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~~, assuming initial and boundary temperature conditions that follow a prescribed topographic history. 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 rock 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~~ As an example, the three-dimensional model simulates the exhumation and post-orogenic topographic decay of the Dabie Shan, ~~China, with constraints from fission-track and (U-Th)/He ages.~~ East China, whose post-orogenic evolution has been constrained by previous thermochronological data and models. This approach takes advantage of the computational power of machine learning algorithms, offering a valuable alternative to existing analytical and numerical methods, with great adaptability to diverse boundary conditions and easy integration with various optimization schemes.

1 Introduction

Thermochronology has been used extensively in earth sciences for ~~quantifying~~ estimating ages and rates of geological and landscape evolution processes. To interpret the data, how the thermal structure of the crust has changed over time and space must be understood. In the thermochronology community, analytical and ~~finite-element methods have been used to estimate the temperature field, under either~~ numerical methods have advanced the analysis and interpretation of data by providing quantitative estimates of temperature fields, tectonic deformation, and landscape changes under various scenarios (e.g., Stüwe et al., 1994; Braun, 2003). However, these methods also involve strict assumptions and limitations which hinder their application to geological problems in a broader range of geological and landscape settings. The analytical methods are computationally efficient but almost only applicable to problems with constant uplift rates. They also require either the assumption of a steady-state ~~conditions or time-varying topography.~~ temperature field (Stüwe et al., 1994; Brandon et al., 1998) or additional temperature data to quantify the transience of the field (Willett and Brandon, 2013). Numerical methods have the advantage of simulating the transient state of the thermal evolution (Braun, 2003; Braun et al., 2012), but implementing complex boundary conditions in numerical models is a significant challenge. Solving inverse problems with numerical models has a high computational demand which increases exponentially with the dimensionality of the space of unknown parameters.

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~~ (e.g., He and Tartakovsky, 2021; Rasht-Behesht et al., 2022). This mesh-free method benefits from the increasing capabilities of machine learning algorithms. Here we show that PINNs can be used to solve the heat transfer equation in the crust for given rock uplift and landscape histories. ~~Examples shown in this paper~~ Using example models, we show that PINNs can provide estimates of the thermal models with good agreement with numerical solutions, while allowing more flexibility in the model configuration. Moreover, PINNs also supports the simultaneous optimization of the forward and inverse problems. Therefore, these advantages of PINNs indicate the potential of PINNs for solving problems with more intricate boundary conditions. Our examples are implemented using the TensorFlow 2 library ~~, with great flexibility to be adapted for complex boundary conditions~~(Abadi et al., 2015) and global optimizers from the SciPy library (Virtanen et al., 2020), but we expect that similar results can be achieved using other machine learning libraries and optimization schemes.

2 From thermochronological ages to exhumation rates: existing methods

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~~ quantifying the thermal structure in the crust. The age-elevation relationship (AER) is an approach in low-temperature thermochronology for estimating the rock exhumation rates (e.g., Fitzgerald et al., 1995; Fitzgerald and Malusà, 2019), ~~based on the variation of apparent ages between samples from different elevations . The method adopts~~ using the apparent ages collected at different elevations on a vertical profile. Although this method is based on 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~~ it only assumes that the ~~closure temperature of a thermochronometerhad remained~~ timer, i.e., a thermochronometer, starts at the same depth in the crust and does not require to calculate the values of the closure temperature or its depth. Therefore, as rocks of different elevations, from higher to lower, should have passed the closure depth (i.e., depth of the closure temperature isotherm) consecutively, a positive correlation between the samples' cooling ages and elevations is expected. ~~As a~~ For a steady-state cooling history (Willett and Brandon, 2002), this should result in a linear relationship between apparent ages and sample elevations, with a slope equal to the exhumation rate for the time span indicated by the thermochronological ages. However, 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 . 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~~the AER approach is unreliable when the cooling rate is not constant or the thermal field is transient. In addition, application of the method is often limited by the difficulty of sampling on (near-)vertical profiles.

~~The age-elevation relationship can be interpreted in combination with thermal history modeling . The modeling predicts thermochronological ages~~ HeFTy (Ketcham, 2005) and QTQt (Gallagher, 2012) are widely used programs for estimating thermal history models using thermochronological data. The forward modeling functions in the programs can predict thermochronological data for a given ~~time-temperature path based on kinetic models of~~ thermal history path which is defined by a finite number of time-temperature points, based on simulating the noble gas production-diffusion or fission-track ~~annealing or noble gas diffusion; the~~ production-annealing kinetics. The forward model is often used together with ~~an optimization method (e.g.,~~ a Monte Carlo ~~method) in order~~ sampling method 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 simultaneously optimize~~ thermal history that suits observed thermochronological data. When the time-temperature paths of ~~rock~~ multiple samples on a ~~(near-)vertical profile as well as the history of the temperature offset. Assuming~~ vertical profile are made available by modeling, they can be used to infer changes in the geothermal gradient over time: 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~~. Then the exhumation rates can be derived from the modeled thermal history and geothermal gradient~~. Both QTQt and the recent version of HeFTy have the function to simultaneously sample the time-temperature-offset space for multiple samples on a vertical profile, and therefore the result will also yield an optimal history for the geothermal gradient (e.g., Jiao et al., 2014; Jepson et al., 2022). This approach requires no assumption of a steady-state or monotonic cooling history. However, neither of the two programs solves the heat transfer equations nor explicitly models the physical mechanisms that cause these changes. Moreover, at any time in the parameter 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~~Calculating crustal temperatures

To ensure that the estimated exhumation rate is consistent with the physics of heat transfer, the thermal structure of the crust can be solved as a forward model using analytical methods or numerical models in one, two, or three dimensions. For example, the finite-difference code Tc1D (Whipp, 2022) computes the 1D temperature model of the crust for various initial and boundary conditions, and can be applied to predict thermochronological ages under different tectonic and geomorphic processes.

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 the closure temperature equation, this approach has ~~be~~ been 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~~ in the previous paragraph all assume a constant exhumation rate for the time period from the date indicated by ~~the cooling age of thermochronology~~ a thermochronometer's cooling age 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 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; 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 rock uplift on estimated exhumation rates in active orogens. To constrain the exhumation pattern and history of the central 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~~ has been used to predict apatite fission-track and (U-Th)/He data.

~~For simulating the~~ In the low-temperature thermochronology community, simulation of the three-dimensional (3D) crustal thermal field evolution ~~, PECUBE has been the only computer code widely used in the low-temperature thermochronology community. This finite-element~~ has been mostly performed using PECUBE (Braun, 2003; Braun et al., 2012), a finite-element code written in FORTRAN. This 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~~ and ages of various thermochronometers. The ~~forward modeling is often combined with the~~ program has restricted formats for defining boundary conditions and integrating data, and therefore is not straightforward to impose complex boundary conditions that are not already implemented. For example, the code uses a simplified kink-band model to compute the rock velocity field driven by the displacement on a single fault or an array of faults with the same strike, and therefore cannot simulate other kinematic models without significant modifications. PECUBE can also operate in the inverse mode, which uses the Neighborhood Algorithm (Sambridge, 1999a, b), an optimization method coded in C, to search for the best-fit values 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~~parameters that specify the input tectonic or topographic scenarios. However, the search relies on solutions of a large number of forward problems, and thus can be very demanding for computational power for high-dimensional problems with a large finite-element grid or many time steps.

3 Physics-informed neural networks

~~The physics-informed neural network (PINN) was~~ Neural networks perform complex and nonlinear data operations by mimicking the structure and function of biological neurons. In deep learning, multiple neural layers are combined as deep neural networks (DNN) to decipher high-level information from the raw data. In scientific machine learning, deep learning is incorporated with fundamental physical laws, aiming to develop reliable, scalable, and physics-consistent machine learning models to facilitate new discoveries from scientific data. Physics-informed neural networks (PINNs) were 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 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~~. Two separate optimizers are used for the loss function and data misfit. In following sections we will demonstrate the setup of the neural 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 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~~example, Waheed et al. (2021) used PINNs to solve the eikonal equation in 2D for predicting traveltimes of seismic waves in isotropic and ~~anistropic~~ anisotropic 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

4.1.1 Method and problem setup

In our first example, the 1D forward model describes the thermal profile evolution of a 30 km-thick crust from 0 to 100 Myr, with the surface and basal temperatures fixed at 0 and 600°C, respectively. The model exhumation rate remained constant at 0.05 km/Myr from 0 to about 60 Myr, after which it increased to 0.6 km/Myr. The rapid exhumation continued for about 20 million years, and around 80 Myr the exhumation slowed down to 0.1 km/Myr and remained at this rate for the rest of the model time (Figure 2a).

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~~model time (Myr), 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

$$t \in [0, \cancel{50}100] \quad \text{and} \quad z \in [0, \cancel{20}30], \tag{2}$$

for which Dirichlet boundary conditions are imposed as

$$T(t,0) = \cancel{500}600 \quad \text{and} \quad T(t,\cancel{20}30) = 0, \tag{3}$$

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

$$T(0,z) = \cancel{500}600 - \cancel{25z}20z. \tag{4}$$

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

$$u(t) = u_0 + \cancel{\frac{\Delta u_1}{1+e^{-(t-t_1)/\Delta t}}}\frac{u_1 - u_0}{1+e^{-(t-t_1)/\Delta t}} + \frac{u_2 - u_1}{1+e^{-(t-t_2)/\Delta t}}, \tag{5}$$

in which $u_0 = 0.05$ km/Myr, ~~$\Delta u_1 = 0.6$~~ $u_1 = 0.6$ 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~~

~~500C, respectively. The model uplift rate remained constant at 0.05~~ $u_2 = 0.1$ km/~~Myrfrom 0 to about 30 Ma, after which it increased by 0.6 to 0.65 km~~/~~Myr;~~ Myr, $t_1 = 60$ Myr, $t_2 = 80$ Myr, and $\Delta t = 1$ Myr. Δt is an inertia factor that imposes a time length for ~~transition between the two uplift rates (Figure 2a)~~the transition between two rock uplift rates.

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~~. In our experiments presented in the paper, assigning α to 1 ~~for all experiments shown in this paper~~ensures effective optimization of the PINN models. 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, 2015) is used for minimizing the loss function (Equation 6). The model is set up with ~~1000~~ 2,000 randomly sampled collocation points~~, and 150 points for~~. Another 500 points are used to impose the initial and boundary conditions~~(50~~, including 100, 200, and 200 points for the initial ~~geotherms, 50 for~~ thermal structure, the surface temperature, and ~~50 for~~ the basal temperature~~)~~, respectively (Figure 2b).

~~The~~ For the solution presented here~~was optimized through 8~~, the DNN is optimized through 50,000 iterations~~with a constant learning rate of~~. The initial learning rate is set at 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 DNN at various stages of the training, in comparisonwith the solution of~~ which decays at a factor of 0.9 for every 1000 steps, i.e., at the ith step the learning rate is $0.001 \times 0.9^{i/1000}$.

As a comparison, we also solve the same problem using the finite-element method (FEM)~~which used~~, which provides estimates of the heat transfer equations by discretizing the time-space domain using 101 points on the crustal profile and a time step of ~~0.5 Ma~~1 Myr. 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~~

methods as

$$l^2 = \cancel{\sqrt{\sum_{i=1}^{n}\epsilon_i^2},}\sqrt{\sum_{i=1}^{n}(T^i_{pinn}-T^i_{fem})^2} \text{ and} \tag{9}$$

$$l^\infty = \max_{i=1}^{n}(|\cancel{\epsilon_i}T^i_{pinn}-T^i_{fem}|), \tag{10}$$

~~in which ϵ_i is the difference in~~ in which T^i_{pinn} and T^i_{fem} are the temperatures predicted by the PINN and FEM at locations of each element used in FEM and n is the total number of elements~~used in FEM~~. Based on the temperature histories estimated, we also predict the thermochronological ages of the present-day surface rocks 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). Then the percent errors in ages between the predictions are calculated as

$$\delta = \left|\frac{a_{pinn}-a_{fem}}{a_{fem}}\right| \times 100\%, \tag{11}$$

in which a_{pinn} and a_{fem} are ages calculated by the thermal histories estimated using the PINN and FEM, respectively.

4.1.2 Results

In the first 3,000 iterations, the model loss $\mathcal{L}$ has decreased rapidly from >10,~~i.e., 101 in this example. At~~ 000 to <10 (Figure 3). After that, the $\mathcal{L}_{|}$ experiences some fluctuations but $\mathcal{L}$ continues to decrease smoothly and becomes more stable after 30,000 iterations. Figure 4 shows the thermal profiles at different model times predicted by the PINN at various stages of the training, in comparison to the FEM solutions. Between the final PINN and the FEM solutions at model times of ~~5, 10, 25, and 50 Ma,~~ 40, 70, and 100 Myr, the l^2 ~~are calculated at 30, 16, 40, and 45~~calculated for 101 elements are 3.8, 51.0, and 30.0 °C, ~~and~~ respectively, and the l^∞ ~~at 4, 2, 5, and 8~~are 0.6, 11.4, and 3.5 °C, ~~indicating reasonable consistence between the solutions of the~~ respectively, indicating reasonable consistency between 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~~ For the rock at the present-day surface, the PINN also predicts a time-temperature path very consistent with the FEM result (Figure 5a)~~is calculated from the trained DNN and the input uplift function~~ , with a temperature difference <20 °C for any point on the path and <10 °C for most part of the history (Figure ~~2a). Based on the extracted time-temperature path, thermochronological ages are computed using an empirical model of the~~ 5b). For the same rock, the predicted difference in cooling ages between the two modeling methods are <7% for apatite (U-Th)/He and apatite fission-track ~~annealing and the noble gas diffusion kinetics solved by a finite-difference method . As a result, ages computed from thermal histories predicted by the PINN and FEM show good consistency~~ and <3% for other thermochronometers (Figure 5~~b), with l^2 and l^∞ of the computed age differences at 1.3 and 0.8 Ma, respectively~~d).

4.2 Inverse problem

4.2.1 Method and model setup

We transform the forward model to an inverse problem by considering u_0, ~~Δu_1, and~~ u_1, u_2, t_1 ~~in the~~ , and t_2 in the rock 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~~ five unknown parameters are updated and optimized during the iterations by reducing the misfit between predicted and observed ages. The age misfit function is defined by

$$\cancel{\mathrm{Misfit}}\,\phi = \frac{1}{N_a}\sum_{i=1}^{N_a} \cancel{\frac{p_i - o_i}{\sigma_i}} \left(\frac{a_p^i - a_o^i}{\sigma_o^i}\right)^2, \tag{12}$$

where ~~p_i~~ a_p^i is the predicted age, ~~o_i~~ a_o^i is the observed age, ~~σ_i~~ σ_o^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 of the synthetic observed data, 0.05 km/Myr, 0.6 km/Myr, 0.1 km/Myr, ~~and 30 Ma,~~ 60 Myr and 80 Myr are used as "true" values of u_0, ~~Δu_1, and~~ u_1, u_2, t_1 ~~in the~~ , and t_2 in the rock uplift function (Equation 5), respectively.

~~While the~~ The *Adam* algorithm is used ~~for fitting~~ to fit $\{t_f^i, z_f^i\}_i^{N_f}$ for loss functions derived from the physical law and boundary conditions~~, to avoid local minima~~. To circumvent local minima, the search for optimal values of the unknown parameters in the rock uplift function is conducted using a ~~stochastic, derivative-free approach, i.e., the differential evolution (DE) algorithm~~ modified Lipschitzian approach, DIRECT (Jones et al., 1993; Jones and Martins, 2021), which is a global, derivative-free optimization algorithm implemented in the 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 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.~~ DIRECT operates by progressively partitioning the search space into smaller hyperrectangles and evaluating the objective function at the center of each. In the subsequent iteration, hyperrectangles with optimal solutions are selected and subdivided for further evaluation. This process is repeated to locate the global minimum in the search space.

~~Different strategies can be used for~~ The configuration of the DIRECT algorithm requires adjusting a single parameter, ϵ, which defines the minimal acceptable improvement in the function value from the current best solution to the next potentially optimal hyperrectangle for division. A larger ϵ guides the search to explore a broader domain, whereas a smaller ϵ directs a more local exploitation. To optimize the memory usage in our application, we use a strategy in which, after a certain number of iterations, if a parameter value of the optimal solution has no significant change (less than 5% within the search domain) 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 . 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 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 (~~DIRECT search recommences within a narrowed search space, reducing the domain for this parameter by 10%.

More specifically, we define the initial search space for the parameters u_0, u_1, ~~10~~u_2, ~~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, and t_2 within ranges of 0–2 km/Myr, 0–2 km/~~Myr, and 0–50 Ma, respectively.~~ Myr, 0–2 km/Myr, 0–100 Myr, and 0–100 Myr, respectively. We start the *Adam* optimization of the DNN model by assigning the midpoint values of the search space to the unknown parameters. After n_{adam} iterations, we use the refined DNN model to explore the space of unknown parameters using the DIRECT approach over n_{direct} iterations. This process is repeated for multiple cycles to optimize the DNN model and seek the optimum parameter values . In the presented example, the initial n_{adam} is set to 100 and increased by 50% for each DIRECT cycle, up to a ceiling of 10,000. The learning rate for *Adam* follows the same schedule used for the forward model, i.e., at the ith iteration the learning rate is $0.001 \times 0.9^{i/1000}$. For each cycle, ϵ is maintained at 0.1, and the maximum number of DIRECT iterations, n_{direct}, is limited to 30 with the maximum number of function evaluated fixed at 1,000. Cumulatively, the optimization encompasses >55,000 *Adam* iterations and >290 DIRECT iterations, based on >15,000 age misfit function evaluations.

We also address the inverse problem using a Markov chain Monte Carlo (MCMC) approach. In the MCMC sampling process, the space of unknown parameters is searched iteratively, and at each iteration the parameter values with higher likelihood are accepted to the Markov chain. After sampling, the chain will be used to estimate the probability distribution of the parameters. To evaluate the parameter values, the forward thermal model is simulated with the FEM based on 101 elements on the profile and a time step of 1 Myr. The ~~optimization performances are~~ log-likelihood of a forward model is calculated by comparing the predicted thermochronological ages with the observation, as

$$\log(L) = -\sum_{i=1}^{N_a} \left(\frac{\ln(2\pi)}{2} + \ln(\sigma_o^i) + 0.5 \left(\frac{a_p^i - a_o^i}{\sigma_o^i} \right)^2 \right). \tag{13}$$

In our analysis, the space of five unknown parameters is explored by an affine invariant ensemble sampler (Goodman and Weare, 2010) within predefined domains. The search is conducted by 20 walkers, with each starting from a random location in the parameter space and moving for 50,000 steps. Ultimately, the generated chain comprises 1,000,000 combinations of parameter values from the search space.

4.2.2 Results

The optimization performance is shown as evolution of the loss from the physical law and boundary conditions (Figure 6a)~~and~~ , as well as 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~~ The loss function curve, despite minor fluctuations between DIRECT cycles (Figure 6a), shows a reduction pattern similar to that of the forward model (Figure 3). Figure 7 shows the sampled values used to evaluate the age misfit function. During the DIRECT search, the minimum age misfits decrease progressively until around 200 DIRECT iterations (Figure 6b), after which the age misfits and parameter values of the optimal solution stabilize (Figure 7). Finally, the searches of all five parameters are narrowed to ranges centered around the "true" values, except a relatively larger uncertainty on the estimated rate of the rapid uplift, u_1. ~~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 increase the efficiency of convergence .~~

The MCMC sampling coupled with FEM model has also identified optimal values for the unknown parameters. Among the 20 walkers deployed, three failed to reach equilibrium at the end of the sampling, whereas the remaining walkers have achieved stationary states after around 17,000 iterations (Figure 8). Similar to the result of the DIRECT search, a large uncertainty exists for the estimate of u_1 (Figure ~~7g–i~~8b).

Figure ~~9 shows the~~ 9 presents predictions of the optimized PINN model and the "best-fit" ~~models in the final generations of DE for the three runs, which yield the lowest misfits to the age data. These modelscan 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~~ FEM model, the model with the maximum likelihood from the MCMC chain. For both models, the predicted cooling paths of the rock at the current surface align closely with the "true" thermal history (Figure 9a). The temperature discrepancy at any given time between the PINN solution and the "true" thermal history is <17°C and typically <8°C (Figure 9b), whereas that between the "best-fit" ~~models predict uplift and thermal histories very consistent with~~ FEM model and 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~~ thermal history is <25°C. Moreover, the optimized PINN and "best-fit" FEM models predict similar thermochronological ages for the surface sample, both consistent with the data computed from the synthetic "true" history (Figure 9c), with the maximum difference at 5.3% and 3.7%, respectively (Figure 9d).

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~~(Figure 10) ~~,~~ to demonstrate the application of PINNs for solving the 3D temperature field under changing topography. Compared to the 1D model, the 3D model requires configuration of more complex boundary conditions and is computationally more intensive. The Dabie Shan in East China is a Mesozoic mountain range ~~which has undergone slow~~ that underwent rapid exhumation during the Triassic-Jurassic orogeny along the boundary between the North and South China blocks (Nie et al., 1994; Hacker et al., 1995; Ratschbacher et al., 2006). Based on apatite and zircon (U-Th)/He and apatite fission-track ages, Reiners et al. (2003) suggested that since the Late Cretaceous, the post-orogenic exhumation of the Dabie Shan has been slow and could be constant at a rate of 0.01–0.06 km/Myr. By investigating the same dataset using numerical methods, Braun and Robert (2005) estimated the post-orogenic exhumation and topographic ~~decay~~ evolution of the Dabie Shan, and inferred that since the Late Cretaceous~~.~~, the mountain range has been exhumed at a mean rate of 0.01–0.04 km/Myr, and that its topographic relief has reduced by a factor of 2.5–4.5 during the last 60–80 million years. Therefore, thermochronological data provide good constraints on the post-orogenic exhumation of the Dabie Shan, which can be simulated by a simple model with constant exhumation rates, making the mountain range an ideal natural laboratory for testing new thermochronological modeling methods (e.g., Fox et al., 2014).

5.1 Forward solution

5.1.1 Method and model setup

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

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial z} - \kappa\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) = 0, \tag{14}$$

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~~ last phase of the Mesozoic orogeny and the post-orogenic exhumation history of the mountain~~and~~, and therefore solve its thermal history since 150 Ma. The model space is defined as

$$t \in [0, 150], \quad x \in [0, 159], \quad \cancel{x}y \in [0, 199], \quad \text{and} \quad z \in [0, 36.5], \tag{15}$$

in which the range of the z domain ~~includes~~ is configured to include a crustal thickness ($h_c = 30$ km) below sea level and the ~~surface~~ potential maximum elevation of the mountain range during the model run (h_s). Temperature at the ~~model base~~ base of the model 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

$$T(t,x,y,0) = 600 \quad \text{and} \quad T(t,x,y,z_s(t,x,y)) = 15 - 5h_s(t,x,y), \tag{16}$$

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

$$z_s = h_s + 30. \tag{17}$$

~~We follow and~~

To simulate the surface relief change~~by modifying the amplitude of topography, such that~~, we follow Braun and Robert (2005) and define a scenario, in which between 0 to 70 Myr (equivalent to geological time 150 and 80 Ma) the surface relief was four times as high as that of today, and after that the topography has gradually decreased in a linear fashion towards the current level. The topography is imposed as

$$h_s(t,x,y) = w(t)h_s(0,x,y), \tag{18}$$

in which w is a time-dependent amplification factor for topography (Figure 11b), prescribed by

$$w(t) = \begin{cases} w_0, & t \in [0, t_d] \\ 1 + (w_0 - 1)\dfrac{150 - t}{150 - t_d}, & t \in (t_d, 150] \end{cases} \tag{19}$$

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~~ Myr, 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 elevation model (GEBCO 2014) and resampled to 2 km resolution. ~~Equation 19 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.,

$$T(0,x,y,z) = T(0,x,y,0) - \frac{T(0,x,y,0) - T(0,x,y,z_s(0,x,y))}{z_s(0,x,y)} z. \tag{20}$$

Based on conclusions of Reiners et al. (2003) and Braun and Robert (2005), we assume a spatially uniform rock uplift model and define it using a piecewise function (Figure 11a),

$$u(t) = \begin{cases} u_0, & t \in [0, t_1] \\ u_1, & t \in (t_1, 150], \end{cases} \tag{21}$$

where $u_0 = 0.6$ km/Myr, $u_1 = 0.05$ km/Myr, and $t_1 = 40$ ~~Ma~~Myr. Therefore, ~~on geological time,~~ the rock uplift remained constant at 0.6 km/Myr between ~~150 and 110 Ma~~0 and 40 Myr, and then the rate decreased to 0.05 km/Myr and remained constant until the ~~present-day~~end of model.

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

$$\mathcal{L}_f = \frac{1}{N_f}\sum_{i=1}^{N_f} |f(t_f^i, x_f^i, y_f^i, z_f^i)|^2, \tag{22}$$

$$\mathcal{L}_b = \frac{1}{N_b}\sum_{i=1}^{N_b} |T_b^i - T(t_b^i, x_b^i, y_b^i, z_b^i)|^2, \tag{23}$$

in which $f(t,x,y,z)$ stands for the left-hand side of the 3D heat transfer equation (Equation 14).

The DNN used for the 3D model ~~consists of six hidden layers~~comprises 16 hidden layers, each 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~~

tanh activation function and *Adam* optimizer are employed, similarly to those used for the 1D model. ~~To impose the initial and boundary conditions, we select~~ An exponential decay function, $0.001\times0.9^{i}/10000$, is used to define the learning rate. We selected 2,000 random points to ~~set up~~ define the initial thermal field (~~$t=0$;~~ Figure 12a), ~~whereas~~ and used 2,000 and ~~5~~10,000 ~~randomly sampled points are used to represent~~ points to impose the temperatures at the model base ~~($z=0$)~~ and on the surface topography~~($z=z_s$)~~, respectively (Figure 12b). We ~~set up~~ configured the model with ~~50~~100,000 collocation points (Figure 12c), ~~with half of them~~ half of which are allocated randomly within the orogenic phase ($t\in[0,40]$) and the ~~other half~~ remainder during the post-orogenic phase ($t\in(40,150]$).

~~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 13 shows the performance of the optimization. To verify~~ To evaluate the accuracy of the solution of the PINN, we also computed the thermal field evolution of the Dabie Shan using ~~a~~ the FEM code, ~~i.e.,~~ PECUBE (Braun, 2003). The FEM model is set up on a $120\times120\times31$ ($x\times y\times z$) grid, ~~and the heat transfer is~~ with the thermal field solved at a time step of 1 Myr. The rock uplift history and boundary conditions are ~~set up as the same as~~ identical to those in the PINN model. ~~Figure 14 compares~~

5.1.2 Results

Figure 13 shows the loss curves through the optimization process, which present a progressive decrease until stabilization after approximately 300,000 iterations. Figure 14 presents the temperature profiles on three transects across the model at 40 and 150 ~~Ma (model time)solved by the the~~ Myr computed by the PINN, which reveal significant perturbations of the crustal thermal structures caused by the rock uplift and surface topography, resembling the patterns of the FEM solution. At 40 Myr, at the majority of locations (90%), the temperature differences predicted by the PINN and FEM ~~. On each transect at 40 Ma~~are <10°C. On these transects, the l^2 and l^∞ of the ~~misfits between the temperatures predicted by~~ temperature discrepancies between the solutions of the two methods are ~~in the ranges of 238–324~~240–426 °C and ~~20–40~~26–35 °C, respectively~~, which are calculated at locations of 3100 elements; at~~; these values are calculated from the solutions at the 3131 element locations used in the FEM (Figure 15). At 150 ~~Ma when the~~ Myr when the rock uplift rate has decreased and the topographic relief is lower, the ~~two norms are reduced to 66–92~~ discrepancies between the PINN and FEM solutions are significantly reduced, with the l^2 and l^∞ decreasing to 70–112°C and 8–10 °C, respectively.

~~We also extracted thermal histories for all rock points that ended at~~ Figure 16 shows the misfits in the AHe and AFT ages for the present-day surface ~~and computed apatite fission-track and (U-Th)/He ages . The calculated agesbased on the solutions from~~ rocks at 42,000 locations, which are calculated from their thermal histories predicted by the PINN and FEM. For both thermochronometers, the predicted age misfits at all locations are <20%. At most locations (90%), the misfits between AHe ages predicted by the two methods are <6% and those between AFT ages are <10%. Compared to AHe ages, the higher misfits in predicted AFT ages reflect larger discrepancies in the temperature solutions between the PINN and FEM ~~yield consistent patterns both in the map area (Figure 16) and on the elevation profiles~~ for the periods of high rock uplift rate and significant topographic relief (Figure ~~17). The l^2 and l^∞ norms for the AFT age misfits predicted by~~ 15). At the sites with observed data, the misfits in predicted AHe, AFT, and ZHe ages between 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~~ solutions are <13%, <15%, and ~~12 Ma, respectively~~<2%, respectively, which generally fall within the error margins or clustering ranges of the measured data (Figure 17).

5.2 Inverse problem

~~We test the capability~~

5.2.1 Model setup

To test the effectiveness of PINNs in ~~solving~~ constraining the exhumation models using real data, we formulate an inverse problem~~with real data. To define the problem, we set the onset time and rate of the~~, aiming to estimate the post-orogenic uplift and topographic evolution of the Dabie Shan using published low-temperature thermochronological ages. In our analysis, we seek to constrain the orogenic rock uplift rate (u_0), the ending time of the rapid uplift (t_1 ~~and~~), and the rock uplift rate during the post-orogenic period (u_1), as described in Equation 21. In addition, we estimate the amplification factor for the topography during the orogenic phase (w_0) and the ~~initial topographic amplification (w_0~~ onset time for topographic decay (t_d) as described in Equation 19~~) as unknown parameters. The three parametersare sampled in the ranges of 0–100 Ma, 0–0.1~~. We confine the search for these five parameters, u_0, u_1, w_0 , t_1, and t_d within the ranges 0–1 km/Myr, 0–1 km/Myr, ~~and~~ 0–6, 0–50 Myr, and 50–100 Myr, 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~~ guide the optimization process, thermochronological age data reported by Reiners et al. (2003)(Figure 10b) are used~~to guide the optimization; only~~, considering only the ages <150 Ma~~are used.~~.

Similar to the ~~inversion of the~~ 1D model ~~,~~ inversion, we use both the *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 generation of DE contains 50 iterations of *Adam*. A total of 320,~~ DIRECT algorithms to optimize the DNN model and explore the search space. For the *Adam* iterations, we set the learning rate using the same exponential decay function applied in the forward 3D model. In the DIRECT search, we use the “restart-and-zoom-in” strategy akin to that employed in the forward 3d model, but increase the number of cycles to 40. Each DIRECT cycle includes up to 30 iterations and a maximum of 1,000 ~~iterations were run~~function evaluations. Altogether, the inversion process comprises >500,000 *Adam* iterations and >1,000 DIRECT iterations, evaluating >40,000 age misfit functions.

As a ~~comparison~~reference, we also ~~searched~~ explore the multi-dimensional ~~space defined by the unknown parameters~~ parameter space using the Neighborhood Algorithm (NA) (Sambridge, 1999a), in which the forward model is ~~solved~~ computed using PECUBE (Braun, 2003). ~~The NA partitioned~~ It is worth noting that compared to the modeling employed by Braun and Robert (2005) in the same region, we do not consider the isostatic rebound in response to the exhumation and thus the result may differ slightly. To reduce the computational demands compared to the model used in the forward problem (Section 5.1), we reduce the FEM grid size to 80×80×31 by increasing the horizontal spacing between elements. The NA divides 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~~ cell centers. The subsequent iteration generates new samples within a subset of cells ~~which contain the best-fitting~~ containing the best-fit models from the previous ~~step~~iteration. This process is repeated ~~many times in order~~ to find the optimal ~~values of the unknown parameters that can reproduce~~ parameter values that fit the observed data. Here ~~our NA search contains a total of 26 iterations, with 300 forward models in the first iteration and~~ the NA search for parameters u_0, u_1, w_0 , t_1, and t_d are within the same domains as predefined for the inversion of the PINN model. The search comprises 150 iterations with each running forward models in 60 ~~in each of the rest. After evaluating all models in each step~~cells, out of ~~the 60 cells we resample 50 with the best-fitting models to generate new combinations of parameter values~~ which the 75% best cells are resampled for the next iteration. In total, the NA inversion includes 9,060 forward PECUBE runs.

~~Performance of the optimization of the~~

5.2.2 Results

The training process of the PINN is shown in Figure 18 ~~, and sampled parameters~~ as evolution of the thermal model loss and the misfits in predicted ages. The parameter values evaluated by the DIRECT search are shown in Figure 19. ~~It 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~~ Notably, the loss function curve displays more fluctuation compared to the forward model, but it maintains a general decreasing trend. This fluctuation reduces as the learning rate decreases (Figure ~~19a-b), respectively, whereas that of w_0 to >4 after 125 generations of DE (~~18a). Upon the stabilization of the loss function after approximately 200,000 ~~forward models; Figure 19c). These ranges are in general consistency with the results of the NA search~~*Adam* and 500 DIRECT iterations, the predicted minimum age misfits have a declining pattern for around 250 DIRECT iterations and show minimal changes thereafter (Figure 18b). During the parameter search, convergence mostly occurs within the initial 100 and then from 500 to 1,000 DIRECT iterations (Figure 19). ~~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 sampled values, which will be a direction for the future development (Section 6.2)~~Finally, estimates for the parameters u_0, u_1, w_0 , t_1, and t_d are confined to the ranges of 0.7–0.95 km/Myr, 0–0.18 km/Myr, 2–5.5, 21–45 Myr, and 74–91 Myr, respectively.

The ~~results of our inverse modeling are in general consistency with t's conclusions based on inversion of the same dataset, which suggested that during the post-orogenic evolution since the Late Cretaceous,~~ NA search based on forward PECUBE models exhibits patterns comparable to those observed in the optimization of the PINN model. In specific, the ~~Dabie Shan experienced slow uplift at~~ search for u_0 also converges rapidly to a range of 0–0.1 km/Myr, while the other four sampled parameters show significant convergence only after approximately 80 iterations. In the NA search the parameters u_0, u_1, w_0, t_1, and t_d are ultimately confined within the ranges of 0.8–1 km/Myr, 0.01–0.~~04~~.13 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 's 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~~, 4.1–5.1, 33–44 Myr, and 88–97 Myr, respectively (Figure 20). Except that the parameter w_0 has been confined to a narrower range by the NA, the search results by the two inverse models overlap over significant ranges.

6 Discussion

6.1 ~~Advantage of PINNs~~Comparison between the PINN and numerical method

~~We have shown that PINNs can be applied~~ We have demonstrated that in the study of thermochronology, ~~to solve heat transfer in the crust for providing 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 dataassimilation methods which may require expansive computations 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 .~~ PINNs are comparably effective as conventional numerical methods. In both the 1D and 3D forward models we presented, the mean temperature difference across all points at any given time between the PINN and FEM solutions is <0.5°C, with the maximum difference at specific points ranging between 10 and 30°C, which appear to occur during periods of rapid rock uplift and exhumation (Figures 5b and 15). This level of discrepancy is minor compared to the typical uncertainties in thermal histories derived from thermochronological data. For 90% surface samples in our examples, temperature discrepancies between the PINN and FEM solutions resulted in <10% misfits in the predicted ages. Larger misfits up to 20% were found in cooling ages of samples that record rapid rock exhumation in areas with high topographic relief (Figures 10b and 17b). These misfits in age estimates are at comparable magnitude to inherent uncertainties in measured age data, and can be less significant than the data dispersion caused by the variability in the kinetics of noble gas diffusion or fission-track annealing.

6.2 ~~Current limitations and future directions~~

For simple forward models, conventional numerical methods do not require optimization, thus having an advantage in the computational efficiency over PINNs. For instance, solving the 1D forward model in Section 4.1 using the FEM required a few milliseconds on a modern CPU (Intel Xeon Gold 5420+), whereas training the PINN model with 80 neurons for 50,000 iterations on a GPU (NVIDIA RTX 6000 Ada) took several hundred seconds, which is five orders of magnitude longer than the time required for the FEM solution. This contrast diminishes as the model's dimensionality, size, and complexity increase. To illustrate, simulating the 3D forward model of the Dabie Shan in Section 5.1 on a $120\times120\times31$ grid over 150 time steps took approximately 2,400 seconds (s) using PECUBE, whereas training the PINN with 320 neurons over 400,000 iterations required around 32,000 s, merely one order of magnitude longer than the FEM solution.

~~Our examples show that, at least for a relatively simple time-space domain,a basic configuration of~~ For inverse problems, the computational time escalates as the parameter space expands, which necessitates the evaluation of a larger number of numerical models. In Section 4.2, the MCMC sampling which required the ~~optimization (e.g., *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~~evaluation of 1,000,000 FEM models, consumed approximately 8,000 s. This duration represents an increase by six orders of magnitude from a single forward solution. In contrast, the inverse analysis using DIRECT and the PINN model, which evaluated over 15,000 age misfit functions, was completed within 7,000 s, marking an increase in time by only one order of magnitude from the PINN forward model solution. For the 3D Dabie Shan model (Section 5.2), the NA inversion, using 30 CPU cores and coupled with a PECUBE model on an 80×80×31 grid, required approximately 300,000 s of clock time to explore the search space and evaluate 9,060 misfit functions. The cumulative computation time across all processors amounted to ∼9,000,000 s which is four orders of magnitude greater than the time used for solving a single forward model. Conversely, the DIRECT search and the PINN model optimization evaluated 38,130 age misfit functions within 560,000 s, which is increased by one order of magnitude from the 3D forward model.

Although the computational performance metrics presented herein are not derived from rigorous benchmarks, they provide a preliminary comparison of the efficiency of PINNs and traditional numerical methods. It is evident that for inverse problems, PINNs have an advantage due to their seamless integration of parameter search and forward model optimization within one single training process. This integration prevents PINNs from an exponential increase in computational time as the dimensionality of the inverse problem increases, illustrating the method's potential for inverse analysis of expansive models with complex boundary conditions, which can be unattainable with traditional numerical methods. ~~On the one hand, it is an ongoing effort to improve the algorithm and setup for efficient optimization of PINNs . 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 .~~

~~Our demonstrations have only considered the uplift history as space-independent rates~~

6.2 Directions for future development

In our applications, we have modeled the rock uplift as a vector in the vertical direction ~~.~~ independent of spatial locations. However, thermo-tectonic research often requires considering rock uplift and exhumation rates that vary both temporally and spatially. Previous work (e.g., Cai et al., 2021b) has ~~shown that implementation of a full~~ demonstrated the feasibility of incorporating a comprehensive 3D velocity field in the ~~neural networks is possible~~DNN. However, ~~in cases with sharp~~ the effectiveness of PINNs in scenarios with abrupt changes in the velocity ~~fields, the capability and efficiency of PINNs to solve PDEs remain to be tested or improved~~ field requires further exploration or improvement (e.g., Rasht-Behesht et al., 2022). This ~~is crucial for tectonic studies in which the~~ capability is vital for studying tectonically active regions, where fault activities can significantly disrupt exhumation and geothermal patterns~~have been perturbed by fault activities, and therefore should~~ , and thus will be a focus of our future development of the method.

Our current ~~solution of~~ approach to the inverse problem ~~includes a final~~ yields an optimized solution of the heat ~~model with values of unknown parameters~~transfer model. Although the ~~direct search process (i.e., DE) results in an ensemble~~ search process results in a collection 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~~ associated with the optimized model. To address this, a Bayesian neural network ~~, where~~ can be employed in the PINN framework. This way, the parameters in the neural network would follow Gaussian distributions ~~, instead of a~~ in contrast to the deterministic parameters of a traditional fully-connected network ~~where the parameters are all deterministic in the PINNs framework~~ (Yang et al., 2021). ~~By doing this,~~ This modification would allow the Bayesian PINNs ~~can adapt to~~ to accommodate noisy data and provide ~~estimates of the uncertainties~~ uncertainty estimates for the inferred parameters in the PDEs.

We have outlined the basic configurations for optimizing PINNs and their integration with the DIRECT search algorithm for inverse analysis. However, we expect that model domains of greater complexity, which require a denser array of collocation points, may need more fine tuning of the training strategies. Enhancing the algorithm and its configuration for the efficient optimization of PINNs is an active field of research (Jagtap et al., 2020; Shukla et al., 2021). For the inverse analysis, the flexibility in problem formulation with PINNs facilitates the incorporation of various search and optimization algorithms. Our ongoing efforts include integrating and testing these algorithms, to ensure stable training outcomes from PINNs.

7 Conclusions

We have presented applications of physics-informed neural networks (PINNs) in solving crustal heat transfer problems for thermochronology. By harnessing the computational power of deep learning, PINNs offer a flexible and efficient alternative to traditional analytical and numerical techniques. The method allows a straightforward integration of initial and boundary conditions as well as observed data, and can be used together with gradient-based and derivative-free optimization methods to solve both forward and inverse problems. By applying PINNs to a 1D synthetic model and a 3D natural model of the Dabie Shan under varying tectonic and topographic conditions, we have demonstrated the potential of this approach for estimating the thermal and exhumation histories of the crust during tectonic and landscape evolution. Future work will focus on improving the optimization strategies for inverse models, quantifying the model uncertainties, and expanding the application of PINNs to more intricate geomorphic, geologic, and geodynamic scenarios.

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 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

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

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/10.1130/0016-7606(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: Application to the Dabie Shan, China, Earth Surface Processes and Landforms, 30, 1203–1225, https://doi.org/10.1002/esp.1271, 2005.

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.

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.

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. Thermokinematic model of exhumation, erosion, and thermochronometer interpretation, Journal of Geophysical Research: Solid Earth, 108, 2173, https://doi.org/10.1029/2001JB001723, 2003.

Fitzgerald, P. G. and Malusà, M. G.: Concept of the exhumed partial annealing (retention) zone and age-elevation profiles in thermochronology, 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, 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.

Goodman, J. and Weare, J.: Ensemble samplers with affine invariance, Communications in Applied Mathematics and Computational Science, 5, 65–80, https://doi.org/10.2140/camcos.2010.5.65, 2010.

Hacker, B. R., Ratschbacher, L., Webb, L., and Shuwen, D.: What brought them up? Exhumation of the Dabie Shan ultrahigh-pressure rocks, Geology, 23, 743, https://doi.org/10.1130/0091-7613(1995)023<0743:WBTUEO>2.3.CO;2, 1995.

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., 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.

Jones, D. R. and Martins, J. R. R. A.: The DIRECT algorithm: 25 years Later, Journal of Global Optimization, 79, 521–566, https://doi.org/10.1007/s10898-020-00952-6, 2021.

Jones, D. R., Perttunen, C. D., and Stuckman, B. E.: Lipschitzian optimization without the Lipschitz constant, Journal of Optimization Theory and Applications, 79, 157–181, https://doi.org/10.1007/BF00941892, 1993.

Ketcham, R. A.: Forward and inverse modeling of low-temperature thermochronometry data, Reviews in Mineralogy and Geochemistry, 58, 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. L.: Adam: A method for stochastic optimization, in: ICLR 2015, San Diego, https://doi.org/10.48550/arXiv.1412.6980, 2015.

Lagaris, I., Likas, A., and Fotiadis, D.: Artificial neural networks for solving ordinary and partial differential equations, IEEE Transactions 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.

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.

Nie, S., Yin, A., Rowley, D. B., and Jin, Y.: Exhumation of the Dabie Shan ultra-high-pressure rocks and accumulation of the Songpan-Ganzi flysch sequence, central China, Geology, 22, 999, https://doi.org/10.1130/0091-7613(1994)022<0999:EOTDSU>2.3.CO;2, 1994.

Raissi, M., Perdikaris, P., and Karniadakis, G. E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, 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.

Ratschbacher, L., Franz, L., Enkelmann, E., Jonckheere, R., Pörschke, A., Hacker, B. R., Dong, S., and Zhang, Y.: The Sino-Korean–Yangtze suture, the Huwan detachment, and the Paleozoic–Tertiary exhumation of (ultra)high-pressure rocks along the Tongbai-Xinxian-Dabie Mountains, in: Ultrahigh-pressure metamorphism: Deep continental subduction, Geological Society of America, https://doi.org/10.1130/2006.2403(03), 2006.

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.

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.

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.

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., 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.

Whipp, D.: HUGG/TC1D: v0.1, https://doi.org/10.5281/ZENODO.7124272, 2022.

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 with noisy data, Journal of Computational Physics, 425, 109 913, 2021.

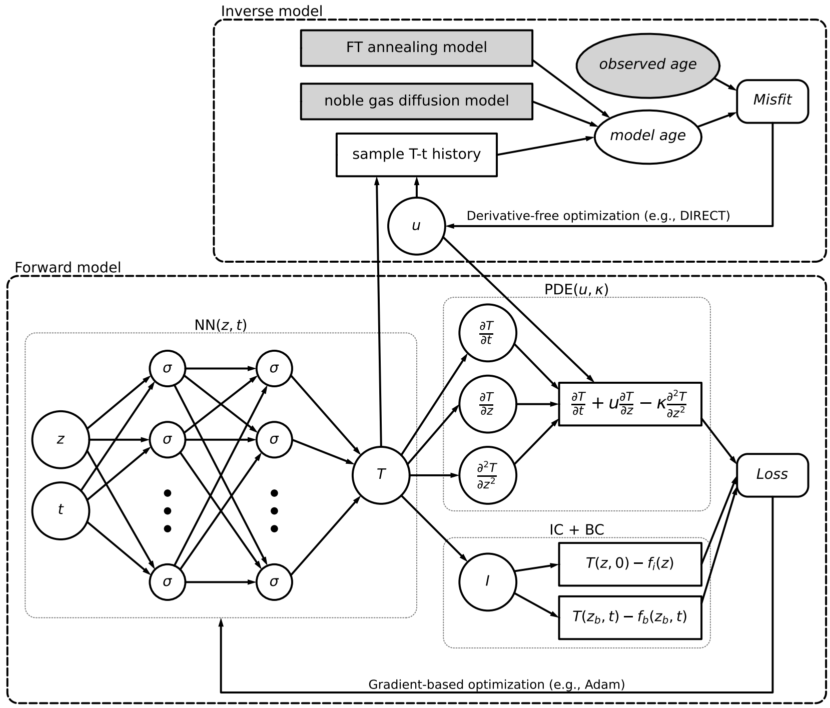

Figure 1. Work flow for forward and inverse modeling of 1D thermal profile for thermochronology. f_i and f_b are functions for the initial and boundary conditions of the model, respectively.

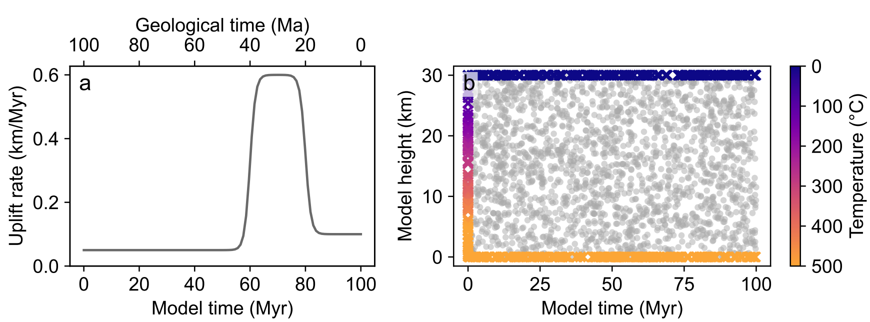

Figure 2. Setup of the 1D thermal model of the crust. (a) ~~Uplift~~ Rock uplift function of the model. (b) Collocation (gray dots) and boundary condition points (colored cross) ~~for~~ of the PINN model.

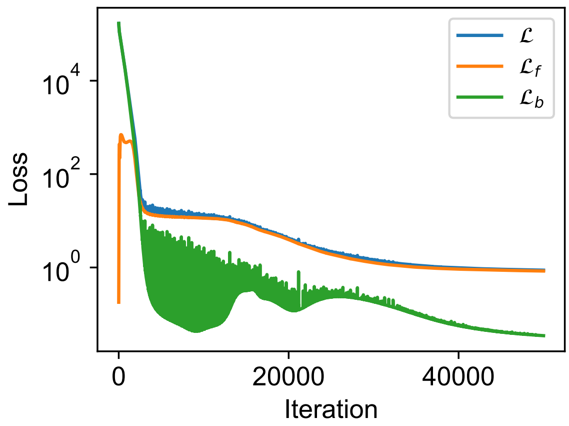

Figure 3. Loss curves ~~for~~ of the 1D forward model. $\mathcal{L}_f$ and $\mathcal{L}_b$ are losses from the heat-transfer equation and the boundary conditions, respectively, whereas $\mathcal{L}$ is the total loss.

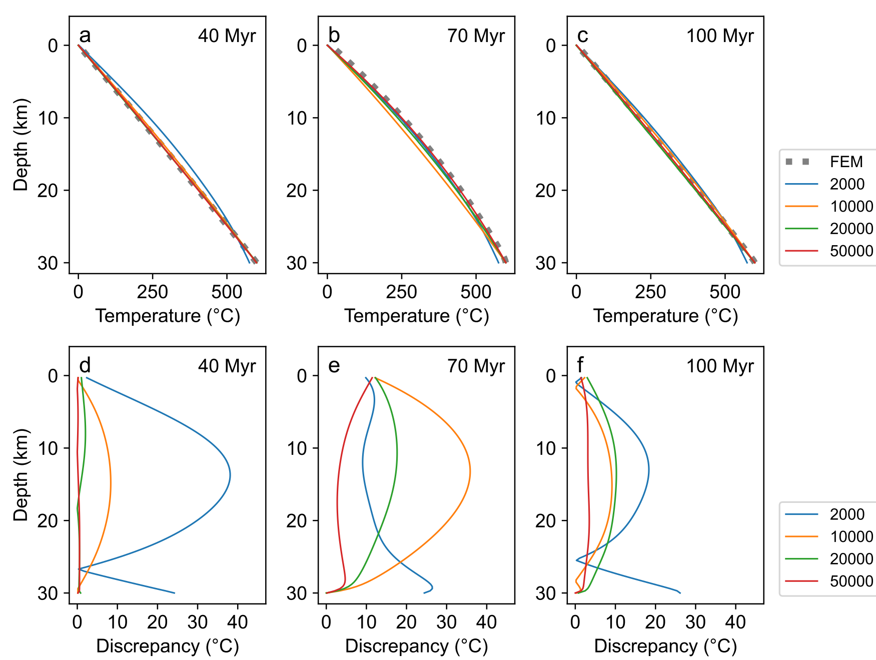

Figure 4. Predicted thermal profiles of the crust at different model times. Colored lines depict ~~solutions using~~ the ~~PINNs~~ PINN solutions after ~~1000,~~ 2000, ~~4000~~10000, 20000, and ~~8000~~ 50000 training iterations of training, respectively. Dashed line ~~represents~~ indicates the solution using the ~~finite-element method (~~FEM ~~)~~solution.

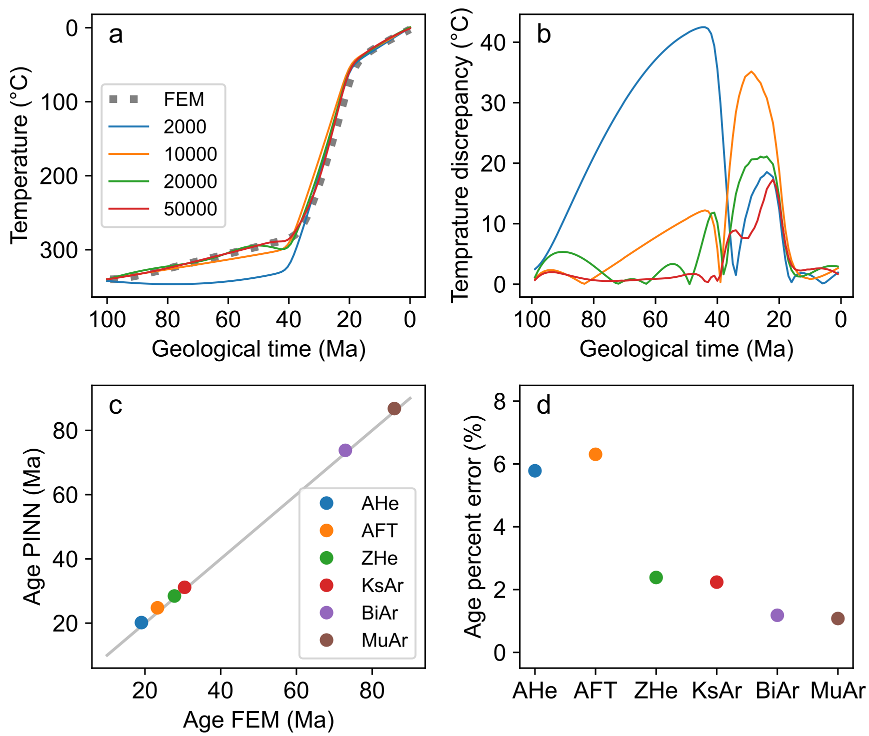

Figure 5. Comparison of the thermal histories and ages computed from the PINN and FEM solutions of the forward model. (a) Predicted ~~temperature-time~~ time-temperature paths and cooling ages for rock sample ended at the surface. ~~(a) Predicted thermal history.~~ Colored lines represent ~~predictions by~~ the ~~PINNs model~~ PINN predictions after ~~1000,~~ 2000, ~~4000~~10000, 20000, and ~~8000~~ 50000 iterations~~of training~~, respectively. Dashed line ~~represents prediction using~~ indicates the FEM solution. (b) ~~Predicted cooling ages~~Temperature discrepancy between the PINN and the FEM solutions. Line color depicts the iteration numbers performed by the PINN, ~~where~~ as shown in the legend of (a). (c) Thermochronological ages computed from the thermal models solved by the PINN and FEM. (d) Percent error between the ages predicted by the PINN and FEM. 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.

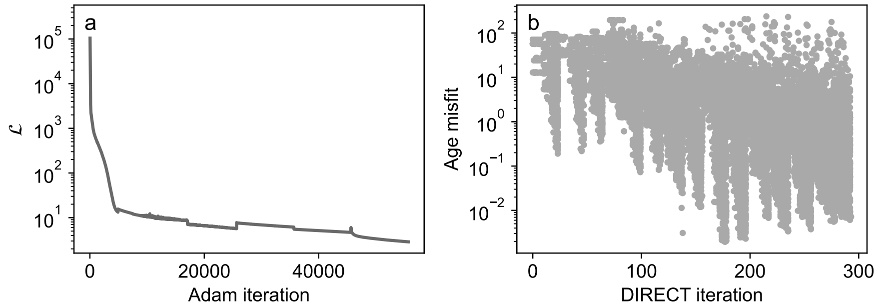

Figure 6. ~~Loss and misfit curves for~~ Optimization process of the 1D inverse model. ~~For clear display, the curves are plotted using a randomly sampled subset (1%) of all iterations. (~~(a) Loss ~~curves for different configuration~~ function curve of the optimization. (b) Age ~~misfit curves. n_g and n_p are generation number and population size of $differential\ evolution$ (DE), respectively, and n_a is~~ misfits evaluated by the ~~*Adam* iteration number in each generation of DE~~DIRECT search.

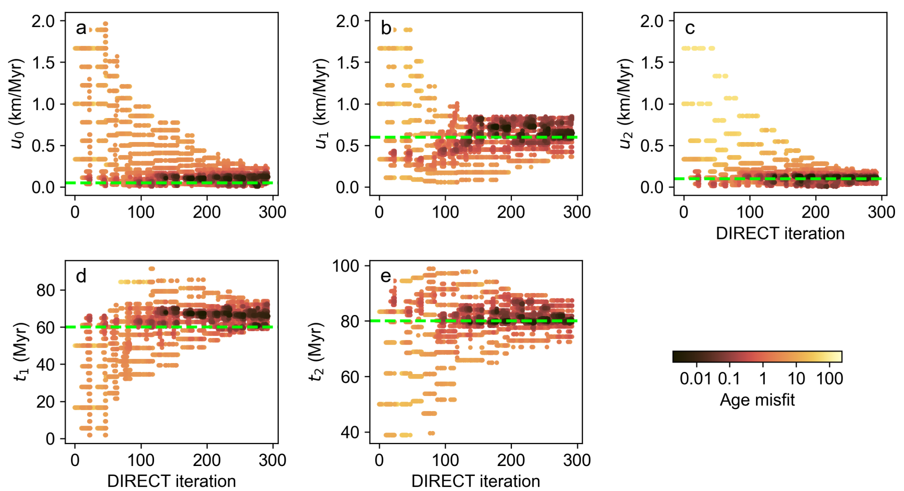

Figure 7. ~~A random subset (1%) of all sampled parameter~~ Parameter values sampled by the DIRECT method coupled with the PINN model for the 1D inverse model. Dots are ~~color coded~~ color-coded according to ~~predicted~~ evaluated function value (age ~~misfits~~misfit). Dashed horizontal lines indicate the "true" parameter values.

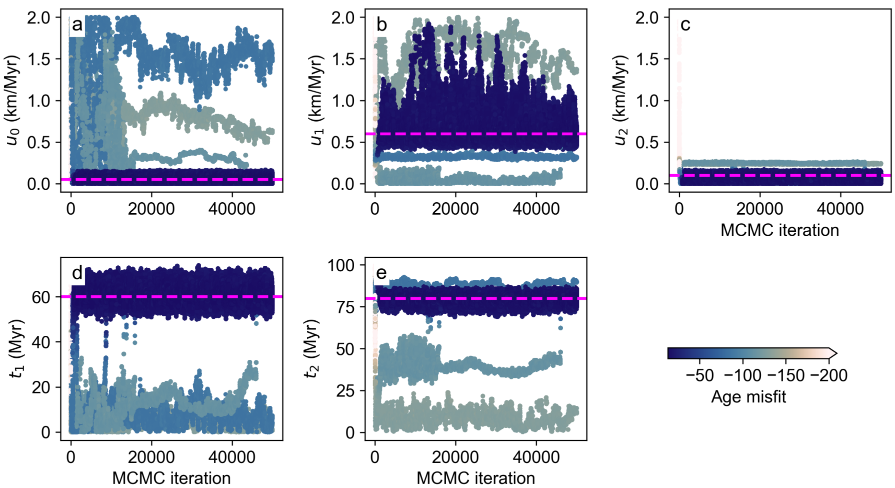

Figure 8. ~~Prediction of~~ Parameter values sampled by the ~~optimized~~ MCMC method coupled with FEM model for the 1D ~~thermal~~ inverse model. ~~(a) Uplift rate histories predicted by~~ Dots are color-coded according to the ~~three inverse~~ log-likelihood of the forward models. Dashed ~~line represents~~ horizontal lines indicate the “true~~uplift model~~” parameter values.~~(b) Predicted thermal histories. Dashed line represents solution with FEM. (c) Thermochronological ages predicted by PINNs versus synthetic data.~~

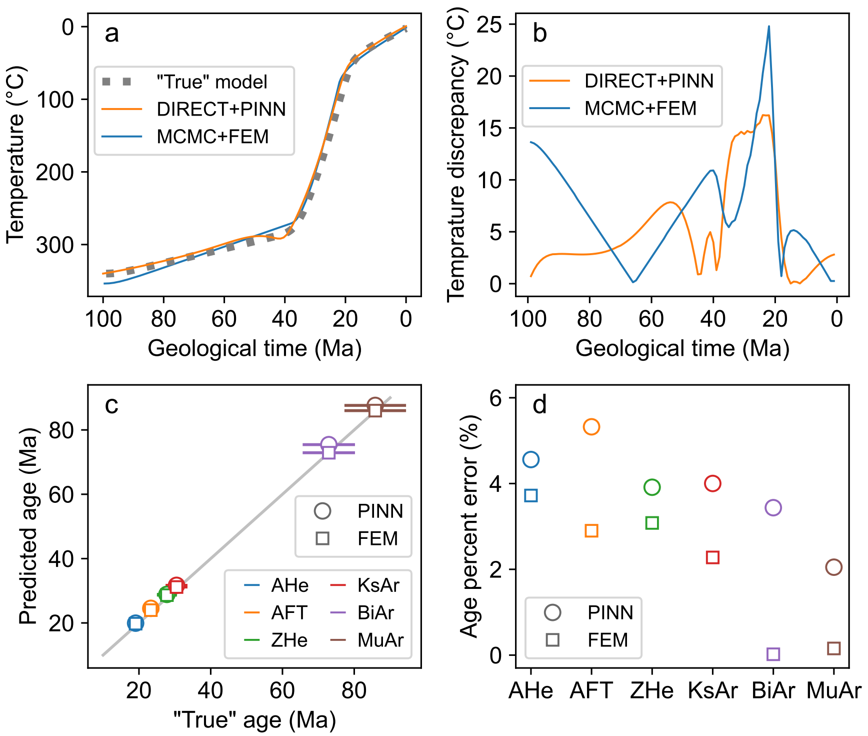

Figure 9. Thermal histories and ages predicted by the optimized PINN model and the "best-fit" FEM model from the MCMC sampling chain. (a) Predicted time-temperature paths and cooling ages for a surface rock sample at the end of the model history. (b) Temperature discrepancy between the predicted time-temperature paths and the "true" model. (c) Thermochronological ages calculated from the PINN solution and the "best-fit" FEM model, compared to the synthetic data from the "true" thermal history. (d) Percent error between the predicted ages and synthetic data.

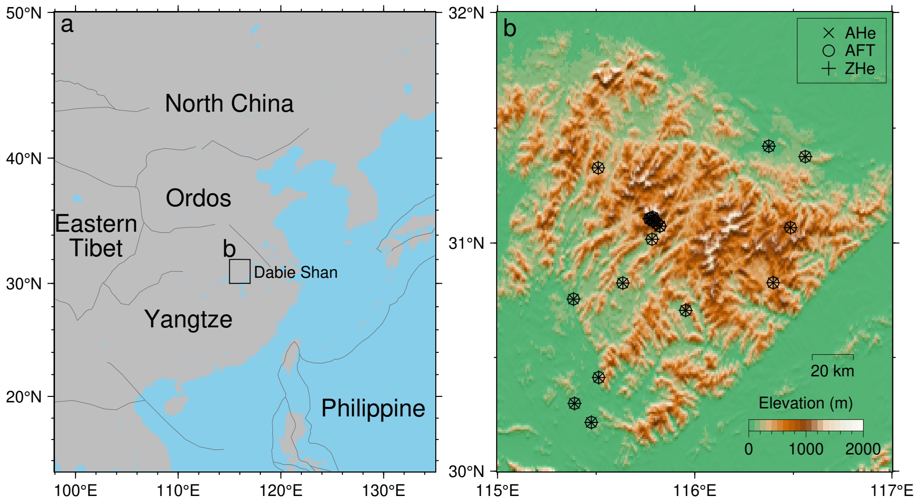

Figure 10. (a) Location of the Dabie Shan, eastern China. (b) Topography of the Dabie Shan and locations of thermochronological data.

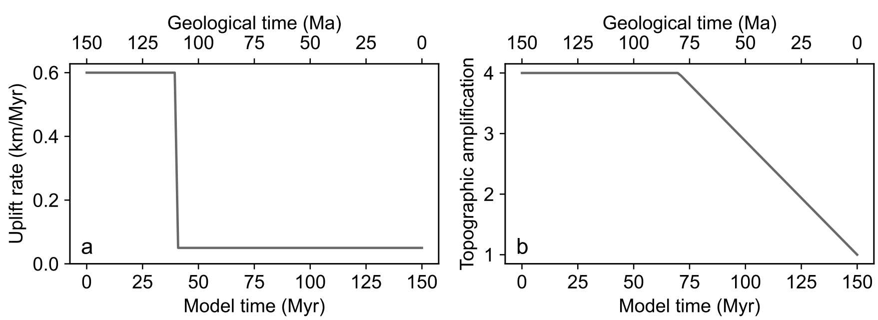

Figure 11. Inputs of the 3D thermal evolution model of the Dabie Shan. (a) ~~Uplift~~ Rock uplift model. (b) Topographic amplification model.

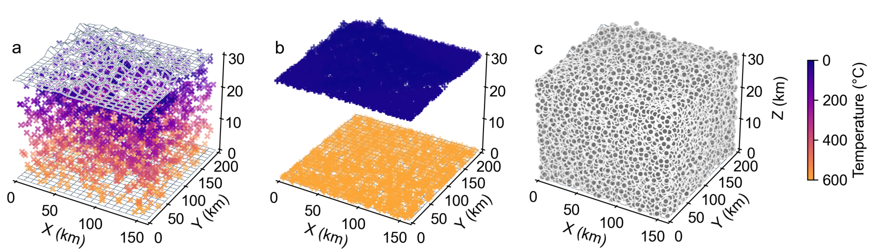

Figure 12. The initial and boundary conditions and collocation points for the 3D thermal model of the Dabie Shan. Control points are projected onto the $\{x, y, z\}$ space. (a) The initial thermal field. (b) Boundary temperatures at the surface and base of the model. (c) Collocation points.

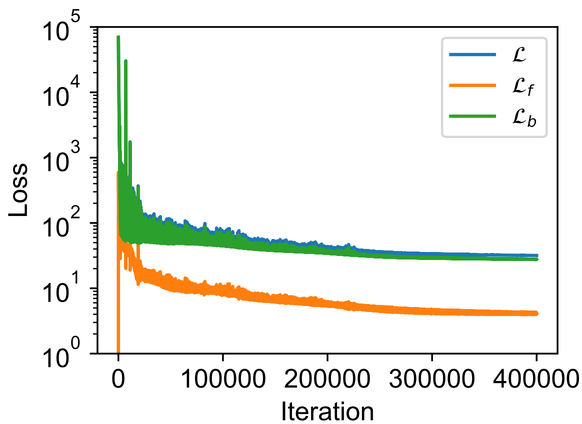

Figure 13. Loss curves for the 3D forward model.

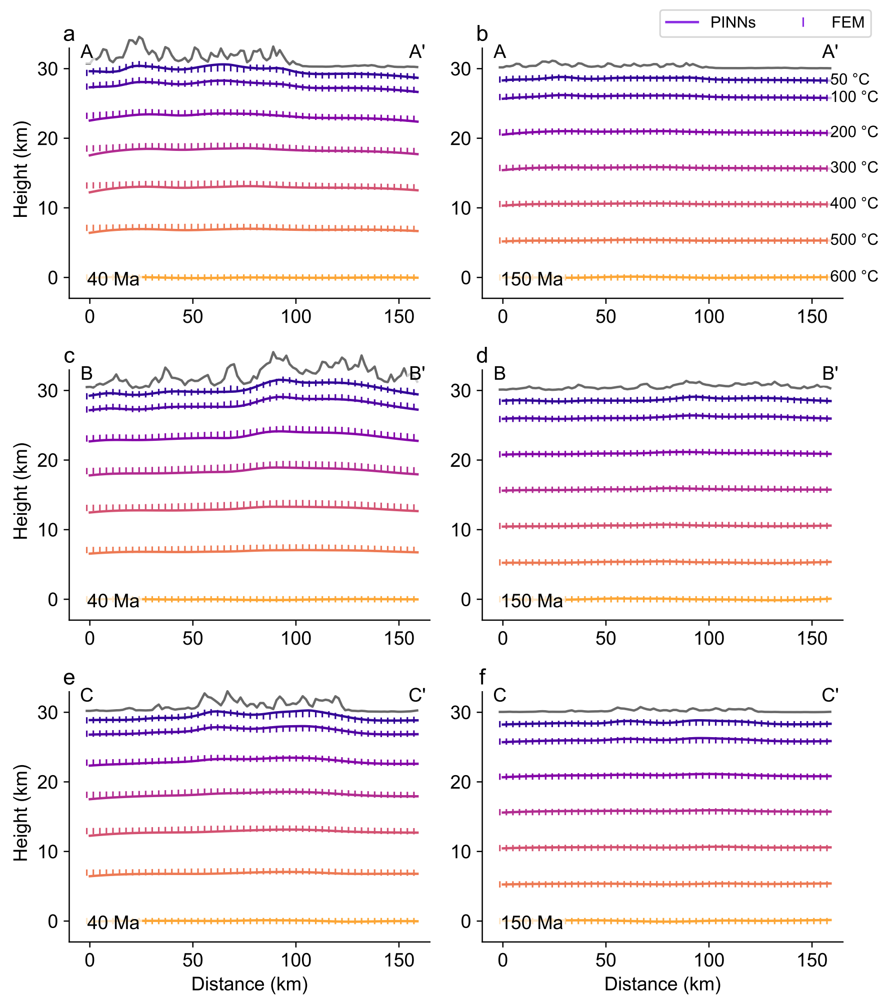

Figure 14. Predicted isotherms on three transects across the Dabie Shan at model times of 40 and 150 ~~Ma~~Myr. Solutions using the PINN and FEM are compared. Locations of the transects are shown in Figure 16.

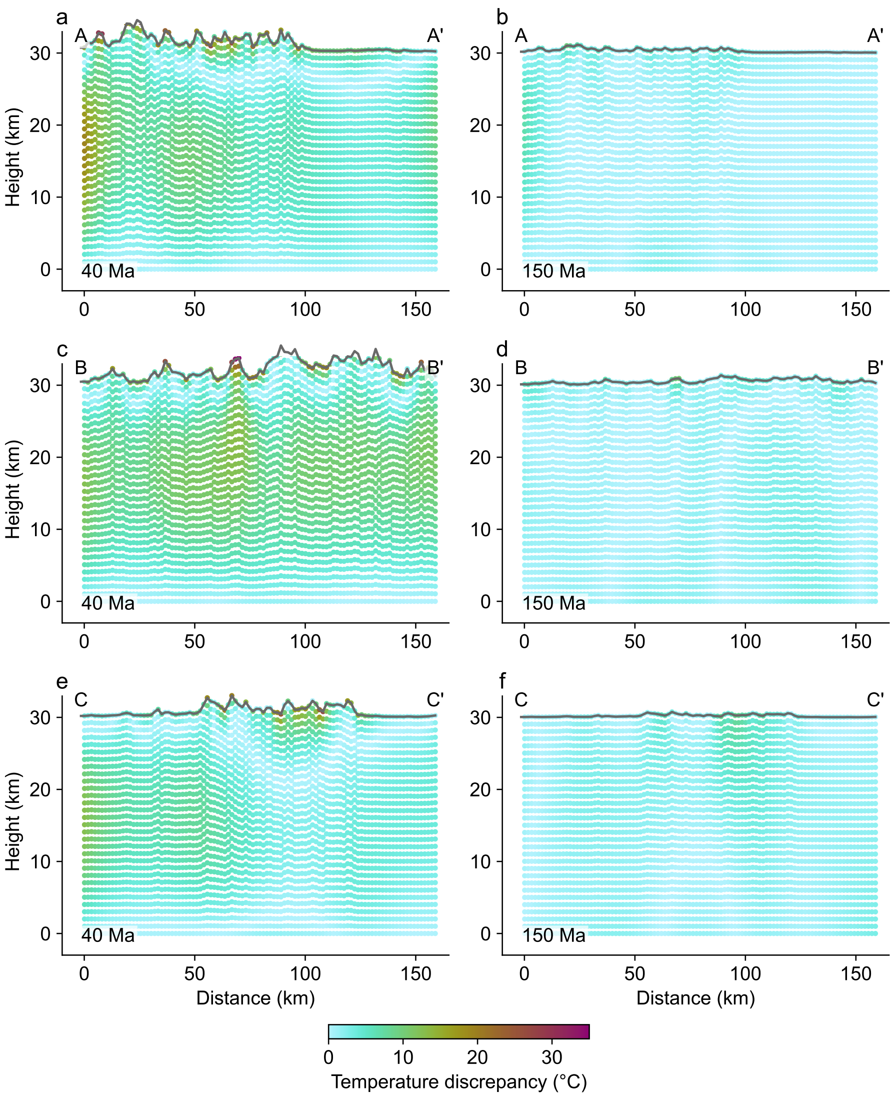

Figure 15. ~~Predicted age maps of~~ Discrepancies in temperature solutions between the PINN and FEM on three transects across the Dabie Shan ~~using the FEM~~ at model times of 40 and ~~PINN~~150 Myr. ~~A-A’, B-B’, and C-C’ indicate locations~~ Locations of the transects are shown in Figure ~~14~~16.

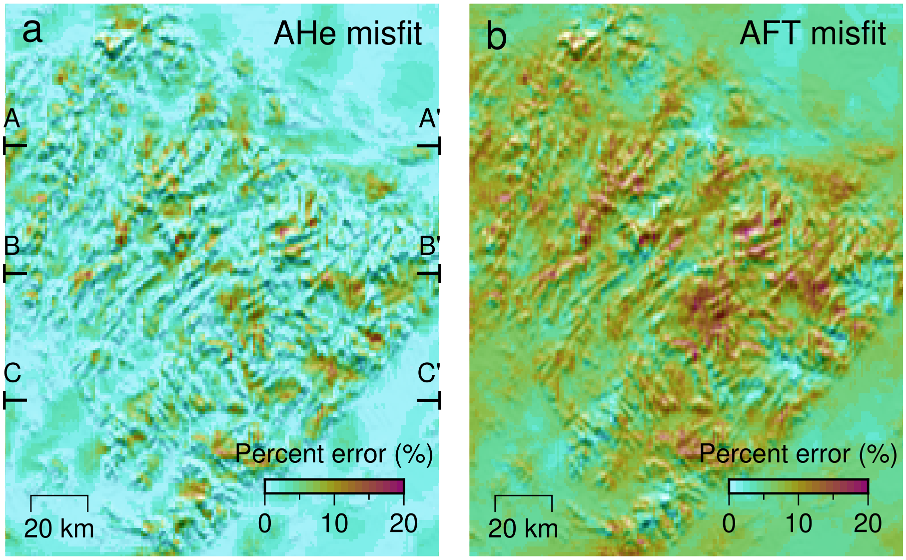

Figure 16. Maps of the Dabie Shan showing the predicted age misfits between the PINN and FEM solutions. A-A', B-B', and C-C' indicate locations of the transects in Figure 14.

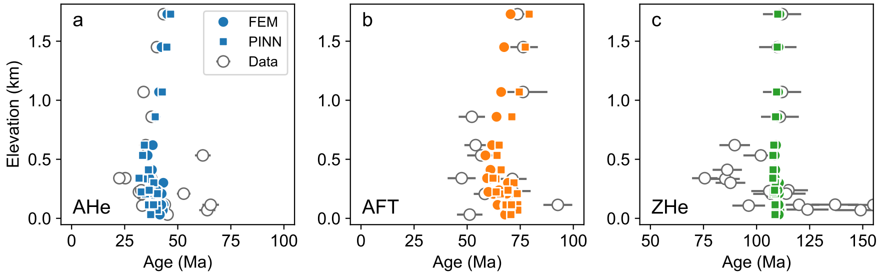

Figure 17. Predicted and observed thermochronological ages versus sample elevations.

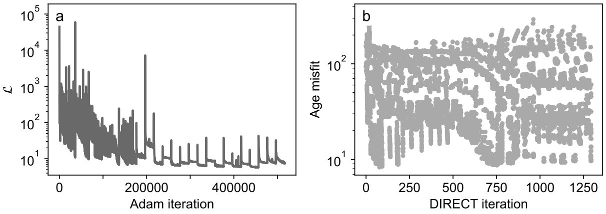

Figure 18. ~~Loss and age misfit curves~~ Optimization process of the 3D inverse model of the Dabie Shan. (a) Loss function curve of the optimization. (b) Age misfits evaluated by the DIRECT search.

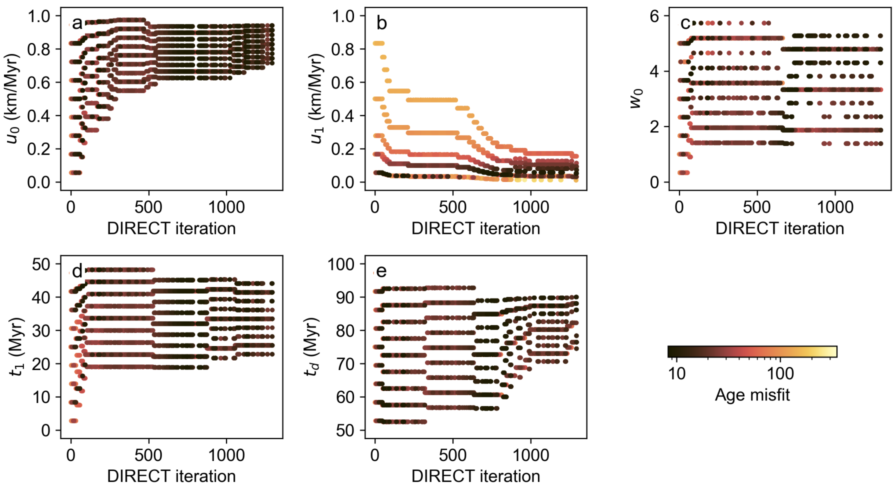

Figure 19. ~~Sampled~~ Evaluated parameter values for the 3D inverse model of the Dabie Shan using ~~(a-e)~~ the ~~PINN coupled with the differential evolution~~ DIRECT algorithm (Jones et al., 1993) and ~~(d-f)~~ the ~~PECUBE coupled with the Neighborhood Algorithm~~ PINN model.~~The results of PINNs shown here are a random subset (1%) of all iterations.~~

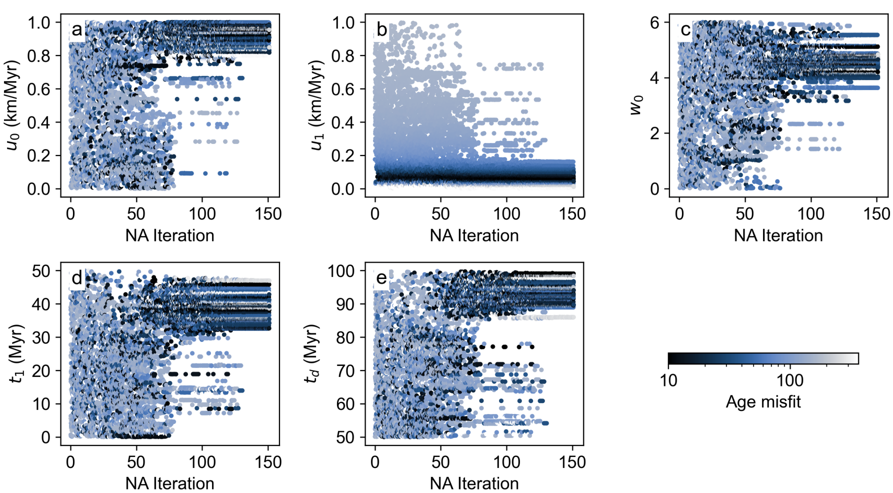

Figure 20. Evaluated parameter values for the 3D inverse model of the Dabie Shan using the Neighborhood Algorithm (Sambridge, 1999a) coupled with the PECBUE model.