Solving crustal heat transfer for thermochronology using physics-informed neural networks

Ruohong Jiao[1], Shengze Cai[2], and Jean Braun[3]

[1]School of Earth and Ocean Sciences, University of Victoria, Victoria, Canada
[2]Institute of Cyber-Systems and Control, College of Control Science and Engineering, Zhejiang University, Hangzhou, China
[3]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany

Correspondence: Ruohong Jiao (rjiao@uvic.ca)

Abstract. We present a deep learning approach based on the physics-informed neural networks (PINNs) for estimating thermal evolution of the crust during tectonic uplift with a changing landscape. The approach approximates the temperature field of the crust with a deep neural network, which is trained by optimizing the heat advection-diffusion equation under boundary conditions such as initial and final thermal structure, topographic history, and surface and basal temperatures. From the trained neural network of temperature field and the prescribed velocity field, one can predict the temperature history of a given rock particle that can be used to compute the cooling ages of thermochronology. For the inverse problem, the forward model can be combined with a global optimization algorithm that minimizes the misfit between predicted and observed thermochronological data, in order to constrain unknown parameters in the uplift history or boundary conditions. We demonstrate the approach with solutions of one- and three-dimensional forward and inverse models of the crustal thermal evolution, which are consistent with results of the finite-element method. The three-dimensional model simulates the post-orogenic topographic decay of the Dabie Shan, China, with constraints from fission-track and (U-Th)/He ages.

1 Introduction

Thermochronology has been used extensively in earth sciences for quantifying ages and rates of geological and landscape evolution processes. To interpret the data, thermal structure of the crust must be understood. In the thermochronology community, analytical and finite-element methods have been used to estimate the temperature field, under either steady-state conditions or time-varying topography.

Recently, the deep learning method of physics-informed neural networks (PINNs) was proposed for solving partial differential equations, and its application has been demonstrated in many research fields including earth sciences. With rapidly increasing computing power, the method offers a valuable alternative to traditional analytical and numerical methods, with great potential to improve with the fast-evolving machine learning algorithm.

Here we demonstrate that PINNs can be used to solve the heat transfer equation in the crust for given uplift and landscape histories. Examples shown in this paper are implemented using the TensorFlow 2 library, with great flexibility to be adapted for complex boundary conditions.

2 From thermochronological ages to exhumation rates: existing methods

25 In this section we provide a brief overview of the widely used methods for interpretation of thermochronological data, highlighting in particular how they deal with the subsurface thermal field and its influence on the observed data.

2.1 Age-elevation relationship and thermal history modeling

Some conventional methods interpret thermochronological data without considering the thermal structure in the crust. The age-elevation relationship is an approach in low-temperature thermochronology for estimating the rock exhumation rates (e.g.,

30 Fitzgerald et al., 1995; Fitzgerald and Malusà, 2019), based on the variation of apparent ages between samples from different elevations. The method adopts the concept of "closure temperature" (Dodson, 1973), which considers the cooling age of a sample to reflect the time when it passed through a certain temperature. To estimate the exhumation rate, the method assumes that the closure temperature of a thermochronometer had remained at the same depth in the crust. Therefore, as rocks of different elevations, from higher to lower, should have passed the closure depth (i.e., depth of the closure temperature isotherm) con-

35 secutively, a positive correlation between the samples' cooling ages and elevations is expected. As a change in the exhumation rate can perturb the isotherms, leading to a transient state of the thermal field, the age-elevation relationship approach is only reliable when the exhumational steady state is achieved for a given thermochronometer (Willett and Brandon, 2002). In the case of steady state, cooling ages and elevations of samples from a topographic relief should yield a linear relationship whose slope is the exhumation rate for the period indicated by the thermochronological ages.

40 The age-elevation relationship can be interpreted in combination with thermal history modeling (e.g., Ketcham, 2005; Gallagher, 2012). The modeling predicts thermochronological ages for a given time-temperature path based on kinetic models of fission-track annealing or noble gas diffusion; the forward model is often used together with an optimization method (e.g., a Monte Carlo method) in order to estimate the optimal ("best-fit") thermal history for an observed dataset of thermochronology. For example, by assigning a temporally varying temperature offset between samples from different elevations, one can

45 simultaneously optimize the time-temperature paths of rock samples on a (near-)vertical profile as well as the history of the temperature offset. Assuming that the elevation offset between samples on the profile has remained constant, the optimal temperature offset history can be used to provide constraints on the evolution of the paleo-geothermal gradient (e.g., Jiao et al., 2014; Jepson et al., 2022). Then the exhumation rates can be derived from the modeled thermal history and geothermal gradient. This approach requires no assumption of a steady-state or monotonic cooling history. However, at any time in the parameter

50 space the modeling procedure uses linear interpolation to determine the temperatures of samples according to their positions on the profile, which is not consistent with the crustal thermal profile in the case of rapid exhumation.

2.2 Solving the subsurface thermal field

To ensure that the estimated exhumation rate is consistent with the physics of heat transfer, the thermal structure of the crust can be solved using analytical methods or numerical models in one, two, or three dimensions.

55 To elucidate the exhumation of the Olympic Mountains on the Cascadia Margin, Brandon et al. (1998) designed an approach
to compute the apatite fission-track age for one-dimensional (1D) steady-state exhumation. For a given exhumation rate, the
approach uses an analytical approximation (Stüwe et al., 1994) of the heat advection-diffusion equation to solve the thermal
profile of the crust, which is then used to find the exhumation rate and the closure temperature (dependent on cooling rate;
Dodson, 1973) through a numerical iteration. By using different kinetic parameters for track annealing or noble gas diffusion in

60 the closure temperature equation, this approach has be updated to estimate exhumation rates from various thermochronometers
(Reiners et al., 2003; Van Der Beek and Schildgen, 2023). To avoid the assumption of a steady-state scenario, Willett and
Brandon (2013) proposed an alternative approach to solve the thermal profile of earth over a half-space domain, which requires
constraints on the final geothermal gradient to quantify the transience of the thermal profile.

To solve the 1D heat transfer problem, the methods mentioned above all assume a constant exhumation rate for the time

65 period from the date indicated by the cooling age of thermochronology to the present-day. Fox et al. (2014) developed a linear
inversion method to infer the variation of exhumation rates in time and space. To estimate the temporal variation of rates, the
method also uses a 1D thermal model and the Dodson (1973)'s approximation but considers the closure depth as a summation
of rates over a finite number of time intervals. This discretization leads to more unknown parameters (i.e., exhumation rates)
than data (i.e., cooling ages) and is thus an underdetermined problem. Therefore, the solution of the problem through linear

70 inversion relies on some independent knowledge of the unknowns (e.g., *a priori* mean exhumation rate and the variance on this
mean) to construct the covariance matrix of the exhumation rates. In the covariance matrix, Fox et al. (2014) also introduces a
spatial correlation function in order to smooth the exhumation rates in space.

The 1D thermal models ignore the heat transport in horizontal directions. To capture the perturbation of isotherms by surface
topography, instead of a sample's true elevation the mean elevation filtered by the wavelength of topography (Stüwe et al., 1994;

75 Mancktelow and Grasemann, 1997) has been used in the 1D models for solving the thermal profile of the crust (Brandon et al.,
1998; Willett and Brandon, 2013; Fox et al., 2014; Van Der Beek and Schildgen, 2023). Two-dimensional (2D) numerical
modeling has also been used to solve the temperature fields under undulating topography during exhumation. Mancktelow and
Grasemann (1997) used a 2D finite-difference method to compute the transient isotherms to assess the influence of topography
and uplift on estimated exhumation rates in active orogens. To constrain the exhumation pattern and history of the central

80 Wasatch Mountains using thermochronological data, Ehlers et al. (2003) coupled a 2D velocity field and a thermal model to
compute the temperature field evolution in the mountains; the thermal model is solved using a finite-difference method. To
simulate the temperature field around hydrothermal systems, Luijendijk (2019) published a model to solve the subsurface heat
flow in the 2D sections using the finite-element method, which have been used to predict apatite fission-track and (U-Th)/He
data.

85 For simulating the three-dimensional (3D) crustal thermal field evolution, PECUBE (Braun, 2003; Braun et al., 2012) has
been the only computer code widely used in the low-temperature thermochronology community. This finite-element code
solves the 3D heat transfer equation in the crust under a prescribed velocity field and evolving topography, and uses the
solution to compute temperature histories of rocks as well as ages of various thermochronometers. The forward modeling is
often combined with the Neighborhood Algorithm (Sambridge, 1999a, b), an optimization method, to search for best-fit values

90 of the parameters used to define the input scenarios (i.e., tectonic or geomorphic) by iteratively minimizing the misfit between the predicted and observed data.

3 Physics-informed neural networks

The physics-informed neural network (PINN) was recently proposed to solve forward and inverse problems involving partial differential equations (PDEs) (Raissi et al., 2019). In the context of PINNs, a fully-connected neural network is generally

95 employed to approximate the solution of the PDEs, e.g., the temperature, $T(z,t)$, by taking the spatial and temporal coordinates (z and t) as the inputs. As demonstrated in Figure 1, the neural network is composed of multiple hidden layers with trainable parameters and nonlinear activation functions. The parameters in the network can be learned using a loss function based on the governing equations (i.e., PDEs) and the boundary and initial conditions of the PDEs. For inverse problems, we add a misfit term representing the difference between the network output and the observation; two separate optimizers are used for the loss

100 function and data misfit. In following sections we will demonstrate the networks case by case using numerical experiments.

A key procedure in PINNs is to compute the derivatives of the PDEs. To address this issue, PINNs use automatic differentiation (AD) to represent all differential operators that exist in the PDEs. The AD calculates the derivatives of the outputs with respect to the network inputs based on the chain rule, which is different from numerical computations and avoids discretization and truncation errors. After training with the loss function (and data misfit), the trained network can be used to approximate

105 the solution of PDEs at any location and any time.

PINNs have been applied in various disciplines such as fluid mechanics (Cai et al., 2021a; Raissi et al., 2020; Boster et al., 2023), and have demonstrated great potential in solving governing equations of physical laws in complex domains, especially in solving inverse problems. They have also attracted attention in the community of Earth sciences. For examples, Waheed et al. (2021) used PINNs to solve the eikonal equation in 2D for predicting traveltimes of seismic waves in isotropic and anistropic

110 media. He and Tartakovsky (2021) used PINNs to solve the advection-dispersion and Darcy flow equations in 1D and 2D fields with spatially varying hydraulic conductivity. Rasht-Behesht et al. (2022) used PINNs to solve the acoustic wave propagation and full waveform inversions under different and complex boundary conditions.

4 One-dimensional thermal profile: a synthetic model

4.1 Forward solution

115 The 1D heat advection-diffusion equation to solve is given by

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial z} - \kappa\frac{\partial^2 T}{\partial z^2} = 0, \tag{1}$$

where T is the temperature (°C), t is the time (Ma), u is the rock uplift rate (km/Myr), κ is the thermal diffusivity here set at 25 km^2/Myr, and z is the vertical position (km) relative to the model base ($z = 0$). In this model we assume no topographic

change over time, so u is equal to the exhumation rate. We solve the model in the space defined by

$$t \in [0, 50] \quad \text{and} \quad z \in [0, 20], \tag{2}$$

for which Dirichlet boundary conditions are imposed as

$$T(t, 0) = 500 \quad \text{and} \quad T(t, 20) = 0, \tag{3}$$

with a linear gradient prescribed for the initial condition, i.e.,

$$T(0, z) = 500 - 25z. \tag{4}$$

We assume that u is time-dependent and impose its time dependence in terms of a logistic function as

$$u(t) = u_0 + \frac{\Delta u_1}{1 + e^{-(t - t_1)/\Delta t}}, \tag{5}$$

in which $u_0 = 0.05$ km/Myr, $\Delta u_1 = 0.6$ km/Myr, $t_1 = 30$ Ma, and $\Delta t = 1$ Ma. Overall, the forward model describes the thermal profile evolution of a 20 km-thick crust from 0 to 50 Ma, with the surface and basal temperatures fixed at 0 and 500°C, respectively. The model uplift rate remained constant at 0.05 km/Myr from 0 to about 30 Ma, after which it increased by 0.6 to 0.65 km/Myr; Δt is an inertia factor that imposes a time length for transition between the two uplift rates (Figure 2a).

To solve the forward model, we approximate T as a deep neural network (DNN), and such that $T(t, z)$ can be learned by minimizing the loss function

$$\mathcal{L} = \mathcal{L}_f + \alpha \mathcal{L}_b, \tag{6}$$

where $\mathcal{L}_f$ penalizes the residual of the governing equations (i.e., PDEs) and $\mathcal{L}_b$ imposes the boundary conditions and initial conditions of the PDEs; α is a weighting parameter which we set to 1 for all experiments shown in this paper. Here the initial condition is considered as a special boundary condition. $\mathcal{L}_f$ and $\mathcal{L}_b$ are computed by

$$\mathcal{L}_f = \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, z_f^i)|^2, \tag{7}$$

$$\mathcal{L}_b = \frac{1}{N_b} \sum_{i=1}^{N_b} |T_b^i - T(t_b^i, z_b^i)|^2, \tag{8}$$

in which $f(t, z)$ is defined for the left-hand side of the heat transfer equation (Equation 1), $\{t_f^i, z_f^i\}_i^{N_f}$ depict the N_f collocation points for $f(t, z)$, and $\{t_b^i, z_b^i\}_i^{N_b}$ depict the N_b initial and boundary training data on $T(t, z)$.

In our demonstration, we use the TensorFlow 2 library (Abadi et al., 2015) to build the DNN, which includes three hidden layers with 20 neurons in each layer and an additional transform layer to impose the boundary conditions as hard constraints (Lagaris et al., 1998; Lu et al., 2021). The hyperbolic tangent function (*tanh*) is used as the activation function, and the *Adam* optimizer algorithm (Kingma and Ba, 2014) is used for minimizing the loss function (Equation 6). The model is set up with

145 1000 randomly sampled collocation points, and 150 points for the initial and boundary conditions (50 for the initial geotherms, 50 for the surface temperature, and 50 for the basal temperature)(Figure 2b).

The solution presented here was optimized through 8,000 iterations with a constant learning rate of 0.001. In the first 5,000 iterations, the model loss $\mathcal{L}$ has decreased rapidly from >100,000 to <10 (Figure 3). After that, $\mathcal{L}$ becomes more stable and is only reduced by 3 for the next 3,000 iterations. Figure 4 shows the thermal profiles at different model times predicted by the

150 DNN at various stages of the training, in comparison with the solution of the finite-element method (FEM) which used 101 points on the profile and a time step of 0.5 Ma. To compare the final solutions of the PINN and FEM, we compute the square and infinity norms of the misfits between the predictions of the two as

$$l^2 = \sqrt{\sum_{i=1}^{n} \epsilon_i^2}, \tag{9}$$

$$l^\infty = \max(|\epsilon_i|), \tag{10}$$

155 in which ϵ_i is the difference in temperatures predicted by the PINN and FEM at locations of each element and n is the number of elements used in FEM, i.e., 101 in this example. At model times of 5, 10, 25, and 50 Ma, l^2 are calculated at 30, 16, 40, and 45 °C, and l^∞ at 4, 2, 5, and 8 °C, indicating reasonable consistence between the solutions of the two methods (Figure 4). To test the influence on the thermochronological data, we also computed ages of different methods for the rock sample ending at the surface. The sample's thermal history (Figure 5a) is calculated from the trained DNN and the input uplift

160 function (Figure 2a). Based on the extracted time-temperature path, thermochronological ages are computed using an empirical model of the fission-track annealing (Ketcham et al., 2007) and the noble gas diffusion kinetics (summarized by Reiners and Brandon, 2006) solved by a finite-difference method (Braun et al., 2006). As a result, ages computed from thermal histories predicted by the PINN and FEM show good consistency (Figure 5b), with l^2 and l^∞ of the computed age differences at 1.3 and 0.8 Ma, respectively.

165 ## 4.2 Inverse problem

We transform the forward model to an inverse problem by considering u_0, Δu_1, and t_1 in the uplift function (Equation 5) as unknown parameters. All other parameters defining the thermal model remain as the same as in the forward problem (Section 4.1). Implementation of the DNN follows a similar setup as that of a forward model, except that the three unknown parameters are updated and optimized during the iterations by reducing the misfit between predicted and observed ages. The

170 age misfit function is defined by

$$\text{Misfit} = \frac{1}{N_a} \sum_{i=1}^{N_a} \left(\frac{p_i - o_i}{\sigma_i} \right)^2, \tag{11}$$

where p_i is the predicted age, o_i is the observed age, σ_i is the uncertainty on the observation, and N_a is the total number of observed ages. For the problem demonstrated here, synthetic age data predicted from the temperature history computed using

the FEM (Figure 5) are used as data, and a 10% uncertainty (standard deviation) is assumed for each data point. For calculation

175 of the synthetic observed data, 0.05 km/Myr, 0.6 /km/Myr, and 30 Ma, are used as "true" values of u_0, Δu_1, and t_1 in the uplift function (Equation 5), respectively.

While the *Adam* algorithm is used for fitting $\{t_f^i, z_f^i\}_i^{N_f}$ for loss functions from the physical law and boundary conditions, to avoid local minima the search for optimal values of the unknown parameters in the uplift function is conducted using a stochastic, derivative-free approach, i.e., the differential evolution (DE) algorithm (Storn and Price, 1997) implemented in the

180 SciPy library (Virtanen et al., 2020). DE maintains a population of candidate solutions and improves them iteratively according to a measure of the solution quality (e.g., a misfit function). The optimization procedure is not based on the gradient of the forward model, and is therefore suitable for solving multidimensional inverse problems that are non-differentiable, noisy, or variable through time. DE initializes the solution population by generating a set of random candidate solutions within the predefined space. During the optimization, for each base solution in the candidate population, a muted solution is created by

185 adding a weighted difference between two other random candidate solutions from the population to a third random one. The muted solution is then mixed with the base solution to create a crossover solution. The crossover and its corresponding base solution are compared, and only the one with a better quality (i.e., a lower misfit) is selected and kept in the population of candidate solutions. The mutation, crossover, and selection steps constitute a generation of DE, which is repeated many times to explore the parameter space to seek for optimal solutions with the highest quality.

190 Different strategies can be used for the initialization, mutation, and crossover steps of DE. For models presented in this paper, a quasi-random population is initiated using a Sobol sequence and updated using the *"best2bin"* strategy, in which the mutation and crossover are based on the currently best solution, two differential variants, and a binomial experiment (Storn and Price, 1997). Both differential weight and crossover constant are set to fixed values, both at 0.5. We combine *Adam* and DE by applying them at different frequencies: the optimization of the inverse problem contains n_g generations of DE with

195 a population size of n_p, and each candidate solution of the forward model is assessed every n_a number of *Adam* iterations. We tested the optimization performance with different configuration of n_g (120, 240, 100), n_p (128, 64, 32), and n_a (10, 10, 50), with all configurations resulting in a total of ~160,000 iterations. The learning rate for *Adam* is set at 0.001 for the first 100,000 iterations and lowered to 0.0001 for the rest of training. We restrict the searches of u_0, Δu_1, and t_1 within ranges of 0–2 km/Myr, 0–2 km/Myr, and 0–50 Ma, respectively.

200 The optimization performances are shown as evolution of the loss from physical law and boundary conditions (Figure 6a) and misfits between the predicted and observed thermochronological data (Figure 6b). For different combining configurations of *Adam* and DE, the parameter sampling performances are presented in Figure 7. For all three configurations tested, searches of the unknown parameters yield reasonable convergence towards the true values. The results also demonstrate that a larger population size of DE leads to better exploration of the parameter space (Figures 7a-c) whereas a higher value of n_a may help

205 increase the efficiency of convergence (Figure 7g-i).

Figure 8 shows the "best-fit" models in the final generations of DE for the three runs, which yield the lowest misfits to the age data. These models can reproduce the synthetic cooling ages of the six thermochronometers, with l^2 and l^∞ in the ranges of 1.1–1.5 Ma and 0.9–1.4 Ma, respectively. The "best-fit" models predict uplift and thermal histories very consistent with the

"true" model for the fast uplift period, but less accurate for the slow period, suggesting possible less constraints of the data for the older part of the history.

5 Three-dimensional temperature field: post-orogenic decay of the Dabie Shan

In this section, we use an example from the Dabie Shan, eastern China (~~Figuer~~ 9), to demonstrate the application of PINNs for solving the 3D temperature field under changing topography. The Dabie Shan is a Mesozoic mountain range which has undergone slow exhumation and topographic decay since the Late Cretaceous (Reiners et al., 2003; Braun and Robert, 2005).

5.1 Forward solution

Assuming that the rock motion is in the vertical direction, the 3D heat transfer equation is given by

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial z} - \kappa\Big(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\Big) = 0, \tag{12}$$

where x, y, and z are spatial coordinates and other parameters are the same as in Equation 1. We focus on the low-temperature and post-orogenic exhumation history of the mountain and solve its thermal history since 150 Ma. The model space is defined as

$$t \in [0, 150], \quad x \in [0, 159], \quad x \in [0, 199], \quad \text{and} \quad z \in [0, 36.5], \tag{13}$$

in which the z domain includes a crustal thickness ($h_c = 30$ km) below sea level and the surface elevation of the mountain range (h_s). Temperature at the model base is fixed at 600°C. The earth surface temperature is set at 15°C at sea level and calibrated according to the elevation using an atmospheric lapse rate at -5°C/km. Therefore, the boundary conditions are imposed as

$$T(t,x,y,0) = 600 \quad \text{and} \quad T(t,x,y,z_s(t,x,y)) = 15 - 5h_s(t,x,y), \tag{14}$$

in which z_s and h_s are the vertical coordinate and surface elevation, respectively, with a relationship

$$z_s = h_s + 30. \tag{15}$$

We follow Braun and Robert (2005) and simulate the surface relief change by modifying the amplitude of topography, such that

$$h_s(t,x,y) = w(t)h_s(0,x,y), \tag{16}$$

in which w is a time-dependent amplification factor for topography (Figure 10b), prescribed by

$$w(t) = \begin{cases} w_0, & t \in [0, t_d] \\ 1 + (w_0 - 1)\dfrac{150 - t}{150 - t_d}, & t \in (t_d, 150] \end{cases} \tag{17}$$

where w_0, set at 4, is the amplification of topography at the beginning of model ($t = 0$) and t_d, set at 70 Ma, is the model time when the post-orogenic decay of the topography starts. The present-day topography ($h_s(0, x, y)$) is extracted from the global

235   elevation model (GEBCO 2014) and resampled to 2 km resolution. Equation 17 describes a scenario in which between 150 and 80 Ma ago (i.e., model time from 0 to 70 Ma) the surface relief was four times as high as that of today, and after 80 Ma the topography has gradually decreased in a linear fashion towards the current level.

The initial thermal field is imposed using a linear interpolation of temperatures between the model base and earth surface, i.e.,

240    $$T(0, x, y, z) = T(0, x, y, 0) - \frac{T(0, x, y, 0) - T(0, x, y, z_s(0, x, y))}{z_s(0, x, y)} z. \tag{18}$$

Based on conclusions of Reiners et al. (2003) and Braun and Robert (2005), we assume a spatially uniform uplift model and define it using a piecewise function (Figure 10a),

$$u(t) = \begin{cases} u_0, & t \in [0, t_1] \\ u_1, & t \in (t_1, 150], \end{cases} \tag{19}$$

where $u_0 = 0.6$ km/Myr, $u_1 = 0.05$ km/Myr, and $t_1 = 40$ Ma. Therefore, on geological time, the rock uplift remained constant

245   at 0.6 km/Myr between 150 and 110 Ma, and then the rate decreased to 0.05 km/Myr and remained constant until the present-day.

Similar to the solution of the 1D model, we approximate T as a DNN and train it using the same loss function (Equation 6). The loss functions from 3D heat transfer law and boundary (and initial) conditions become

$$\mathcal{L}_f = \frac{1}{N_f} \sum_{i=1}^{N_f} |f(t_f^i, x_f^i, y_f^i, z_f^i)|^2, \tag{20}$$

250    $$\mathcal{L}_b = \frac{1}{N_b} \sum_{i=1}^{N_b} |T_b^i - T(t_b^i, x_b^i, y_b^i, z_b^i)|^2, \tag{21}$$

in which $f(t, x, y, z)$ stands for the left-hand side of the 3D heat transfer equation (Equation 12).

The DNN used for the 3D model consists of six hidden layers with 20 neurons on each and an additional layer to transform the outputs. The activation function, optimizer, and learning rate all remain the same as for the 1D model. To impose the initial and boundary conditions, we select 2,000 random points to set up the initial thermal field ($t = 0$; Figure 11a), whereas

255   2,000 and 5,000 randomly sampled points are used to represent the temperatures at the model base ($z = 0$) and on the surface topography ($z = z_s$), respectively (Figure 11b). We set up the model with 50,000 collocation points (Figure 11c), with half of them allocated randomly within the orogenic phase ($t \in [0, 40]$) and the other half post-orogenic ($t \in (40, 150]$).

The model is optimized through 100,000 iterations with learning rates of 0.01 and 0.001 for the first 20,000 and the rest iterations, respectively. Figure 12 shows the performance of the optimization. To verify the accuracy of the solution of the

260   PINN, we also computed the thermal field evolution of the Dabie Shan using a FEM code, i.e., PECUBE (Braun, 2003). The FEM model is set up on a $120 \times 120 \times 31$ ($x \times y \times z$) grid, and the heat transfer is solved at a time step of 1 Myr. The uplift history

and boundary conditions are set up as the same as in the PINN model. Figure 13 compares the temperature profiles on three transects across the model at 40 and 150 Ma (model time) solved by the the PINN and FEM. On each transect at 40 Ma, the l^2 and l^∞ of the misfits between the temperatures predicted by the two methods are in the ranges of 238–324 °C and 20–40 °C,

265 respectively, which are calculated at locations of 3100 elements; at 150 Ma when the topographic relief is lower, the two norms are reduced to 66–92 °C and 8–10 °C, respectively.

We also extracted thermal histories for all rock points that ended at the present-day surface and computed apatite fission-track and (U-Th)/He ages. The calculated ages based on the solutions from the PINN and FEM yield consistent patterns both in the map area (Figure 14) and on the elevation profiles (Figure 15). The l^2 and l^∞ norms for the AFT age misfits predicted

270 by the PINN and FEM, which are calculated using 42,000 surface samples, are 943 and 17 Ma, respectively; the two norms for AHe age misfits of the two methods are 424 and 12 Ma, respectively.

5.2 Inverse problem

We test the capability of PINNs in solving an inverse problem with real data. To define the problem, we set the onset time and rate of the post-orogenic uplift (t_1 and u_1 in Equation 19) and the initial topographic amplification (w_0 in Equation 17)

275 as unknown parameters. The three parameters are sampled in the ranges of 0–100 Ma, 0–0.1 km/Myr, and 0–6, respectively. To limit the number of dimensions for inversion, we keep the uplift rate during the orogenic phase (u_0) and the time of topographic decay (t_d) as fixed. Thermochronological age data reported by Reiners et al. (2003)(Figure 9b) are used to guide the optimization; only ages <150 Ma are used. Similar to the inversion of the 1D model, *Adam* and DE are combined for optimization. The inversion result presented here is based on 200 generations of DE with a population size of 32, and each

280 generation of DE contains 50 iterations of *Adam*. A total of 320,000 iterations were run.

As a comparison, we also searched the multi-dimensional space defined by the unknown parameters using the Neighborhood Algorithm (NA) (Sambridge, 1999a), in which the forward model is solved using PECUBE (Braun, 2003). The NA partitioned the parameter space into Voronoi cells, and during each iteration the method evaluates the model misfit according to the parameter values at the center of each cell. Then for the next iteration, new samples will be generated within a subset of cells

285 which contain the best-fitting models from the previous step. This process is repeated many times in order to find the optimal values of the unknown parameters that can reproduce the observed data. Here our NA search contains a total of 26 iterations, with 300 forward models in the first iteration and 60 in each of the rest. After evaluating all models in each step, out of the 60 cells we resample 50 with the best-fitting models to generate new combinations of parameter values for the next iteration.

Performance of the optimization of the PINN is shown in Figure 16, and sampled parameters are shown in Figure 17. It

290 appears that estimates of u_1 and t_1 using DE converge quickly to ranges of 0.04–0.06 km/Myr and 35–70 Ma (Figure 17a-b), respectively, whereas that of w_0 to >4 after 125 generations of DE (200,000 forward models; Figure 17c). These ranges are in general consistency with the results of the NA search (Figure 17). However, while parameter search using NA can converge continuously and eventually to a relatively narrow range, it appears that sampling using DE may reach a stable population and stop converging. Further evaluation of the optimization performance requires estimates of the probability functions of the

295 sampled values, which will be a direction for the future development (Section 6.2).

The results of our inverse modeling are in general consistency with tBraun and Robert (2005)'s conclusions based on inversion of the same dataset, which suggested that during the post-orogenic evolution since the Late Cretaceous, the Dabie Shan experienced slow uplift at 0.01–0.04 km/Myr and a topographic decay by a factor of 2.5 to 4.5. It is worth noting that the uplift in our model combines the background rock uplift and the isostatic compensation whereas Braun and Robert (2005)'s

300 model accounted for the two separately, and thus it is reasonable for our estimates of the uplift rate and the magnitude of erosion-induced topographic decay to be higher.

6 Discussion

6.1 Advantage of PINNs

We have shown that PINNs can be applied in the study of thermochronology, to solve heat transfer in the crust for providing

305 constraints on the uplift and exhumation processes of the crust. Our examples demonstrate that, built on the machine learning frameworks such as the TensorFlow and PyTorch, solutions of forward and inverse problems of crustal heat transfer are straightforward. The implementation is flexible and similar for both the forward and inverse problems, making it easy to impose and modify boundary conditions based on various assumptions and observations, such as topographic history, geothermal fields, and thermochronological data. Compared to the data assimilation methods which may require expansive computations

310 for an inverse problem, PINNs may be more efficient as the method solves the equations in one training process. In addition, PINNs can seamlessly integrate different kinds of experimental data (e.g., temperature data and velocity data) with the governing equations, making it possible to integrate multi-fidelity data in real experiments (Meng and Karniadakis, 2020).

6.2 Current limitations and future directions

Our examples show that, at least for a relatively simple time-space domain, a basic configuration of the optimization (e.g.,

315 *tanh* activation function, *Adam* optimizer, constant or piecewise learning rate, etc.) on a CPU platform can lead to efficient training of PINNs. However, we expect that for more complex domains that require a higher density of collocation points, better strategies may be required for training the models. On the one hand, it is an ongoing effort to improve the algorithm and setup for efficient optimization of PINNs (Jagtap et al., 2020). On the other hand, implementation of the PINNs on a multi-GPU cluster is a viable approach to increase the computation efficiency for larger problems (Shukla et al., 2021).

320 Our demonstrations have only considered the uplift history as space-independent rates in the vertical direction. Previous work (e.g., Cai et al., 2021b) has shown that implementation of a full 3D velocity field in the neural networks is possible. However, in cases with sharp changes in the velocity fields, the capability and efficiency of PINNs to solve PDEs remain to be tested or improved (e.g., Rasht-Behesht et al., 2022). This is crucial for tectonic studies in which the exhumation and geothermal patterns have been perturbed by fault activities, and therefore should be a focus of future development of the method.

325 Our current solution of the inverse problem includes a final solution of the heat model with values of unknown parameters. Although the direct search process (i.e., DE) results in an ensemble of sampled parameter values, they are not appraised

to quantify the uncertainties of the optimized parameters. To add estimates of the parameter uncertainties, one can employ a Bayesian neural network, where the parameters in the neural network follow Gaussian distributions, instead of a fully-connected network where the parameters are all deterministic in the PINNs framework (Yang et al., 2021). By doing this, the

330 Bayesian PINNs can adapt to noisy data and provide estimates of the uncertainties for the inferred parameters in the PDEs.

Code availability. Codes for examples shown in the paper, implemented using the TensorFlow 2 library, are available from https://github. com/jiaor/PINNs_Chron/.

Data availability. Apatite and zircon (U-Th)/He and apatite fission-track ages from the Dabie Shan were published by Reiners et al. (2003).

Author contributions. RJ conceived the idea, developed the code, carried out the tests. SC tested the codes. All authors wrote the manuscript

335 together.

Competing interests. The authors have declared no competing interests.

Acknowledgements. RJ's research is supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 2019-04405). We thank Stan Dosso for his feedback on the manuscript.

References

340 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Good-
fellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S.,
Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F.,
Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: large-scale machine learning on heterogeneous
systems, 2015.

345 Boster, K. A., Cai, S., Ladrón-de Guevara, A., Sun, J., Zheng, X., Du, T., Thomas, J. H., Nedergaard, M., Karniadakis, G. E., and Kelley,
D. H.: Artificial intelligence velocimetry reveals in vivo flow rates, pressure gradients, and shear stresses in murine perivascular flows,
Proceedings of the National Academy of Sciences, 120, e2217744 120, 2023.

Brandon, M. T., Roden-Tice, M. K., and Carver, J. I.: Late Cenozoic exhumation of the Cascadia accretionary wedge in the Olympic
Mountains, northwest Washington State, Bulletin of the Geological Society of America, 110, 985–1009, https://doi.org/10.1130/0016-
350 7606(1998)110<0985:LCEOTC>2.3.CO;2, 1998.

Braun, J.: Pecube: a new finite-element code to solve the 3D heat transport equation including the effects of a time-varying, finite amplitude
surface topography, Computers and Geosciences, 29, 787–794, https://doi.org/10.1016/S0098-3004(03)00052-9, 2003.

Braun, J. and Robert, X.: Constraints on the rate of post-orogenic erosional decay from low-temperature thermochronological data: Applica-
tion to the Dabie Shan, China, Earth Surface Processes and Landforms, 30, 1203–1225, https://doi.org/10.1002/esp.1271, 2005.

355 Braun, J., Beek, P. V. D., and Batt, G.: Quantitative thermochronology: numerical methods for the interpretation of thermochronological data,
Cambridge University Press, 1 edn., https://doi.org/10.1017/CBO9780511616433, 2006.

Braun, J., van der Beek, P., Valla, P., Robert, X., Herman, F., Glotzbach, C., Pedersen, V., Perry, C., Simon-Labric, T., and Prigent, C.:
Quantifying rates of landscape evolution and tectonic processes by thermochronology and numerical modeling of crustal heat transport
using PECUBE, Tectonophysics, 524-525, 1–28, https://doi.org/10.1016/j.tecto.2011.12.035, 2012.

360 Cai, S., Mao, Z., Wang, Z., Yin, M., and Karniadakis, G. E.: Physics-informed neural networks (PINNs) for fluid mechanics: A review, Acta
Mechanica Sinica, 37, 1727–1738, 2021a.

Cai, S., Wang, Z., Fuest, F., Jeon, Y. J., Gray, C., and Karniadakis, G. E.: Flow over an espresso cup: inferring 3-D velocity and pressure
fields from tomographic background oriented Schlieren via physics-informed neural networks, Journal of Fluid Mechanics, 915, A102,
2021b.

365 Dodson, M. H.: Closure temperature in cooling geochronological and petrological systems, Contributions to Mineralogy and Petrology, 40,
259–274, https://doi.org/10.1007/BF00373790, 1973.

Ehlers, T. A., Willett, S. D., Armstrong, P. A., and Chapman, D. S.: Exhumation of the central Wasatch Mountains, Utah: 2. Thermokine-
matic model of exhumation, erosion, and thermochronometer interpretation, Journal of Geophysical Research: Solid Earth, 108, 2173,
https://doi.org/10.1029/2001JB001723, 2003.

370 Fitzgerald, P. G. and Malusà, M. G.: Concept of the exhumed partial annealing (retention) zone and age-elevation profiles in thermochronol-
ogy, in: Fission-track thermochronology and its application to geology, edited by Malusà, M. G. and Fitzgerald, P. G., pp. 165–189,
Springer International Publishing, https://doi.org/10.1007/978-3-319-89421-8_9, 2019.

Fitzgerald, P. G., Sorkhabi, R. B., Redfield, T. F., and Stump, E.: Uplift and denudation of the central Alaska Range: A case study in the
use of apatite fission track thermochronology to determine absolute uplift parameters, Journal of Geophysical Research: Solid Earth, 100,
375 20 175–20 191, https://doi.org/10.1029/95JB02150, 1995.

Fox, M., Herman, F., Willett, S. D., and May, D. A.: A linear inversion method to infer exhumation rates in space and time from thermochronometric data, Earth Surface Dynamics, 2, 47–65, https://doi.org/10.5194/esurf-2-47-2014, 2014.

Gallagher, K.: Transdimensional inverse thermal history modeling for quantitative thermochronology, Journal of Geophysical Research: Solid Earth, 117, B02 408, https://doi.org/10.1029/2011JB008825, 2012.

380 He, Q. and Tartakovsky, A. M.: Physics-informed neural network method for forward and backward advection-dispersion equations, Water Resources Research, 57, https://doi.org/10.1029/2020WR029479, 2021.

Jagtap, A. D., Kawaguchi, K., and Karniadakis, G. E.: Adaptive activation functions accelerate convergence in deep and physics-informed neural networks, Journal of Computational Physics, 404, 109 136, 2020.

Jepson, G., Carrapa, B., George, S. W. M., Reeher, L. J., Kapp, P. A., Davis, G. H., Thomson, S. N., Amadori, C., Clinkscales, C., Jones, S.,
385 Gleadow, A. J. W., and Kohn, B. P.: Where did the Arizona-Plano go? Protracted thinning via upper- to lower-crustal processes, Journal of Geophysical Research: Solid Earth, 127, https://doi.org/10.1029/2021JB023850, 2022.

Jiao, R., Seward, D., Little, T. A., and Kohn, B. P.: Thermal history and exhumation of basement rocks from Mesozoic to Cenozoic subduction cycles, central North Island, New Zealand, Tectonics, 33, 1920–1935, https://doi.org/10.1002/2014TC003653, 2014.

Ketcham, R. A.: Forward and inverse modeling of low-temperature thermochronometry data, Reviews in Mineralogy and Geochemistry, 58,
390 275–314, https://doi.org/10.2138/rmg.2005.58.11, 2005.

Ketcham, R. A., Carter, A., Donelick, R. A., Barbarand, J., and Hurford, A. J.: Improved modeling of fission-track annealing in apatite, American Mineralogist, 92, 799–810, https://doi.org/10.2138/am.2007.2281, 2007.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, https://doi.org/10.48550/ARXIV.1412.6980, 2014.

Lagaris, I., Likas, A., and Fotiadis, D.: Artificial neural networks for solving ordinary and partial differential equations, IEEE Transactions
395 on Neural Networks, 9, 987–1000, https://doi.org/10.1109/72.712178, 1998.

Lu, L., Meng, X., Mao, Z., and Karniadakis, G. E.: DeepXDE: a deep learnin library for solving differential equations, SIAM Review, 63, 208–228, https://doi.org/10.1137/19M1274067, 2021.

Luijendijk, E.: Beo v1.0: numerical model of heat flow and low-temperature thermochronology in hydrothermal systems, Geoscientific Model Development, 12, 4061–4073, https://doi.org/10.5194/gmd-12-4061-2019, 2019.

400 Mancktelow, N. S. and Grasemann, B.: Time-dependent effects of heat advection and topography on cooling histories during erosion, Tectonophysics, 270, 167–195, https://doi.org/10.1016/S0040-1951(96)00279-X, 1997.

Meng, X. and Karniadakis, G. E.: A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems, Journal of Computational Physics, 401, 109 020, 2020.

Raissi, M., Perdikaris, P., and Karniadakis, G. E.: Physics-informed neural networks: A deep learning framework for solving forward and
405 inverse problems involving nonlinear partial differential equations, Journal of Computational physics, 378, 686–707, 2019.

Raissi, M., Yazdani, A., and Karniadakis, G. E.: Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations, Science, 367, 1026–1030, 2020.

Rasht-Behesht, M., Huber, C., Shukla, K., and Karniadakis, G. E.: Physics-informed neural networks (PINNs) for wave propagation and full waveform inversions, Journal of Geophysical Research: Solid Earth, 127, https://doi.org/10.1029/2021JB023120, 2022.

410 Reiners, P. W. and Brandon, M. T.: Using thermochronology to understand orogenic erosion, Annual Review of Earth and Planetary Sciences, 34, 419–466, https://doi.org/10.1146/annurev.earth.34.031405.125202, 2006.

Reiners, P. W., Zhou, Z., Ehlers, T. A., Xu, C., Brandon, M. T., Donelick, R. A., and Nicolescu, S.: Post-orogenic evolution of the Dabie Shan, eastern China, from (U-Th)/He and fission-track thermochronology, American Journal of Science, 303, 489–518, https://doi.org/10.2475/ajs.303.6.489, 2003.

415 Sambridge, M.: Geophysical inversion with a neighbourhood algorithm-I. Searching a parameter space, Geophysical Journal International, 138, 479–494, https://doi.org/10.1046/j.1365-246X.1999.00876.x, 1999a.

Sambridge, M.: Geophysical inversion with a neighbourhood algorithm-II. Appraising the ensemble, Geophysical Journal International, 138, 727–746, https://doi.org/10.1046/j.1365-246x.1999.00900.x, 1999b.

Shukla, K., Jagtap, A. D., and Karniadakis, G. E.: Parallel physics-informed neural networks via domain decomposition, Journal of Computational Physics, 447, 110 683, 2021.

Storn, R. and Price, K.: Differential evolution – a simple and efficient heuristic for global optimization over continuous Spaces, Journal of Global Optimization, 11, 341–359, https://doi.org/10.1023/A:1008202821328, 1997.

Stüwe, K., White, L., and Brown, R.: The influence of eroding topography on steady-state isotherms. Application to fission track analysis, Earth and Planetary Science Letters, 124, 63–74, https://doi.org/10.1016/0012-821X(94)00068-9, 1994.

425 Van Der Beek, P. and Schildgen, T. F.: Short communication: *age2exhume* – a MATLAB/Python script to calculate steady-state vertical exhumation rates from thermochronometric ages and application to the Himalaya, Geochronology, 5, 35–49, https://doi.org/10.5194/gchron-5-35-2023, 2023.

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.,

430 Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: fundamental algorithms for scientific computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Waheed, U. B., Haghighat, E., Alkhalifah, T., Song, C., and Hao, Q.: PINNeik: Eikonal solution using physics-informed neural networks, Computers & Geosciences, 155, 104 833, https://doi.org/10.1016/j.cageo.2021.104833, 2021.

435 Willett, S. D. and Brandon, M. T.: On steady states in mountain belts, Geology, 30, 175–178, https://doi.org/10.1130/0091-7613(2002)030<0175:OSSIMB>2.0.CO;2, 2002.

Willett, S. D. and Brandon, M. T.: Some analytical methods for converting thermochronometric age to erosion rate: age to erosion rate, Geochemistry, Geophysics, Geosystems, 14, 209–222, https://doi.org/10.1029/2012GC004279, 2013.

Yang, L., Meng, X., and Karniadakis, G. E.: B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems

440 with noisy data, Journal of Computational Physics, 425, 109 913, 2021.
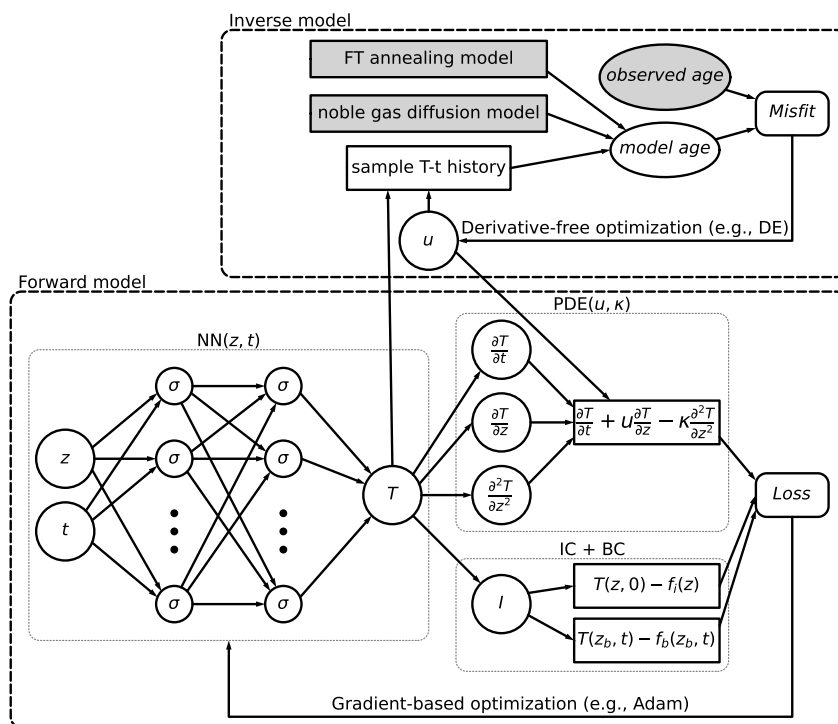
Figure 1. Work flow for forward and inverse modeling of 1D thermal profile for thermochronology. f_i and f_b are functions for the initial and boundary conditions of the model, respectively.
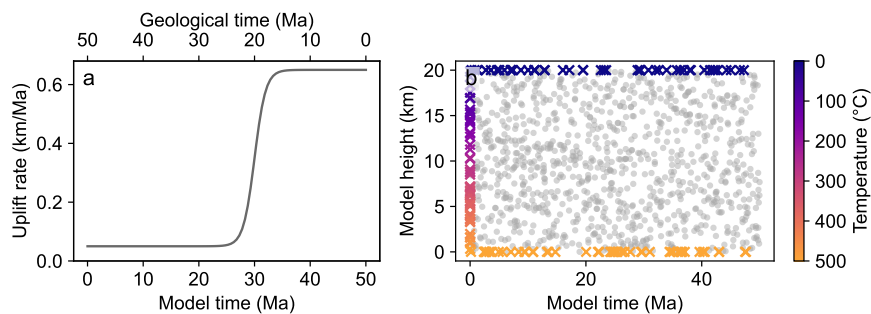
Figure 2. Setup of the 1D thermal model of the crust. (a) Uplift function of the model. (b) Collocation (gray dots) and boundary condition points (colored cross) for the PINN model.
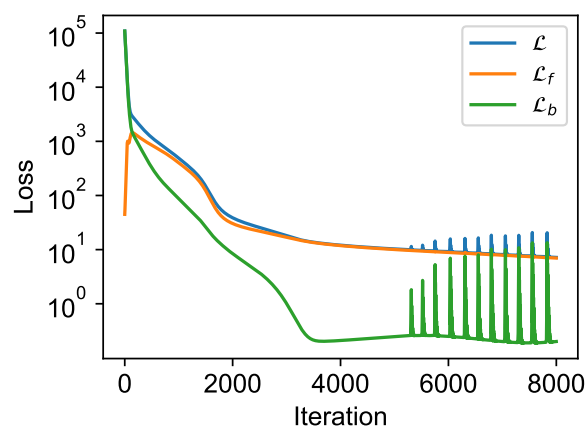
Figure 3. Loss curves for the 1D forward model.
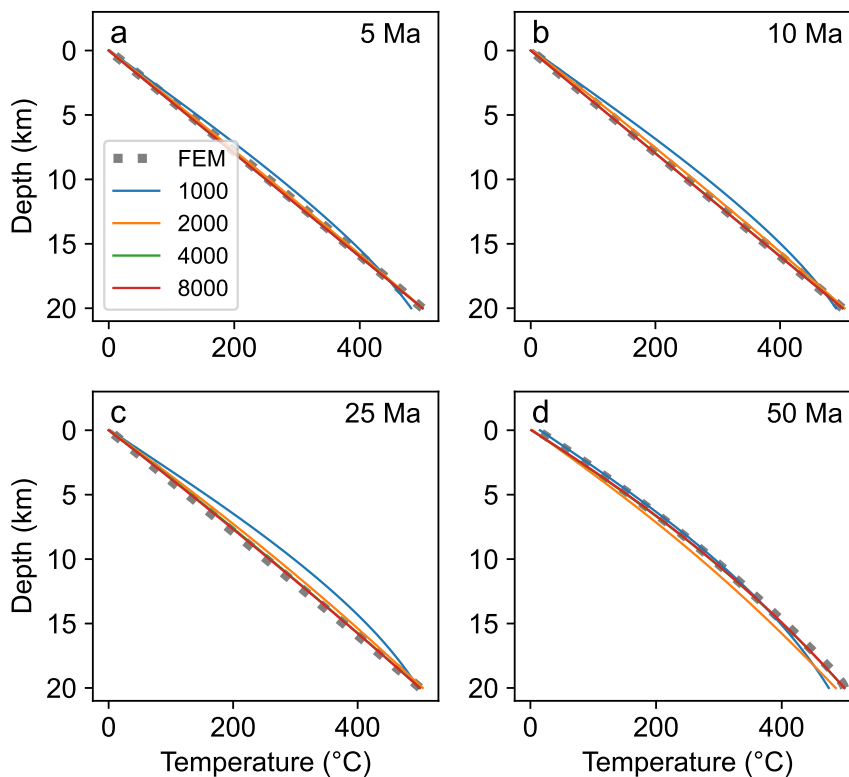
Figure 4. Predicted thermal profiles of the crust at different model times. Colored lines depict solutions using the PINNs after 1000, 2000, 4000, and 8000 iterations of training, respectively. Dashed line represents solution using the finite-element method (FEM).
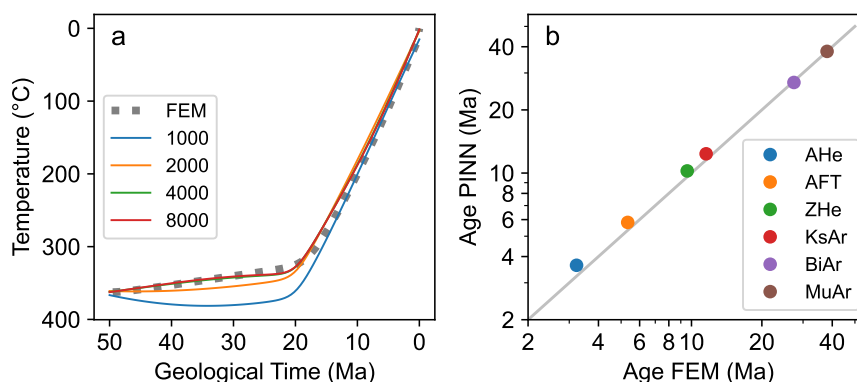
Figure 5. Predicted temperature-time paths and cooling ages for rock sample ended at the surface. (a) Predicted thermal history. Colored lines represent predictions by the PINNs model after 1000, 2000, 4000, and 8000 iterations of training, respectively. Dashed line represents prediction using FEM. (b) Predicted cooling ages, where AHe represents apatite (U-Th)/He; AFT, apatite fission-track; ZHe, zircon (U-Th)/He; KsAr, K-feldspar ^{40}Ar/^{39}Ar; BiAr, biotite ^{40}Ar/^{39}Ar; MuAr, muscovite ^{40}Ar/^{39}Ar.
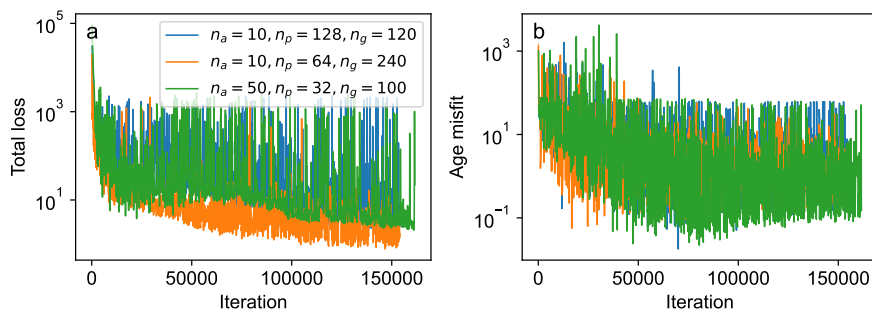
Figure 6. Loss and misfit curves for the 1D inverse model. For clear display, the curves are plotted using a randomly sampled subset (1%) of all iterations. (a) Loss curves for different configuration of the optimization. (b) Age misfit curves. n_g and n_p are generation number and population size of *differential evolution* (DE), respectively, and n_a is the *Adam* iteration number in each generation of DE.
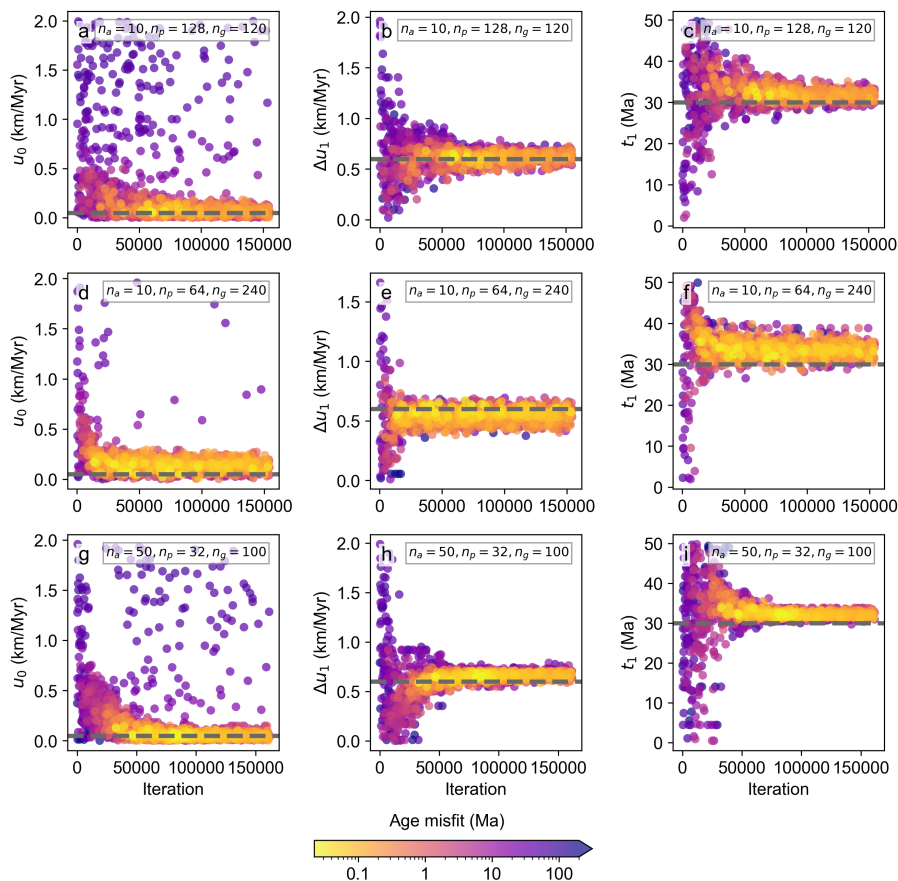
Figure 7. A random subset (1%) of all sampled parameter values for the 1D inverse model. Dots are color coded according to predicted age misfits.
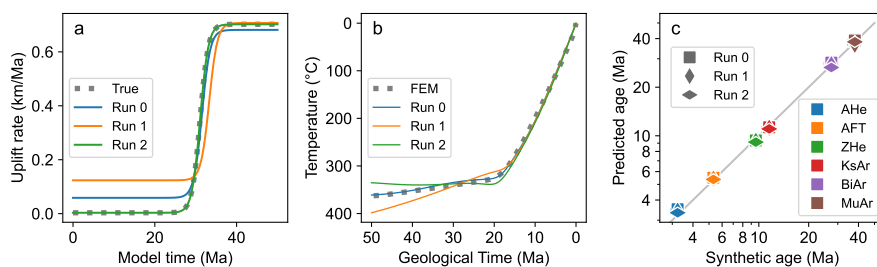
Figure 8. Prediction of the optimized 1D thermal model. (a) Uplift rate histories predicted by the three inverse models. Dashed line represents the true uplift model. (b) Predicted thermal histories. Dashed line represents solution with FEM. (c) Thermochronological ages predicted by PINNs versus synthetic data.
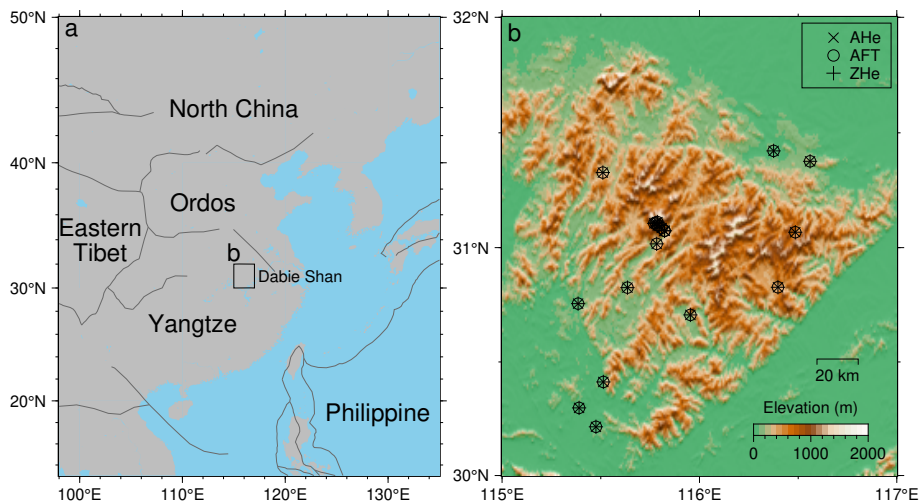
Figure 9. (a) Location of the Dabie Shan, eastern China. (b) Topography of the Dabie Shan and locations of thermochronological data.
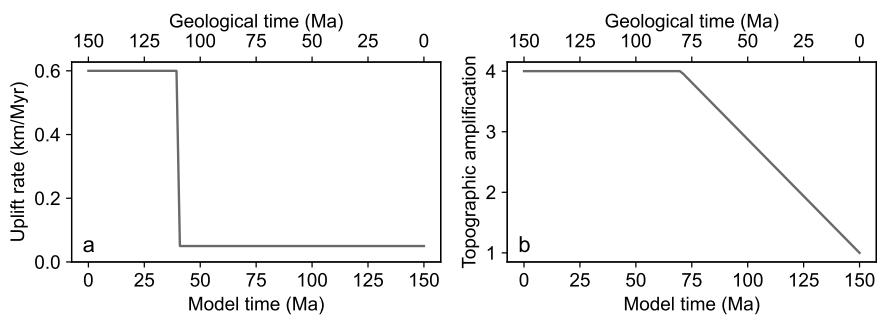
Figure 10. Inputs of the 3D thermal evolution model of the Dabie Shan. (a) Uplift model. (b) Topographic amplification model.
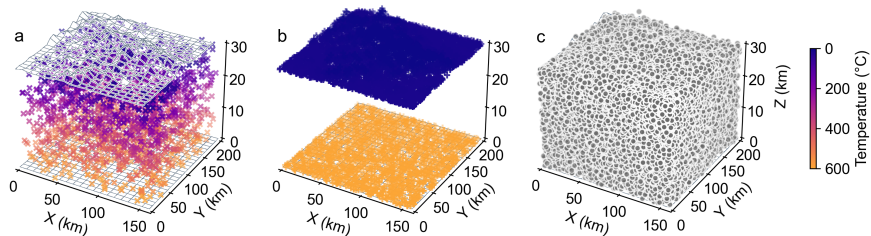
Figure 11. The initial and boundary conditions and collocation points for the 3D thermal model of the Dabie Shan. Control points are projected onto the $\{x, y, z\}$ space. (a) The initial thermal field. (b) Boundary temperatures at the surface and base of the model. (c) Collocation points.
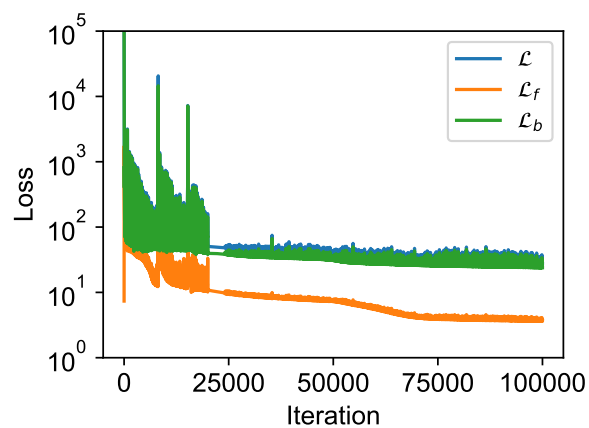
Figure 12. Loss curves for the 3D forward model.
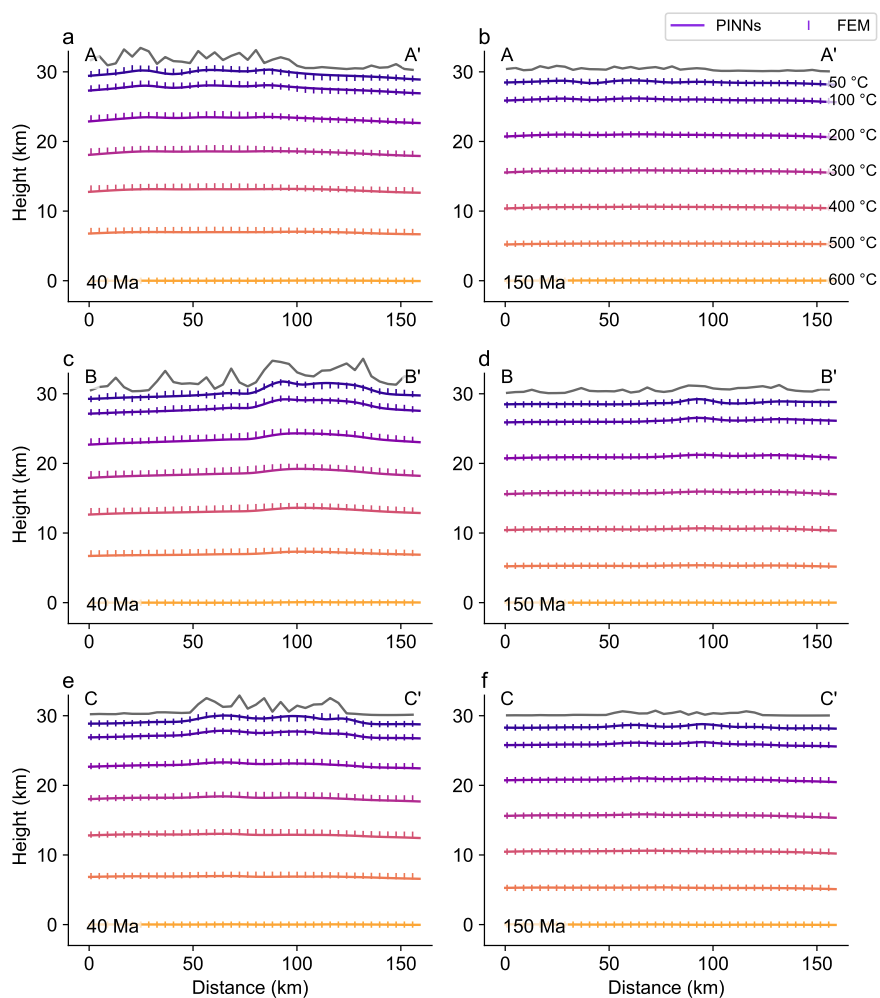
Figure 13. Predicted isotherms on three transects across the Dabie Shan at model times of 40 and 150 Ma. Solutions using the PINN and FEM are compared. Locations of the transects are shown in Figure 14.
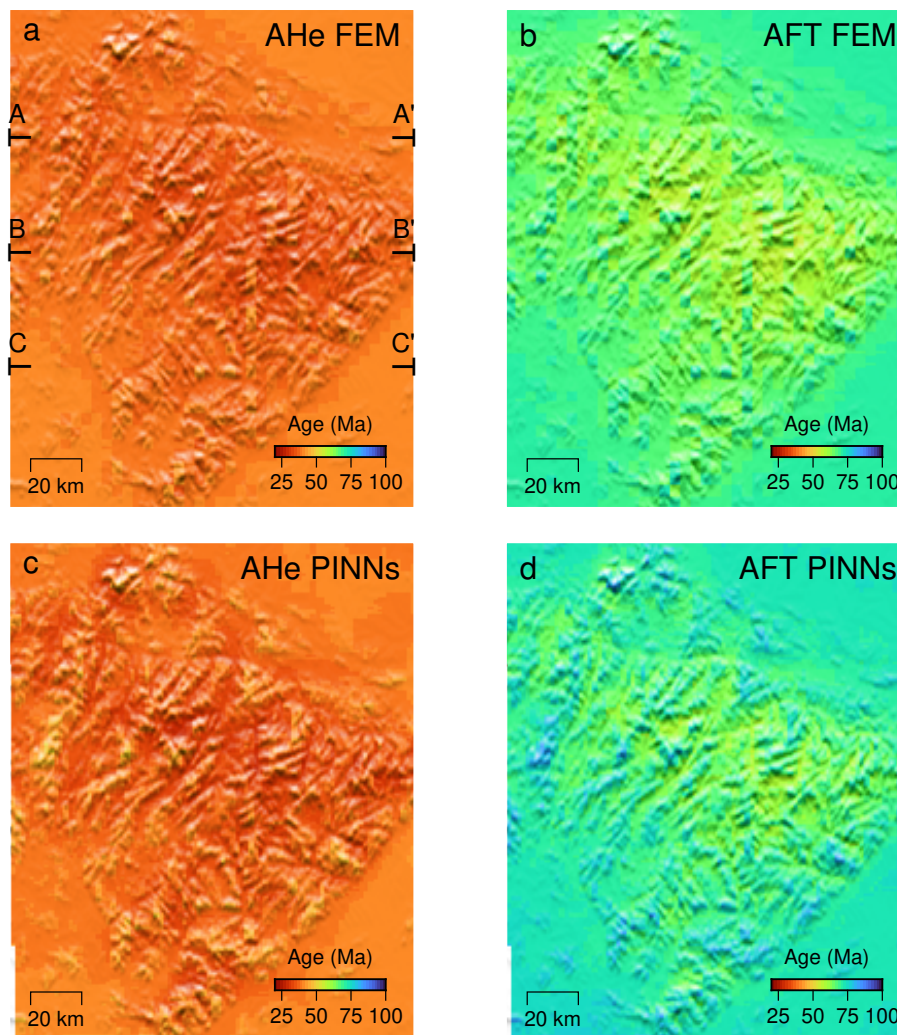
Figure 14. Predicted age maps of the Dabie Shan using the FEM and PINN. A-A', B-B', and C-C' indicate locations of the transects in Figure 13.
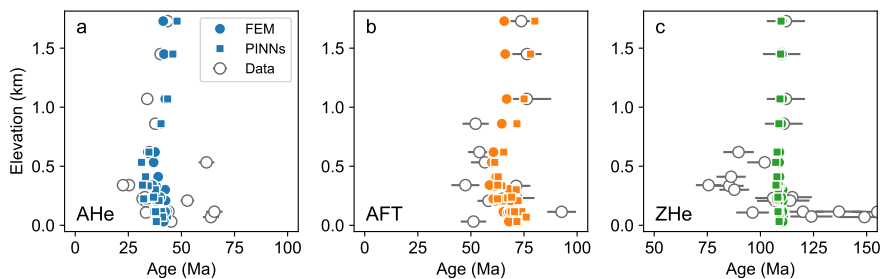
Figure 15. Predicted and observed thermochronological ages versus sample elevations.
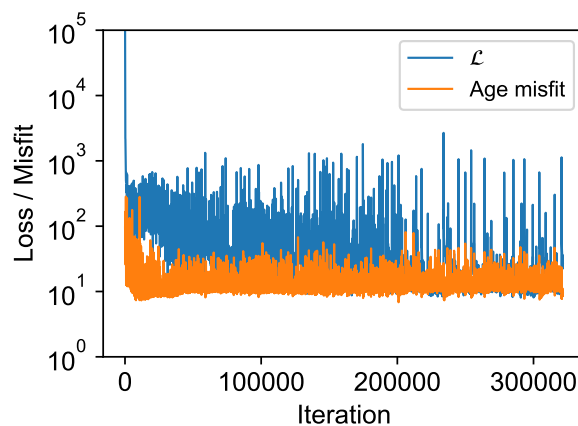
Figure 16. Loss and age misfit curves of the 3D inverse model.
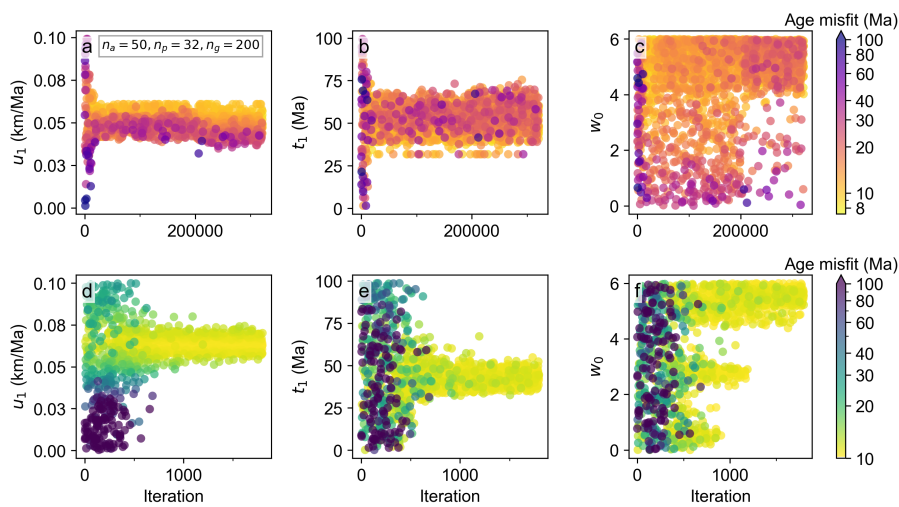
Figure 17. Sampled parameter values for the 3D inverse model using (a-c) the PINN coupled with the differential evolution and (d-f) the PECUBE coupled with the Neighborhood Algorithm (Sambridge, 1999a). The results of PINNs shown here are a random subset (1%) of all iterations.