would like to thank two anonymous reviewers for helping to improve this paper through their remarks and comments.

*Financial support.* This research has been supported by the ESM consortium.

The article processing charges for this open-access publication were covered by the Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences.

*Review statement.* This paper was edited by Rohitash Chandra and reviewed by two anonymous referees.

## References

- Aretz-Nellesen, N., Grepl, M. A., and Veroy, K.: 3D-VAR for parameterized partial differential equations: a certified reduced basis approach, *Adv. Comput. Math.*, 45, 2369–2400, 2019.
- Baroni, G. and Tarantola, S.: A General Probabilistic Framework for uncertainty and global sensitivity analysis of deterministic models: A hydrological case study, *Environ. Modell. Softw.*, 51, 26–34, 2014.
- Baş, D. and Boyacı, I. H.: Modeling and optimization I: Usability of response surface methodology, *J. Food Eng.*, 78, 836–845, 2007.
- Benner, P., Gugercin, S., and Willcox, K.: A Survey of Projection-Based Model Reduction Methods for Parametric Dynamical Systems, *SIAM Rev.*, 57, 483–531, 2015.
- Bezerra, M. A., Santelli, R. E., Oliveira, E. P., Villar, L. S., and Escalera, L. A.: Response surface methodology (RSM) as a tool for optimization in analytical chemistry, *Talanta*, 76, 965–977, 2008.
- Böhm, R., Auer, I., Schöner, W., Ganekind, M., Gruber, C., Jurkovic, A., Orlik, A., and Ungersböck, M.: Eine neue Webseite mit instrumentellen Qualitäts-Klimadaten für den Grossraum Alpen zurück bis 1760, *Wiener Mitteilungen*, 216, 7–20, 2009.
- Box, G. E.: Robustness in the Strategy of Scientific Model Building, in: *Robustness in statistics*, Elsevier, Amsterdam, The Netherlands, 201–236, 1979.
- Cannavó, F.: Sensitivity analysis for volcanic source modeling quality assessment and model selection, *Comput. Geosci.*, 44, 52–59, 2012.
- Cherpeau, N. and Caumon, G.: Stochastic structural modelling in sparse data situations, 21, 233–247, 2015.
- Cloke, H., Pappenberger, F., and Renaud, J.-P.: Multi-Method Global Sensitivity Analysis (MMGSA) for modelling floodplain hydrological processes, *Hydrological Processes: An International Journal*, 22, 1660–1674, 2008.
- Degen, D., Veroy, K., and Wellmann, F.: Certified reduced basis method in geosciences, *Computat. Geosci.*, 24, 241–259, <https://doi.org/10.1007/s10596-019-09916-6>, 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., 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.
- Doherty, J. E. and Hunt, R. J.: Approaches to highly parameterized inversion: a guide to using PEST for groundwater-model calibration, US Department of the Interior, US Geological Survey Reston, 2010.
- Elison, P., Niederau, J., Vogt, C., and Clauser, C.: Quantification of thermal conductivity uncertainty for basin modeling, *AAPG Bull.*, 103, 1787–1809, 2019.
- Fan, Y. and Van den Dool, H.: A global monthly land surface air temperature analysis for 1948 – present, *J. Geophys. Res.-Atmos.*, 113, <https://doi.org/10.1029/2007JD008470>, 2008.

- Feng, L., Palmer, P. I., Parker, R. J., Deutscher, N. M., Feist, D. G., Kivi, R., Morino, I., and Sussmann, R.: Estimates of European uptake of CO<sub>2</sub> inferred from GOSAT X<sub>CO<sub>2</sub></sub> retrievals: sensitivity to measurement bias inside and outside Europe, *Atmos. Chem. Phys.*, 16, 1289–1302, <https://doi.org/10.5194/acp-16-1289-2016>, 2016.
- Fernández, M., Egufá, P., Granada, E., and Febrero, L.: Sensitivity analysis of a vertical geothermal heat exchanger dynamic simulation: Calibration and error determination, *Geothermics*, 70, 249–259, 2017.
- Feyen, L. and Caers, J.: Quantifying geological uncertainty for flow and transport modeling in multi-modal heterogeneous formations, *Adv. Water Resour.*, 29, 912–929, 2006.
- Floris, F., Bush, M., Cuypers, M., Roggero, F., and Syversveen, A. R.: Methods for quantifying the uncertainty of production forecasts: a comparative study, *Petrol. Geosci.*, 7, S87–S96, 2001.
- 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, in: *Large-Scale Inverse Problems and Quantification of Uncertainty*, John Wiley and Sons, Hoboken, New Jersey, USA, 7, 123–149. <https://doi.org/10.1002/9780470685853.ch7>, 2010.
- Freyermark, J., Sippel, J., Scheck-Wenderoth, M., Bär, K., Stiller, M., Fritsche, J.-G., and Kracht, M.: The deep thermal field of the Upper Rhine Graben, *Tectonophysics*, 694, 114–129, 2017.
- Fuchs, S. and Balling, N.: Improving the temperature predictions of subsurface thermal models by using high-quality input data. Part 1: Uncertainty analysis of the thermal-conductivity parameterization, *Geothermics*, 64, 42–54, 2016.
- Ghasemi, M. and Gildin, E.: Model order reduction in porous media flow simulation using quadratic bilinear formulation, *Computat. 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.
- Grepl, M.: *Reduced-basis Approximation and A Posteriori Error Estimation for Parabolic Partial Differential Equations*, Ph.D. thesis, Massachusetts Institute of Technology, 2005.
- Herman, J. and Usher, W.: SALib: An open-source Python library for Sensitivity Analysis, *J. Open Source Softw.*, 2, 97, 2017 (code available at: <https://salib.readthedocs.io/en/latest/>, last access: 18 November 2021).
- Hesthaven, J. S., Rozza, G., and Stamm, B.: *Certified reduced basis methods for parametrized partial differential equations*, Springer-Briefs in Mathematics, Springer, Berlin, Germany, 2016.
- Hill, M. C. and Tiedeman, C. R.: *Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty*, John Wiley and Sons, Hoboken, New Jersey, USA, 2006.
- Houghton, J. T., Ding, Y., Griggs, D. J., Noguera, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C.: *Climate change 2001: the scientific basis*, The Press Syndicate of the University of Cambridge, Cambridge, UK, 2001.
- Iglesias, M. and Stuart, A. M.: *Inverse Problems and Uncertainty Quantification*, SIAM News, <https://homepages.warwick.ac.uk/~masdr/BOOKCHAPTERS/stuart19c.pdf>, 2–3, 2014.
- Jülich Supercomputing Centre: JUWELS: Modular Tier-0/1 Supercomputer, *Jülich Supercomputing Centre, JLSRF*, 5, A135, <https://doi.org/10.17815/jlsrf-5-171>, 2019.
- Kärcher, M., Boyaval, S., Grepl, M. A., and Veroy, K.: Reduced basis approximation and a posteriori error bounds for 4D-Var data assimilation, *Optim. Eng.*, 19, 663–695, 2018.
- Khuri, A. I. and Mukhopadhyay, S.: *Response surface methodology*, Wiley Interdisciplinary Reviews: Computational Statistics, 2, 128–149, 2010.
- Lehmann, H., Wang, K., and Clauser, C.: Parameter identification and uncertainty analysis for heat transfer at the KTB drill site using a 2-D inverse method, *Tectonophysics*, 291, 179–194, 1998.
- Lerch, F. J.: Optimum data weighting and error calibration for estimation of gravitational parameters, *B. Geod.*, 65, 44–52, 1991.
- Linde, N., Ginsbourger, D., Irving, J., Nobile, F., and Doucet, A.: On uncertainty quantification in hydrogeology and hydrogeophysics, *Adv. Water Resour.*, 110, 166–181, 2017.
- Locarnini, M. M., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., Reagan, J. R., Seidov, D., Weathers, K. W., Paver, C. R., and Smolyar, I.: *World ocean atlas 2013: Volume 1, Temperature*, <https://doi.org/10.7289/V55X26VD>, 2013.
- Luijendijk, E., Winter, T., Köhler, S., Ferguson, G., von Hagke, C., and Scibek, J.: Using thermal springs to quantify deep groundwater flow and its thermal footprint in the Alps and North American Orogens, *Geophys. Res. Lett.*, 47, <https://doi.org/10.1029/2020GL090134>, 2020.
- Magrin, A. and Rossi, G.: Deriving a new crustal model of Northern Adria: the Northern Adria Crust (NAC) model, *Front. Earth Sci.*, 8, 89, <https://doi.org/10.3389/feart.2020.00089>, 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, <https://doi.org/10.1016/j.cageo.2019.01.012>, 2019.
- Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768–772, 2004.
- Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M.: *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley and Sons, Hoboken, New Jersey, USA, 2016.
- 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.
- Permann, C. J., Gaston, D. R., Andrš, 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.
- 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, *J. Fluid Eng.*, 124, 70–80, 2002.
- Przybycin, A. M., Scheck-Wenderoth, M., and Schneider, M.: The 3D conductive thermal field of the North Alpine Foreland Basin: influence of the deep structure and the adjacent European Alps, *Geothermal Energy*, 3, 17, 2015.
- Quarteroni, A., Manzoni, A., and Negri, F.: *Reduced Basis Methods for Partial Differential Equations: An Introduction*, UNITEXT, Springer International Publishing, Berlin, Germany, 2015.
- Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., and Vanrolleghem, P. A.: Uncertainty in the environmental modelling process – a framework and guidance, *Environ. Modell. Softw.*, 22, 1543–1556, 2007.
- 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>, 2018.
- Rousset, M. A., Huang, C. K., Klie, H., and Durlofsky, L. J.: Reduced-order modeling for thermal recovery processes, *Comput. Geosci.*, 18, 401–415, 2014.
- Rozza, G., Huynh, D. B. P., and Patera, A. T.: Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations, *Arch. Comput. Methods E.*, 15, 229, <https://doi.org/10.1007/s11831-008-9019-9>, (2008).
- Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, *Comput. Phys. Commun.*, 145, 280–297, 2002.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S.: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Comput. Phys. Commun.*, 181, 259–270, 2010.
- Schaeffer, A. and Lebedev, S.: Global shear speed structure of the upper mantle and transition zone, *Geophys. J. Int.*, 194, 417–449, 2013.
- Schwarz, R., Pfeifer, N., Pfenningbauer, M., and Mandlbauer, G.: Depth Measurement Bias in Pulsed Airborne Laser Hydrography Induced by Chromatic Dispersion, *IEEE Geosci. Remote S.*, 18, 1332–1336, 2020.
- 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.
- Spoooner, C., Scheck-Wenderoth, M., Götze, H.-J., Ebbing, J., Hetényi, G., and the AlpArray Working Group: Density distribution across the Alpine lithosphere constrained by 3-D gravity modelling and relation to seismicity and deformation, *Solid Earth*, 10, 2073–2088, <https://doi.org/10.5194/se-10-2073-2019>, 2019a.
- Spoooner, C., Scheck-Wenderoth, M., Götze, H.-J., Ebbing, J., and Hetényi, G.: 3D ALPS: 3D Gravity Constrained Model of Density Distribution Across the Alpine Lithosphere. V. 2.0., GFZ Data Services [data set], <https://doi.org/10.5880/GFZ.4.5.2019.004>, 2019b.
- Spoooner, C., Scheck-Wenderoth, M., Cacace, M., and Anikiev, D.: 3D-ALPS-TR: A 3D thermal and rheological model of the Alpine lithosphere, GFZ Data Services [data set], <https://doi.org/10.5880/GFZ.4.5.2020.007>, 2020a.
- Spoooner, C., Scheck-Wenderoth, M., Cacace, M., Götze, H.-J., and Luijendijk, E.: The 3D thermal field across the Alpine orogen and its forelands and the relation to seismicity, *Global Planet. Change*, 193, 103288, <https://doi.org/10.1016/j.gloplacha.2020.103288>, 2020b.
- Tang, Y., Reed, P., Van Werkhoven, K., and Wagener, T.: Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis, *Water Resour. Res.*, 43, 6, <https://doi.org/10.1029/2006WR005813>, 2007.
- Trumpy, E. and Manzella, A.: Geothopica and the interactive analysis and visualization of the updated Italian National Geothermal Database, In: *J. Appl. Earth Obs.*, 54, 28–37, 2017.
- Turcotte, D. L. and Schubert, G.: *Geodynamics*, 2nd edn, Cambridge University Press, Cambridge, UK, 456 pp., 2002.
- van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R.: A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, 324, 10–23, 2006.
- 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, I., 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, *Nature methods*, 17, 261–272, 2020 (code available at: <https://scipy.org>, last access: 18 November 2021).
- Vogt, C., Mottaghy, D., Wolf, A., Rath, V., Pechinig, R., and Clauser, C.: Reducing temperature uncertainties by stochastic geothermal reservoir modelling, *Geophys. J. Int.*, 181, 321–333, 2010.
- Wagner, R. and Clauser, C.: Evaluating thermal response tests using parameter estimation for thermal conductivity and thermal capacity, *J. Geophys. Eng.*, 2, 349–356, 2005.
- 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.
- Wellmann, J. F. and Reid, L. B.: Basin-scale geothermal model calibration: Experience from the Perth Basin, Australia, *Energy Proced.*, 59, 382–389, 2014.
- Zehner, B., Watanabe, N., and Kolditz, O.: Visualization of gridded scalar data with uncertainty in geosciences, *Comput. Geosci.*, 36, 1268–1275, 2010.
- Zhan, C.-S., Song, X.-M., Xia, J., and Tong, C.: An efficient integrated approach for global sensitivity analysis of hydrological model parameters, *Environ. Modell. Softw.*, 41, 39–52, 2013.
- Zlotnik, S., Díez, P., Modesto, D., and Huerta, A.: Proper generalized decomposition of a geometrically parametrized heat problem with geophysical applications, *Int. J. Numer. Meth. Eng.*, 103, 737–758, 2015.