Proceedings World Geothermal Congress 2026
Calgary, Canada, June 8 – 11, 2026

3D Modeling of Terrestrial Temperature and Geothermal Resource Assessment throughout China Using Physics-informed Neural Network

Wei Runqian[1]*, Katsuaki Koike[1] and Taiki Kubo[1]
Department of Urban Management, Graduate School of Engineering, Kyoto University[1]*
wei.runqian.25m@st.kyoto-u.ac.jp

Keywords: temperature-at-depth, heat flow, Curie point depth, surface temperature

ABSTRACT

In the background of the global energy transition moving from conventional sources to renewables, China is increasingly focusing on geothermal energy. This shift is motivated by geothermal potential to supply a significant amount of energy, offering advantages for both environmental sustainability and national energy independence by replacing fossil fuels. Accurate estimation of temperature-at-depth, particularly via data-driven methods, necessitates high-quality borehole temperature data. However, the availability of such accessible datasets is severely limited in China. To address this scarcity, this study tries to map the spatial distribution of subsurface temperature throughout China, employing a physics-informed neural network (PINN) that integrates multiple crustal properties to compensate for the sparse nature of well-temperature logging data. The PINN is governed by the heat conduction equation and Fourier's law. Prediction of temperature-at-depth is performed using an ensemble learning approach, combining Random Forest, XGBoost, and LightGBM, with the surface temperature and Curie point depth for the boundary conditions. The resultant temperature distribution model generates a predicted temperature map for the target depth, which represents a rational value range and accurately depicts the distribution in specific areas. Nevertheless, the overall distribution remains significantly affected by the surface temperature, suggesting a need to incorporate more reasonable constraints derived from subsurface properties in the next step.

1 INTRODUCTION

Geothermal energy serves various specific applications, such as electricity generation, agriculture, and aquafarming. Given the rising global need for renewable power, it is recognized as a highly promising energy source. Although the use of geothermal heat has been widespread in China, geothermal electricity generation remains underdeveloped. This limitation stems primarily from a lack of detailed modeling concerning temperature-at-depth, heat source distribution, and available energy potential. As outlined in the 13th and 14th Five-Year Plans, China is actively promoting the use of geothermal energy, particularly focusing on mid-to-deep geothermal resources and hot dry rock formations, as an alternative, renewable power source.

Well-temperature logging data provides the most direct and crucial input for estimating target temperatures within a data-driven machine learning framework. Previous research by Ieki et al. (2025) demonstrated a reasonable estimation of temperature-at-depth throughout Japan using borehole temperature data as input for a deep neural network. However, data points containing this critical information are scarce in China. Consequently, physical constraints are incorporated to enable the prediction of temperature-at-depth, and the effectiveness of this implementation is validated in the present study.

2 METHODS

2.1 Network Modeling

Physics-informed neural networks (PINNs) are machine learning methods that incorporate physical properties, specifically partial differential equations (PDEs) and boundary conditions, into the learning process of the neural network (Ishitsuka and Lin, 2023), as depicted in Figure 1. In this study, PINN is used as the metamodel in a stacking machine learning process, which is the most important for the estimation.

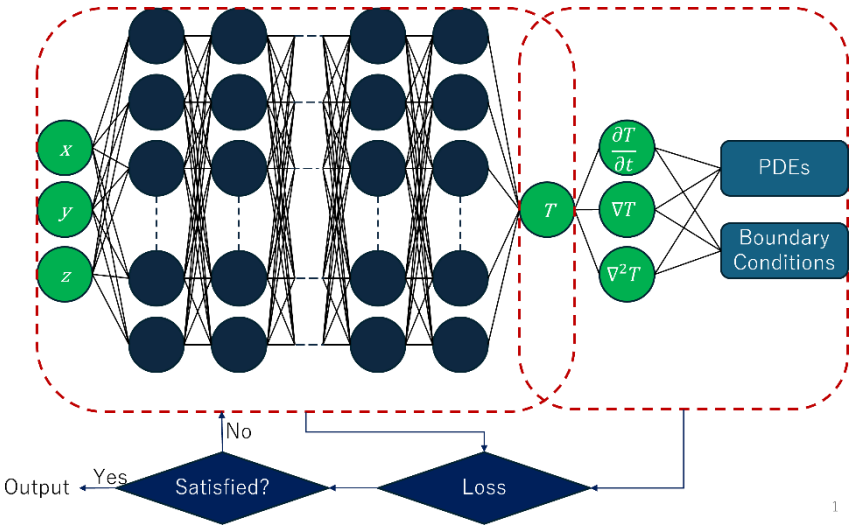


Figure 1. Loss calculation process in PINN.

2.1.1 Stacking Method

The stacking method is a machine learning paradigm that involves combining the predictions from several base models and then feeding these integrated outputs into a final model (the metamodel) to generate the ultimate prediction. Specifically, the base models' outputs serve as inputs for the metamodel to predict the target value. This technique is highly regarded for its ability to enhance the generality and robustness of machine learning models while concurrently mitigating overfitting. For this study, Random Forest, XGBoost, and LightGBM were chosen as the base models. A PINN, structured with a three-layered fully connected network, is employed as the metamodel within the stacking framework.

2.1.2 Physics-informed Neural Network

While a normal neural network approximates the mathematical relation between input and output, a PINN integrates partial differential equations and boundary conditions into loss calculation (Ishituska and Lin, 2023). Thus, physical information can be utilized as prior information. In PINN, the model is optimized to satisfy both observed data and physical information rather than only observed data, as implemented in a normal neural network. The loss calculation process and function are given by the following equations.

$$LOSS_{total} = w_{NN} * LOSS_{NN} + w_{PDE} * LOSS_{PDE} + w_{BC} * LOSS_{BC} \quad (1)$$

where $LOSS_{NN}$, $LOSS_{PDE}$, $LOSS_{BC}$ denote the losses between predicted values from the neural network and observed data, PDE calculated values, and values at the boundary, respectively. The w indicates the weight coefficient for each related $LOSS$ value. In this study, the one-dimensional heat conduction equation and Fourier's law are adopted for the physical rules:

$$\frac{d^2T}{dz^2} = 0 \quad (2)$$

$$q = -K \cdot \frac{dT}{dz} \quad (3)$$

where z represents the depth, and q and K are the heat flow value and thermal conductivity of the rock, respectively. T is the abbreviation for temperature, and z is for depth. For boundary conditions, the temperature at the surface and a fixed temperature value (585°C) at the Curie point depth were utilized. By converting the PDE equations, the physical constraints of the PINN are shown as follows.

$$T_{z=0} = T_{surface} \quad (4)$$

$$T_{Curie} = 585.0 \quad (5)$$

$$T = T_{upper} + \frac{q \cdot (z - z_{upper})}{K} - \frac{A(z - z_{upper})^2}{2K} \quad \text{for sedimentary layers} \tag{6}$$

$$T = T_{upper} + \frac{(z - z_{upper})}{K}(q - AD) + \frac{AD^2}{K}\left\{1 - \exp\left(\frac{-(z - z_{upper})}{D}\right)\right\} \quad \text{for crystalline crust} \tag{7}$$

where $T_{surface}$ and T_{Curie} are the input values of surface temperature and Curie point temperature, both used for a boundary condition. The partial differential equation is founded on one-dimensional steady-state equation, subscript "upper" in T and z represents the state at the upper boundary, and q, D, A, and K denote heat flow, layer thickness, heat production rate, and thermal conductivity, respectively. For sedimentary layers, the heat-generating radioactive materials are assumed to be uniformly distributed, while they are concentrated in shallow parts and exponentially decrease in the crystalline crust (Ma, 2023). Specific heat production rate and layer thickness values are retrieved from the study made by Wang et al. (2023). If the target is in a particular layer, the calculation starts from a known boundary (mostly, ground surface) and proceeds from shallower to deeper boundaries, ultimately calculating the target point value from the boundary immediately above the target point. The calculation uses either equation (6) or (7) above, depending on the layer attributes passed through on the way from boundary to boundary or target. By such calculation, the temperature consistency at the multi-layered structure is guaranteed.

As the neural network, a model consisting of three fully connected hidden layers with rectified linear unit (ReLU) as an activation function is selected. The weight coefficients for the loss of PDE, boundary condition, and neural network are fixed at 5.0, 1.0, and 1.0, respectively.

The borehole temperature dataset from the National Oceanic and Atmospheric Administration (NOAA) was employed as the primary training data. Due to the extreme scarcity of available data points, the training dataset was augmented using temperature-at-depth data linearly interpolated between the ground surface and the Curie point depth. For the estimation purpose, the target area was discretized into 1 km × 1 km horizontal grids. The model is engineered to produce a temperature distribution map for all grids at the targeted depth, with depth contours ranging from the surface down to the Curie point depth. By integrating multiple temperature-at-depth contour maps, a three-dimensional (3D) distribution is subsequently simulated. For computational practicality in this study, the depth interval for the contour maps was fixed at 1 km, meaning temperature predictions were generated at the center of each 1 km × 1 km × 1 km voxel within the target region.

2.2 Feature Engineering

In addition to the borehole temperature data, feature data were incorporated at both the input data and estimation points. These features include the Curie point depth (Xiong et al., 2016), rock type (Pang et al., 2017), heat flow, thermal conductivity, thermal gradient (Hu et al., 2001; Jiang et al., 2019; Jiang et al., 2016; Wang et al., Wang et al., 1988; Wang et al., 1990; Wang et al., 2024), volcano location, gravity acceleration, gravity disturbance, and Bouguer anomaly. Specifically, for the volcano location data, the Euclidean distance to the nearest volcano was calculated and used. For the other features provided in raster format, the data were directly extracted at the corresponding data points.

2.3 Accuracy Evaluation

For accuracy evaluation, a five-part K-fold cross-validation was adopted. In the process, input data are split into training and test data, and the accuracy is evaluated by the difference of predicted values from observed values, PDE calculated values, and boundary values, respectively. Since the boundary condition is only effective at the boundary, loss from the boundary condition is only calculated at the surface and the Curie point depth. For the loss function, the mean squared error (*MSE*) was used as

$$MSE = \frac{1}{n}\Sigma_{i=1}^{n}(y_i - \hat{y_i})^2 \tag{8}$$

where y_i is an observed, boundary, or calculated value, $\hat{y_i}$ is a predicted value, and n is the number of observations. The change of loss values with epochs is depicted in Figure 2.

Generally, the loss calculated with observed values and PDE calculation values decreases as learning proceeds. BC loss shows an increase at the start, and then the loss is almost static until the end of training. The total loss shows a decreasing trend during the whole process.

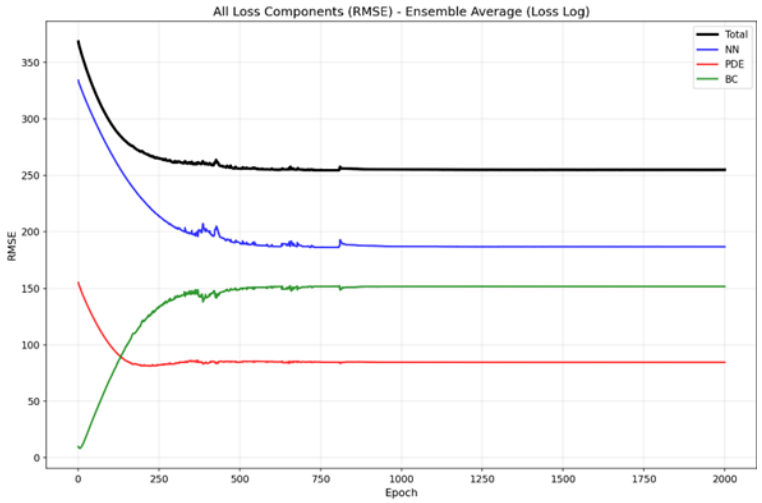


Figure 2. Loss changes with epochs.

3 RESULTS AND DISCUSSION

The resultant temperature distribution map at -3000m is shown in Figure 3. The result reveals high subsurface temperature in the southwestern part of China, especially in Tibet, where the largest and only geothermal plant of China is located. Tibet is known for its tectonic interaction with the Indian plate, which indicates the effect of orogenic activity on the temperature-at-depth.

While this trend matches the fact that the utilization of geothermal energy, including hot springs in southern Tibet, the predicted result does not point out other areas which have geothermal indications, such as the Tengchong area in Yunnan, Taiwan, and the coastal region of Fujian. Generally, the predicted result shows higher values in the Tibet area and a lower trend in the north SinKiang and northeast China, indicating that the result is strongly affected by the surface temperature distribution, and is insufficient to precisely capture promising geothermal resource sites with a high spatial resolution.

As the mismatch between BC and PDE/observation is implied from the loss change log, revise of physical constraints and weight coefficient optimization could be an immediate, applicable improvement. Then, an optimization using the tree-structured Parzen estimator sampler (TPE sampler) is in progress. Furthermore, additional PDEs, such as Navier-Stokes equations for subsurface fluids, or a combination with geostatistics represented by neural kriging (Koike et al., 2022) can be taken into consideration.

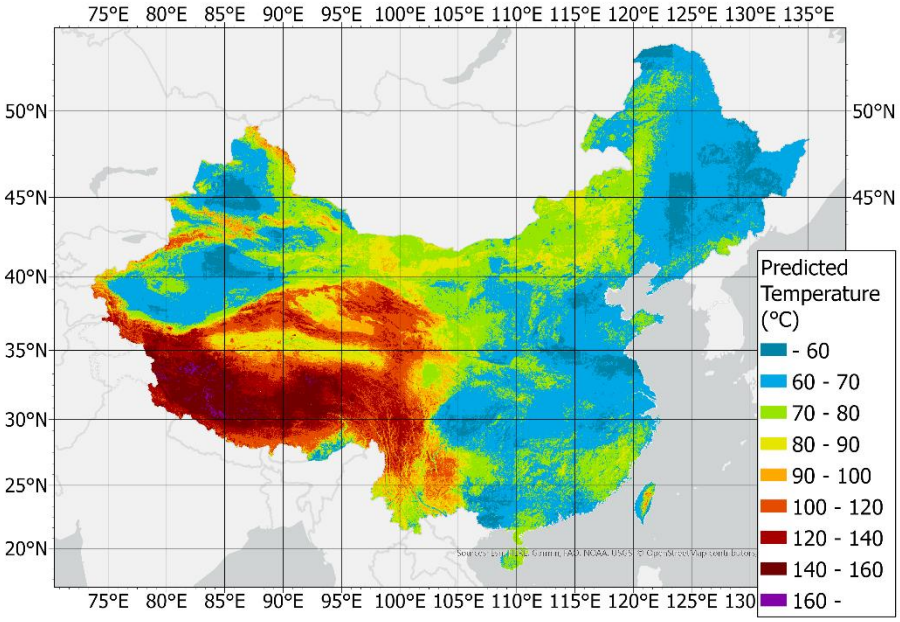


Figure 3. Predicted temperature distribution map at -3000m.

4 CONCLUSION

A physics-informed neural network considering heat-related equations and the surface temperature and Curie point depth as constraints was used to generate a temperature-at-depth map from point-based datasets throughout China. The result showed a reasonable temperature range at the predicted depth range, and a close predicted value with the measured ones in certain areas. However, the trend of wide area temperature distribution was strongly affected by the measured surface temperature, resulting in a mismatch with the actual utilization case of geothermal energy in China. The result was not sufficient to narrow down the potential geothermal point, and further improvement is necessary.

REFERENCES

Hu, S., He, L. and Wang, J.: Compilation of heat flow data in the continental China (3rd edition), *Chinese Journal of Geophysics*, 44(05), (2001), 611-626.

Ieki, Y., Koike, K. and Kubo, T.: Deep learning-based three-dimensional terrestrial temperature modeling throughout Japan incorporating multiple crustal properties and spatial correlation with an application to critical point distribution. *Geothermics*, 131, (2025), 103403.

Ishitsuka, K. and Weiren, L.: Physics-informed neural network for inverse modeling of natural-state geothermal systems. *Applied Energy*, 337, (2023), 120855.

Jiang, G., Hu, S., Shi, Y., Zhang, C., Wang, Z. and Hu, D.: Terrestrial heat flow of continental China: Updated dataset and tectonic implications. *Tectonophysics*, 753, (2019), 36-48.

Jiang, G., Gao, P., Rao, S., Zhang, L., Tang, X., Huang, F., Zhao, P., Pang, Z., He, L., Hu, S. and Wang, J.: compilation of heat flow data in the continental area of China (4th edition). *Chinese Journal of Geophysics*, 59(8), (2016), 2892-2910.

Koike, K., Yono, O., de Sá, V. R., Tomita, S. A., Nozaki, T., Takaya, Y. and Komori, S.: Effectiveness of neural kriging for three-dimensional modeling of sparse and strongly biased distribution of geologic data with an application to seafloor hydrothermal mineralization. *Mathematical Geosciences*, 54, (2022), 1189-1206.

Ma, Y.: Deep geothermal resources in China: Potential, distribution, exploitation, and utilization. *Energy Geoscience*, 4, (2023), 100209.

Pang, J., Ding, X., Han, K., Zeng, Y., Chen, A., Zhang, Y., Zhang, Q. and Yao, D.: The National 1: 1000000 Geological Map Spatial Database. *Geology in China*, 44(S1), (2017), 8-18.

Wang, G., Liu, F., Lin, W., Zhang, W., Yuan, R., Xi, Y., Wei, S., Liao, Y. and Wang, Y.: The crustal heat production rate and crustal and mantle heat flow distribution in the land of China. *CHINESE JOURNAL OF GEOPHYSICS*, 66(12), (2023), 5041-5056.

Wang, J. and Huang, S.: Compilation of heat flow data in the continental area of China (1st edition), *Chinese Journal of Geology*, 02, (1988), 196-204.

Wang, J. and Huang, S.: Compilation of heat flow data in the continental area of China (2nd edition), *SEISMOLOGY AND GEOLOGY*, 12(04), (1990), 351-363.

Wang, Y., Liu, S., Chen, C., Jiang, G., Wu, J., Guo, L., Wang, Y., Zhang, H., Wang, Z., Jiang, X., Zhu, C., Zuo, Y., Hu, J., Zhang, C., He, L., Hu, S., Pang, Z. and Wang, J.: Compilation of terrestrial heat flow data in continental China (5th edition). *Chinese Journal of Geophysics*, 67(11), (2024), 4233-4265.

Xiong, S., Yang, H. and Li, Z.: Characteristics of Chinese continent Curie point isotherm. *CHINESE JOURNAL OF GEOPHYSICS*, 59(10), (2016), 3604-3617.