PROCEEDINGS, 49th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 12-14, 2024
SGP-TR-227

# Evaluation of Machine Learning Models Based on Different Data Preprocessing Methods of Predicting Geothermal Heat Flow in China

Jifu He[1], Kewen Li[1], Lin Jia[1]

[1] China University of Geosciences, Beijing, 100083, China

likewen1268@163.com

**Keywords:** Machine learning, Gradient boosted regression tree, Data preprocessing method, Geothermal heat flow, China.

## ABSTRACT

Estimating geothermal heat flow (GHF) is critical for understanding Earth's thermal regime and facilitating geothermal exploration. Machine learning techniques offer a promising approach for deriving GHF estimates based on various geological and geophysical features. While various studies have employed different data preprocessing methods for global GHF datasets, the influence of these methods on GHF predictions has not been thoroughly investigated. In this study, we assess the impact of four distinct preprocessing techniques applied to GHF datasets on GHF prediction accuracy within China using the Gradient Boosted Regression Trees (GBRT) algorithm. The results showed that more features are needed for the model performance to converge after the dataset is processed on average. Moreover, averaging GHF data considerably influences the model's assessment of feature importance, resulting in lower GHF estimations and a reduced prediction range. We also find that the prediction accuracy of GBRT models on the test set can be substantially improved by applying a low-pass filter to the GHF data. Upon further comparison, the model derived from the low-pass filtered dataset exhibits the highest predictive capability for GHF in China. Lastly, we present eight new GHF maps for China, which consistently indicate high GHF values in Tibet and eastern China, and low GHF values in northwest China. Our models also identify several unobserved high-heat regions, potentially offering valuable insights for geothermal exploration.

## 1. INTRODUCTION

Geothermal heat flow (GHF) serves as a surface manifestation of the thermal state of the Earth's interior and plays a crucial role in geothermal exploration (Chandrasekharam et al., 2022; He et al., 2022; Jiang et al., 2019; Lucazeau, 2019). GHF measurements are typically calculated by multiplying the rock thermal conductivity by the temperature gradient. Nevertheless, acquiring temperature gradient measurements is both practically costly and time-consuming, and the measurements can be influenced by surface perturbations such as water circulation, erosion, and climatic changes (Lucazeau, 2019). Consequently, GHF measurements exhibit a sparse distribution across many parts of the world, such as South America and Africa (Rezvanbehbahani et al., 2017). In China, despite recent increases in GHF measurements, an uneven distribution persists (Fig. 1), with the majority of data points situated in the North China Craton (NCC) and Wuyi-Yunkai Orogen (WYO). In unmeasured regions, GHF is primarily extrapolated via interpolation (Jiang et al., 2019).

GHF is related to geological and geophysical parameters (Bodri & Bodri, 1985; Li et al., 2013; Wang & Li, 2018), making it possible to estimate heat flow in areas without measurement. For instance, using the seismic wave velocity model to estimate GHF in Antarctica (Shapiro & Ritzwoller, 2004) or using thermotectonic age and geological characteristics to extrapolate GHF (Davies & Davies, 2010). However, these methods predominantly rely on a single parameter, capturing only a fraction of the processes impacting GHF. To better characterize the variety of geodynamic environments, an integrated dataset including multiple parameters was built and used to predict GHF based on empirical approaches of GHF extrapolation (Goutorbe et al., 2011). However, empirical methods may not offer optimal predictions and might not necessarily constitute the most suitable choice.

In recent years, machine learning has emerged as a promising alternative for GHF prediction (Rezvanbehbahani et al., 2017). And this method has been applied to estimate GHF in Greenland, Antarctica (Lösing & Ebbing, 2021), and Bohai Bay Basin (He et al., 2022). Different datasets were used in these studies to build the models. The study by Rezvanbehbahani et al. (2017) used a processed dataset in which GHF data were averaged on a 1 × 1 degree grid that reduced the number of data points from 35000 to more than 4000, and these averages were low-pass filtered. Losing & Ebbing (2021) only reduced the resolution of GHF to 0.5 degrees without low-pass filtering and left more data points, about 10000. Different data processing methods will inevitably lead to differences in GHF prediction. However, few studies have examined the uncertainty caused by different datasets on the model. Additionally, which data processing method can provide more accurate predictions is also worth studying.

In this paper, we utilize the GBRT algorithm to generate GHF models for China based on 16 geological and geophysical features. Four datasets processed in different ways are used to construct GBRT models, and the differences between the prediction results of the four models are discussed. Furthermore, the effect of the number of features on model performance is also investigated. Finally, we present several new heat flow maps for China.

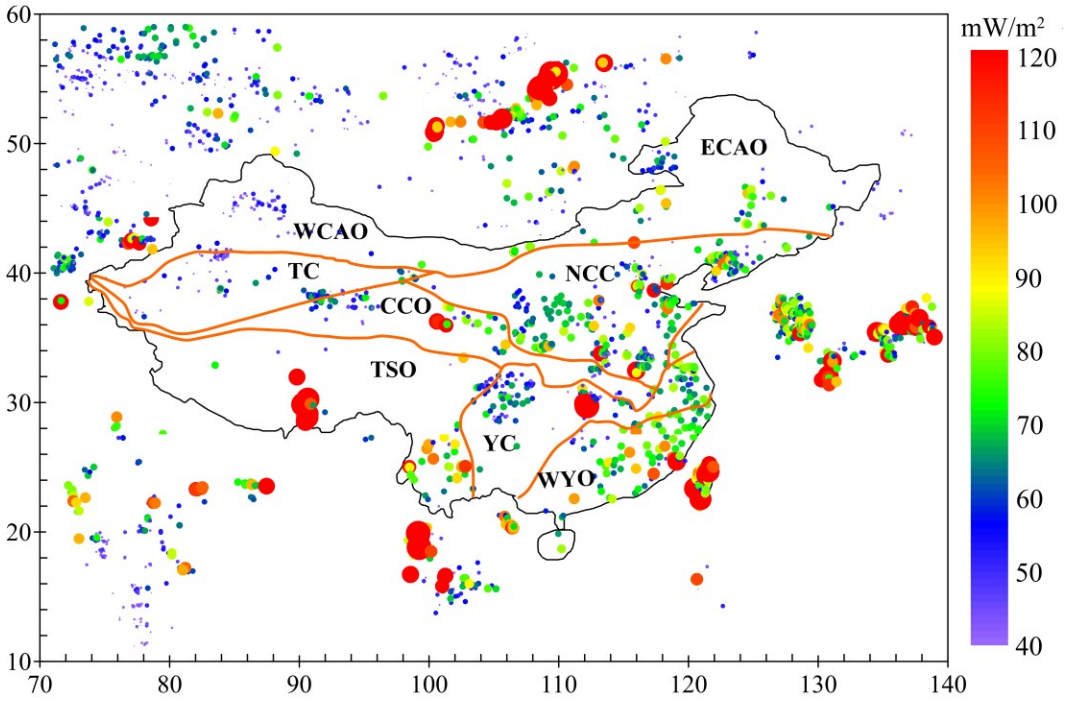


**Figure 1: Distribution of GHF measurements in and around China. The relative size of data points corresponds to the magnitude of GHF. WCAO: Western Central Asian Orogen, TC: Tarim Craton, CCO: Central China Orogen, TSO: Tibet-Sanjiang Orogen, YC: Yangtze Craton, WYO: Wuyi-Yunkai Orogen, NCC: North China Craton, ECAO: Eastern Central Asian Orogen.**

## 2. DATA

### 2.1 Geothermal Heat Flow

Global GHF measurements on the continents are extracted from the International Heat Flow Commission, and measurements lower than 20 mW/m$^2$ or higher than 200 mW/m$^2$ are excluded as they may be influenced by local surface processes (Goutorbe et al., 2011). There were 36325 data remaining after the above screening, and four datasets were constructed using these data (Fig. 2). The first dataset comprises the unprocessed original data. The second dataset features the original data subjected to low-pass filtering, mitigating short-scale spatial variability while maintaining the quantity of data points, resulting in a more constrained GHF distribution (Fig. 2). The third dataset is derived from spatially averaging the original data using a 0.5° × 0.5° grid (similar to Lösing & Ebbing, 2021), consequently reducing the data volume from 36325 to 7740. Lastly, the fourth dataset involves spatial averaging of the original data on a 1° × 1° grid, which reduces the data count from 36,325 to 3,939, followed by low-pass filtering (similar to Rezvanbehbahani et al., 2017).

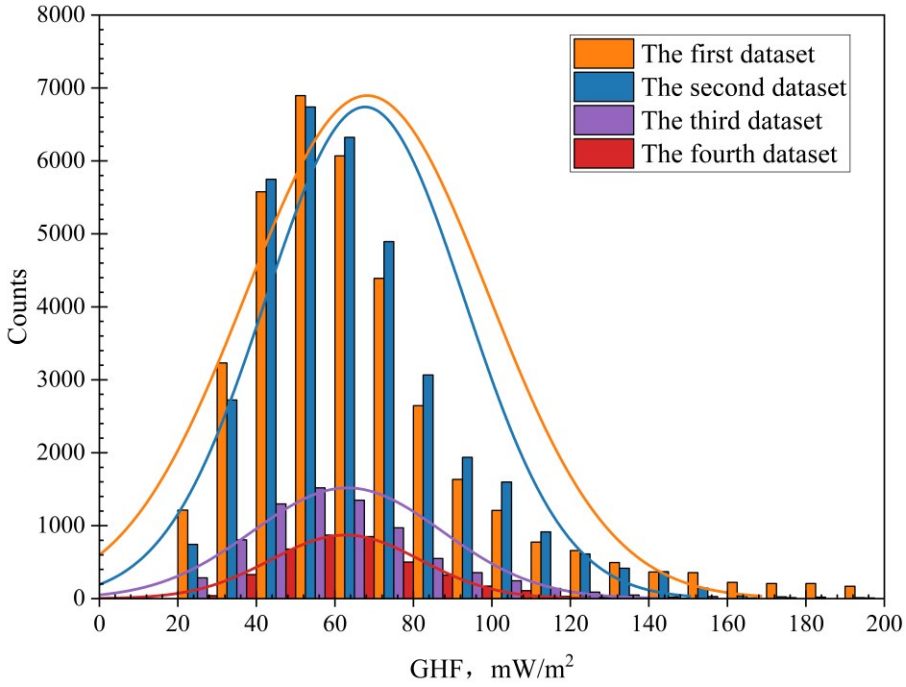


**Figure 2: Histogram of GHF of the four datasets.**

### 2.2 Geological and Geophysical Features

Sixteen geological and geophysical features related to GHF were included in our study (Table 1). Next, we outline these features and their sources. We extracted Moho depth data from the crustal model built by Szwillus et al. (2019) using active seismic data from the USGS GSC database. It has a typical difference of 3 km on the continent compared to CRUST 1.0 but a smoother transition. The sedimentary thickness affects GHF through the radiogenic heat component in the sediments (Goutorbe et al., 2011) and the redistribution of GHF caused by the thermal conductivity difference between the sedimentary caprock and the bedrock (He et al., 2022). We choose the sedimentary thickness from CRUST 1.0 model. Magnetic signals from the crust are affected by the subsurface temperature, which makes it an important feature correlated with GHF. The temperature at the Curie depth demagnetizes the ferromagnetic minerals in the crust. We used the EMAG2 magnetic anomaly model with a 2-arc-minute resolution compiled from satellite, ship, and airborne magnetic measurements (Maus et al., 2009). Gravity anomalies can indicate geological structural features that may influence the distribution of GHF. We used the WGM2012 gravity anomalies model that is derived from EGM2008 and DTU10 global gravity models (Balmino et al., 2012). It has a one-arc-minute resolution and considers the contribution of most surface masses. Lithosphere-asthenosphere boundary (LAB) depth data was extracted from the LITHO1.0 model that was generated by examination and discussion of the model concerning crucial lithospheric parameters, such as average crustal velocity, crustal thickness, upper mantle velocity, and lithospheric thickness (Pasyanos et al., 2014). Radioactive heat production is thought to contribute about half of global GHF (Mareschal et al., 2017). We choose an approximate heat production model established by Goutorbe et al. (2011) that can capture the large-scale trends of surface heat production. Plate boundaries and volcanic fields are where geothermal activity is most intense and affect the surrounding GHF distribution. We introduce the distance parameters to plate boundaries (ridge, trench, and transform) and volcanoes into the study to consider their effects on heat flow. We used distance measures to plate boundaries from the compilation of plate boundaries provided by the University of Texas Institute for Geophysics (Coffin et al., 1997). Unlike other studies (Goutorbe et al., 2011; Lösing & Ebbing, 2021; Rezvanbehbahani et al., 2017), we also included active fault density data into the data set (Table 1). High fault density may indicate abnormal crustal thermal conditions. We infer the global fault density distribution based on the GEM Global Active Faults Database, which contains 13500 faults (Styron & Pagani, 2020).

**Table 1: Data Sets Included in the Feature Vector for Predicting GHF.**

| Number | Features | Reference |
|---|---|---|
| 1 | Moho depth | (Szwillus et al., 2019) |
| 2 | Crustal thickness | (Laske et al., 2013) |
| 3 | Sedimentary thickness | (Laske et al., 2013) |
| 4 | Magnetic anomaly | (Maus et al., 2009) |
| 5 | Topography | (Amante & Eakins, 2009) |
| 6 | Bouguer gravity anomaly | (Balmino et al., 2012) |
| 7 | Isostatic gravity anomaly | (Balmino et al., 2012) |

| | | |
|---|---|---|
| 8 | LAB depth | (Pasyanos et al., 2014) |
| 9 | Heat production province | (Goutorbe et al., 2011) |
| 10 | Distance to ridge | (Coffin et al., 1997) |
| 11 | Distance to trench | (Coffin et al., 1997) |
| 12 | Distance to transform | (Coffin et al., 1997) |
| 13 | Distance to young rift | (Goutorbe et al., 2011) |
| 14 | Distance to volcanoes | (Goutorbe et al., 2011) |
| 15 | Upper mantle density anomaly | (Goutorbe et al., 2011) |
| 16 | Density of active faults | Inferred from (Styron & Pagani, 2020) |

## 3. METHODOLOGY

### 3.1 Gradient Boosted Regression Tree

In this paper, we choose Gradient Boosted Regression Tree (GBRT) algorithm to estimate GHF using the above 16 geological and geophysical features. The GBRT algorithm is a kind of forward-learning ensemble method that obtains predictive results through gradually improved estimations. It combines weaker learners (decision trees) into a single strong learner through iteration (Fig. 3). And each new tree model is built to reduce the residuals of the previous models towards the gradient direction (Fig. 3), which is quite different from the traditional Boost that weights the samples. As a result, GBRT models can reduce bias and outperform other tree-based algorithms in terms of overfitting and computational costs (Fan et al., 2018). In addition, the algorithm does not need to scale the data to perform well. The approximation function for the GBRT can be expressed as

$$F(x) = \sum_{n=1}^{N} \beta_m h(x; a_n) \tag{1}$$

where $x$ represents the input variables and $a_n$ denotes the classifier of each decision tree. Each tree can be defined as $h(x;a_n)$, and $\beta_m$ denotes the weight of each tree.

During the modeling process, 80% of the data were classified as the training set to find the optimal model, and the remaining 20% was used as the test set to evaluate the optimal model. A 4-fold cross-validation was used in the training process. The optimal model was obtained by adjusting the combination of regression tree number, depth, and learning rate.

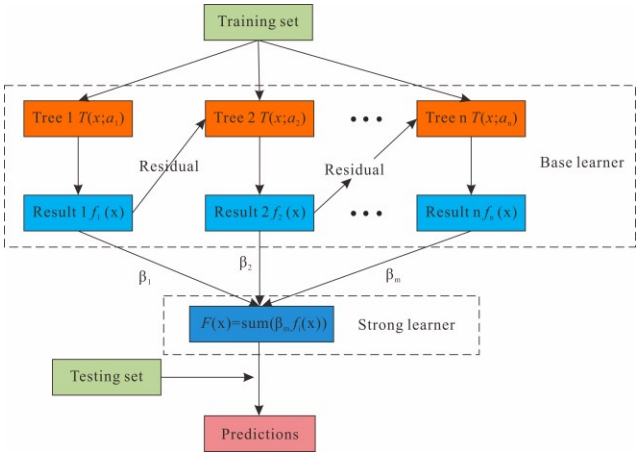


**Figure 3: Diagram of GBRT algorithm (He et al., 2022).**

### 3.2 Evaluation Methods

In this paper, three general statistical indicators are used to evaluate the accuracy of the established model, including coefficient of determination ($R^2$), mean absolute error (MAE), and root mean square error (RMSE).

(1). Coefficient of Determination ($R^2$): the linear correlation between the measured GHF and predicted GHF in the test set, which can be expressed as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \tag{2}$$

where $X_i$ is the actual GHF, $Y_i$ is the predicted GHF, and n is the number of outputs, $\bar{X}$ is the average of $X_i$.

(2). Root Mean Square Error (RMSE): the deviation between measured GHF and predicted GHF, which can be expressed as

$$RMSE = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{3}$$

(3). Mean Absolute Error (MAE): the absolute measure of error between measured GHF and predicted GHF in the test set, which can be expressed as

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - Y_i\right| \tag{4}$$

## 4. RESULTS

### 4.1 Testing Individual Features

The features used in the model will significantly affect the GHF predictions. Therefore, it is necessary to understand the importance of individual features for predicting GHF. Here, we tested the impact of each feature on the predicted performance under different datasets by using individual features to predict GHF. A discernible trend emerges wherein elevated $R^2$ values are typically accompanied by reduced MAE and RMSE (Fig. 4).

Notably, the relative importance rankings of features for the first and second datasets exhibit remarkable similarities, yet they contrast sharply with those for the third and fourth datasets. This is probably due to the fact that the average operation greatly changes the quantity of data and the intervals of GHF values in the third and fourth datasets. Despite these differences, certain features, such as LAB, heat production province, distance to young rifts, and distance to volcanoes, consistently demonstrate substantial importance across all four datasets. These features yield higher $R^2$ scores and lower MAE and RMSE values.

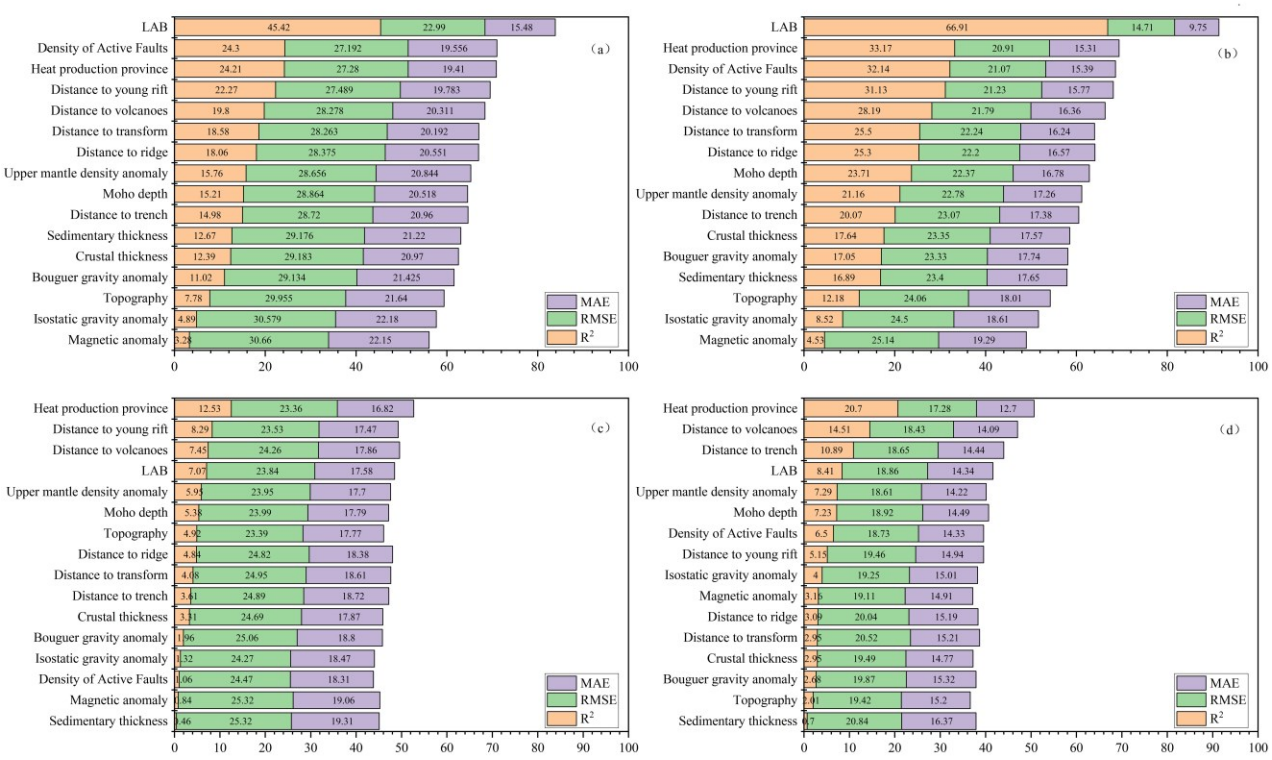


**Figure 4: Scores of individual features on $R^2$, RMSE, and MAE. (a)~(d) correspond to the first to the fourth dataset.**

In addition, we investigated the impact of the number and combinations of features on the performance of models under different datasets by gradually increasing features (Fig. 5). Five modes were employed in increasing features. Mode 1 progressively increases the features according to their scores (Fig. 4) from highest to lowest. Specifically, for the first dataset, we start with the LAB, add the density of active faults subsequently, then the heat production province, and so on. Instead, Mode 2 adds features in order of feature score from lowest to highest, starting with the magnetic anomaly, adding the isostatic gravity anomaly subsequently, then topography, and so on. Modes 3 to 5 add features randomly.

The results reveal convergent behavior in both $R^2$ scores and RMSE across the five modes, which is more pronounced in the first and second datasets (Fig. 5). Initially, $R^2$ scores increase and RMSE values decrease substantially with the addition of features. However, as more features are incorporated, these evaluation metrics plateau, and the subsequent inclusion of features exhibits minimal impact. The convergence threshold for the first and second datasets ranges between 5 and 8 features, whereas for the third and fourth datasets, which were subjected to averaging, the threshold is 9 or higher (Fig. 5 and Fig. 6). Furthermore, Mode 2 requires more features to converge than other modes, primarily due to unsatisfactory individual scores for the initial few features incorporated in this mode. However, for a more consistent comparison, we still employed all the features in predicting GHF in China in the subsequent study.

The impact of feature combinations on models derived from the third and fourth datasets is markedly more significant than their impact on models generated from the first and second datasets. When the number of features included in models for the first and second datasets is greater than eight, the $R^2$ score and RMSE are relatively comparable across all feature combinations (Fig. 5 and Fig. 6). However, for the third and fourth datasets, the difference in model performance under different feature combinations is still quite noticeable even when the number of features is relatively large.

Additionally, it can be seen that models based on different datasets have different $R^2$ scores and RMSE on their respective test sets (Fig. 5 and Fig. 6). Typically, datasets that have undergone low-pass filtering have greater $R^2$ and lower RMSE values, such as the second and the fourth dataset (Fig. 5b and Fig. 6b). Conversely, the averaged dataset has a lower $R^2$ score and RMSE, as exemplified by the third dataset (Fig. 5c and Fig. 6c). However, we cannot perform cross-model comparisons based on these findings since their test sets are inconsistent. Therefore, in the subsequent investigation, we will utilize the same test set to compare the accuracy of several GHF prediction models in China.

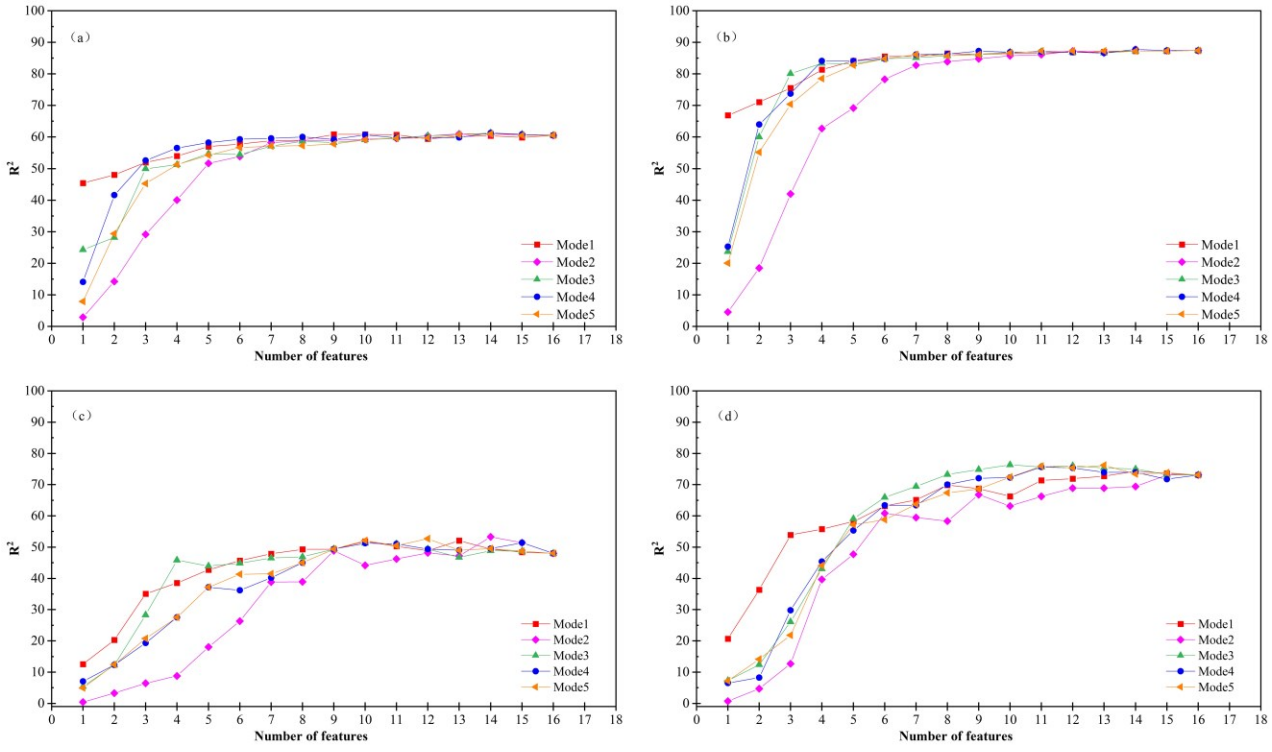


**Figure 5: Effect of the number and combination of features on $R^2$ score of models. (a)~(d) correspond to the first to the fourth dataset.**

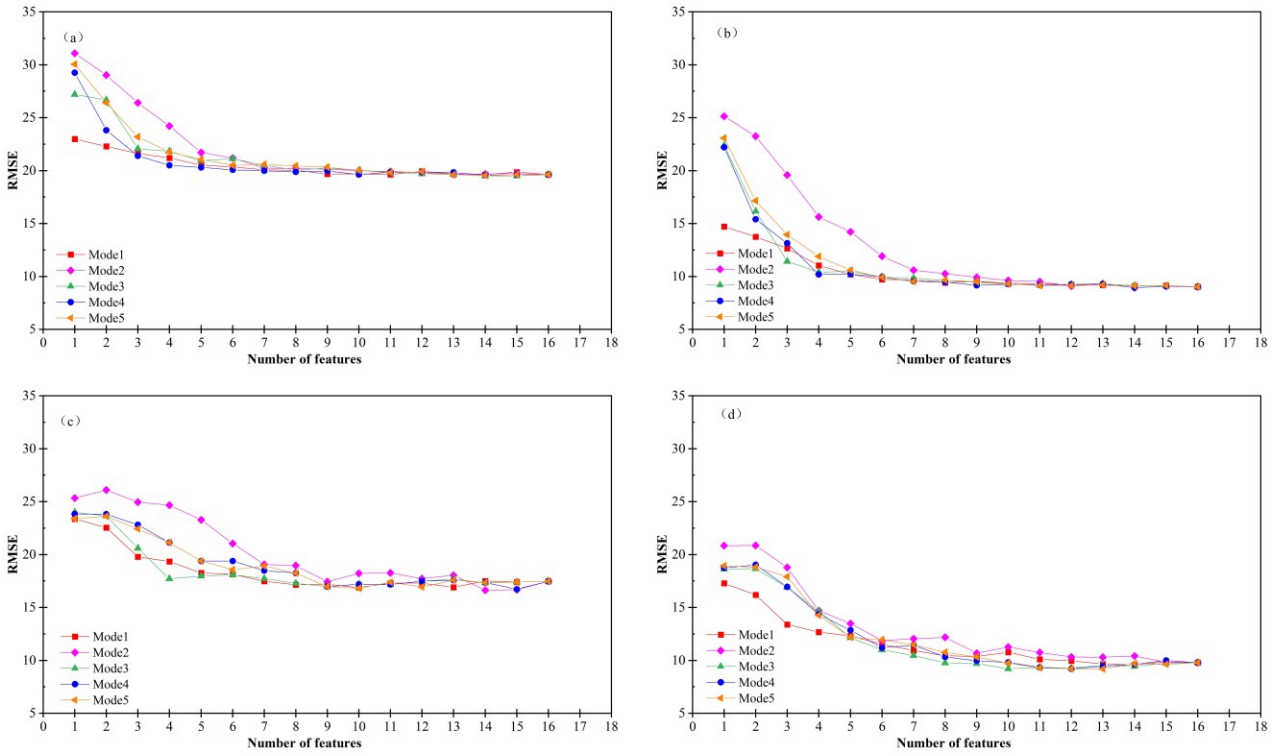


**Figure 6: Effect of the number and combination of features on RMSE of models. (a)~(d) correspond to the first to the fourth dataset.**

### 4.2 Predicting GHF for China using data from outside China

In this section, we first examine the GBRT model's ability to predict China's heat flow in the absence of Chinese heat flow data. We exclude Chinese data from the aforementioned four datasets and employ the residual data to construct models. Figure 7 depicts the performance of GBRT models derived from different datasets on the test set. The model based on the second dataset exhibits the most optimal performance on the test set, followed by the model derived from the fourth dataset (Fig 7). This outcome is likely attributable to the low-pass filtering operation applied to both datasets. However, this does not necessarily imply that models based on the second dataset possess superior predictive capabilities compared to the others, as their test sets are inconsistent. Consequently, in the subsequent study, we will utilize the original Chinese GHF data as the test set to facilitate a more consistent comparison among models.

The GHF predictions of GBRT models based on different datasets are shown in Figure 8. The dots in Figure 8 represent the measured GHF. Overall, the GHF predictions of the model based on the third and fourth data sets are lower than that of models based on the first and second data sets. This is mainly because the average operation reduces the data with high GHF values in the third and fourth data sets (Fig. 2).

It is visible that Yangtze Craton and Tarim Craton are less uncertain than other regions. All the models predict the two tectonic units as relatively cold regions. In contrast, Tibet-Sanjiang Orogen is considerably more uncertain than in other areas. The prediction of hot and cold regions by various models is quite different for Tibet-Sanjiang Orogen (Fig. 8). However, the predicted GHF values of the model based on the first dataset are much closer to the measured values.

Based on the first and second data sets, models predict a banded zone with high GHF in the middle of the North China Craton that goes from southwest to northeast (Fig. 8a and 8b). This is supported by the GHF measurements. However, models based on the third and fourth data sets did not identify this trend (Fig. 8c and 8d). In the eastern part of the Eastern Central Asian Orogen, models based on the first and second datasets predict a broader region with relatively high GHF than other models, which is more consistent with the GHF measurements. For Central China Orogen, models based on the first and second data sets predict a cold region in its western part and a hot region in its eastern part (Fig. 8a and 8b). However, models based on the third and fourth data sets predict a wider cold region (Fig. 8c and 8d). By comparing with the GHF measurements, we believe that the prediction of the model based on the first and second data sets is more in line with the reality.

Furthermore, we counted the performance of GBRT models based on different datasets in predicting the raw GHF measurements in China (Fig. 9). It is evident that GBRT models are able to identify the relative magnitude of GHF between different regions (Fig. 8), but their prediction accuracy for specific GHF values is poor (Fig. 9). In addition, the MAE and RMSE scores of the different models do not differ much, but the $R^2$ score of the model based on the first dataset is significantly higher.

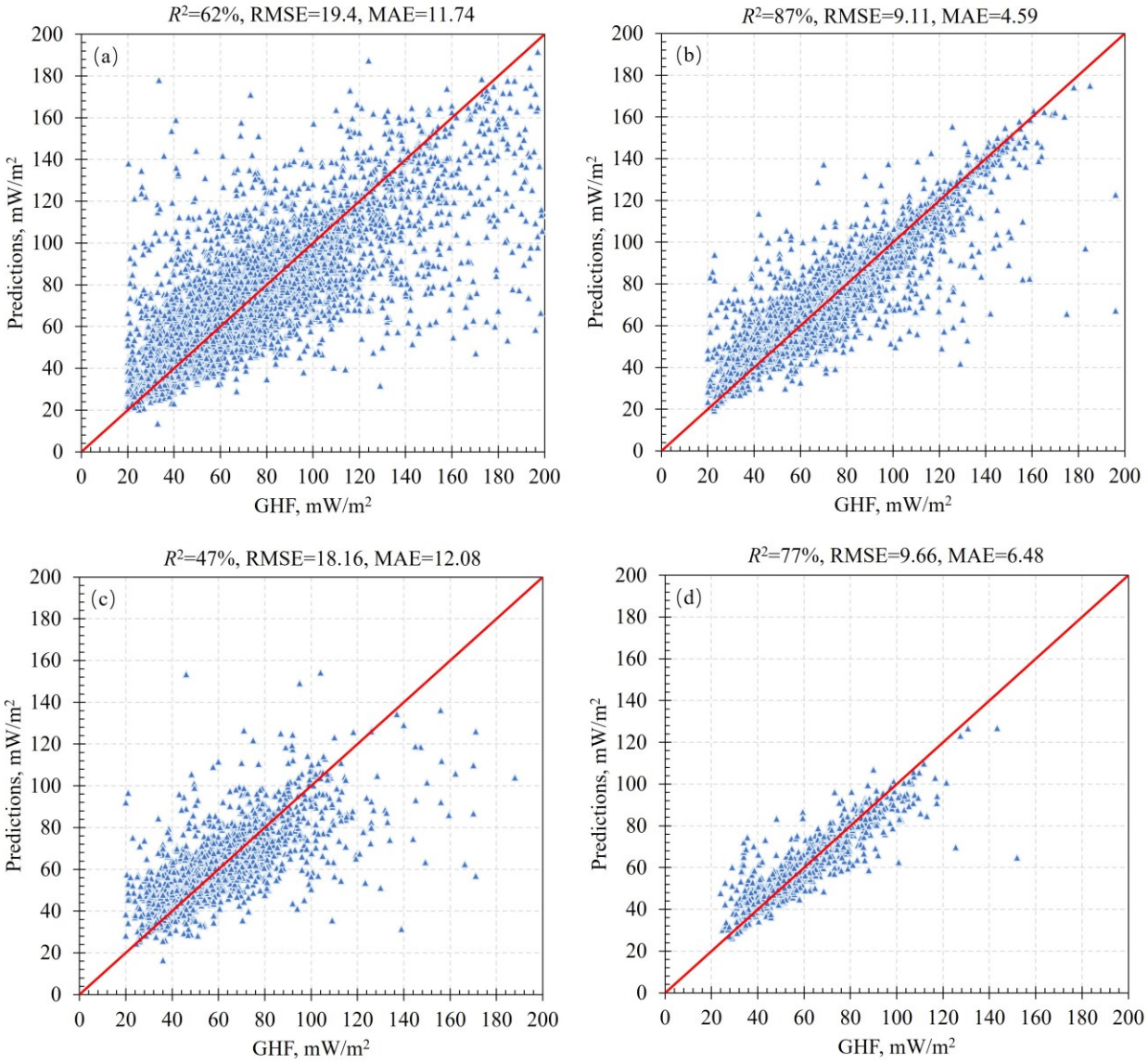


**Figure 7: Performance of models based on different data sets on their respective test sets. (a)~(d) corresponds to the first to the fourth dataset.**

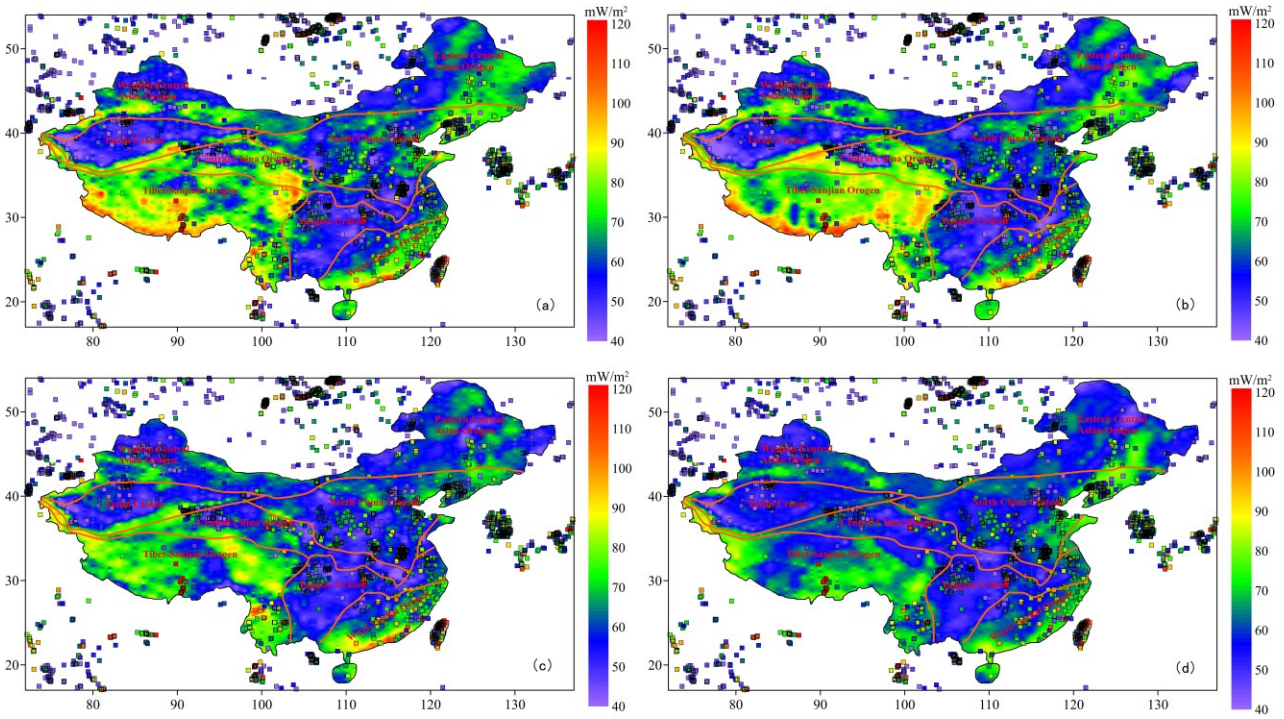


**Figure 8: GHF predictions for China by using data from outside China. (a)~(d) corresponds to the predicted GHF of the models built using the first to the fourth dataset.**

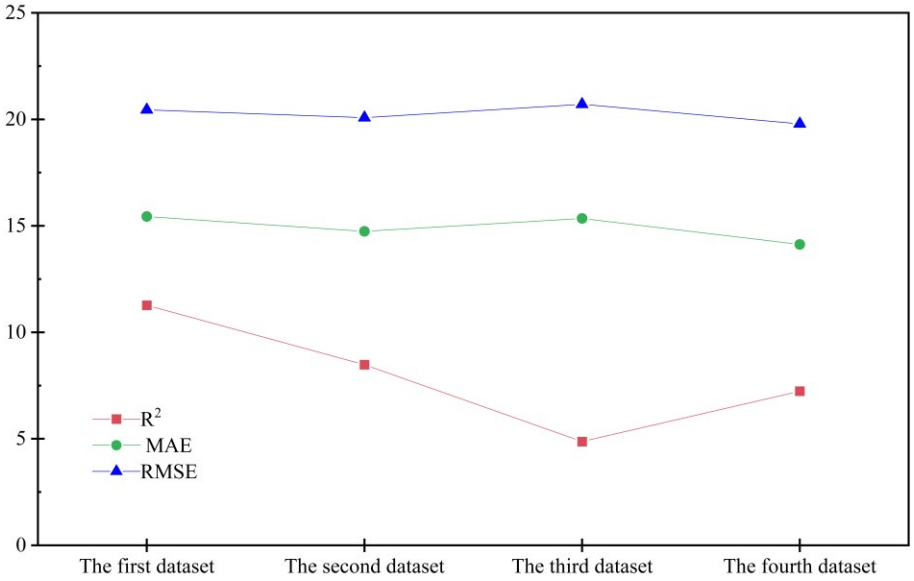


**Figure 9: Performance of GBRT models based on different datasets in predicting GHF measurements in China (None of the datasets contain data from China).**

### 4.3 Predicting GHF for China using the complete dataset

In this part, we constructed GBRT models utilizing the entire dataset incorporating China data. Four GBRT models, Model A, Model B, Model C, and Model D, were developed based on the first through fourth data sets. Figure 10 illustrates the GHF predictions of several models. They exhibit comparable GHF distribution characteristics to those predicted using data from outside China (Fig. 8 and Fig. 10). They all predicted relatively high GHF for the eastern boundary of mainland China, Tibet-Sanjiang Orogen, and eastern Central China Orogen and relatively low GHF for Tarim Craton and Yangtze Craton. Nevertheless, many differences exist between them. Tibet-

Sanjiang Orogen has the largest GHF prediction discrepancy, where the GHF values predicted using the entire data set are substantially greater. Similarly, the GHF predictions by models A~D are also higher than those by using data from outside China in many areas of eastern China. On the contrary, the GHF predictions of model A~D are lower in the West Central Asia Orogen and along the boundary of Tarim Craton and Central China Orogen.

Differences could also be seen across Models A, B, C, and D with regard to their predicted GHF. Model C predicts a larger area of high GHF in the Tibet-Sanjiang Orogen compared to the other models. Unfortunately, there aren't enough local observations to back up this prediction. Local GHF measurements and Models A, B, and C suggest a banded high GHF zone extending NE in the western East Central Asia Orogen. However, Model D misses this pattern. Model A predicts a high GHF zone at the boundary between the Tarim Craton and the Central China Orogen and the boundary between the Tarim Craton and the Western Central Asia Orogen, where there are no GHF observations (Fig. 10a). These high GHF regions gradually disappear with the deepening of data set processing. And Model D does not identify these regions (Fig. 10d).

In order to better compare the influence of different data sets on heat flow prediction, the performance of different models in predicting the raw GHF data, the mean GHF values in each 0.5°×0.5° cell, and the mean GHF values in each 1°×1° cell were calculated (Fig. 11). It shows that Model B has the best predictive performance in all three cases, with the highest $R^2$ score and the lowest RMSE and MAE, followed by Model A. And Model D performed the worst. Furthermore, the predictive performance of all models improved as the heat flow resolution decreased. And all models have the best performance in predicting the mean GHF values in 1°×1°cells (Fig. 11). This indicates that current models are effective in identifying large-scale GHF changes but limited in identifying small-scale GHF changes.

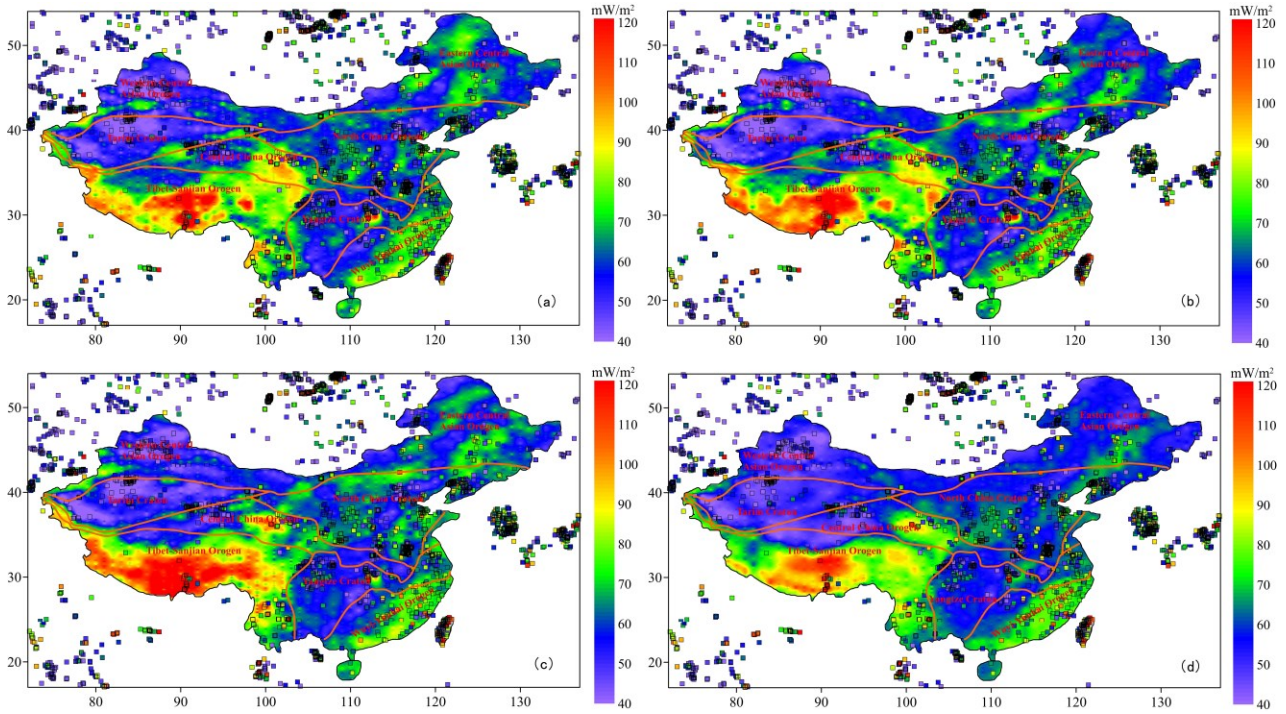


**Figure 10: Predicted heat flow for China by using the complete dataset. (a)~(d) corresponds to the predicted GHF by Model A~D.**

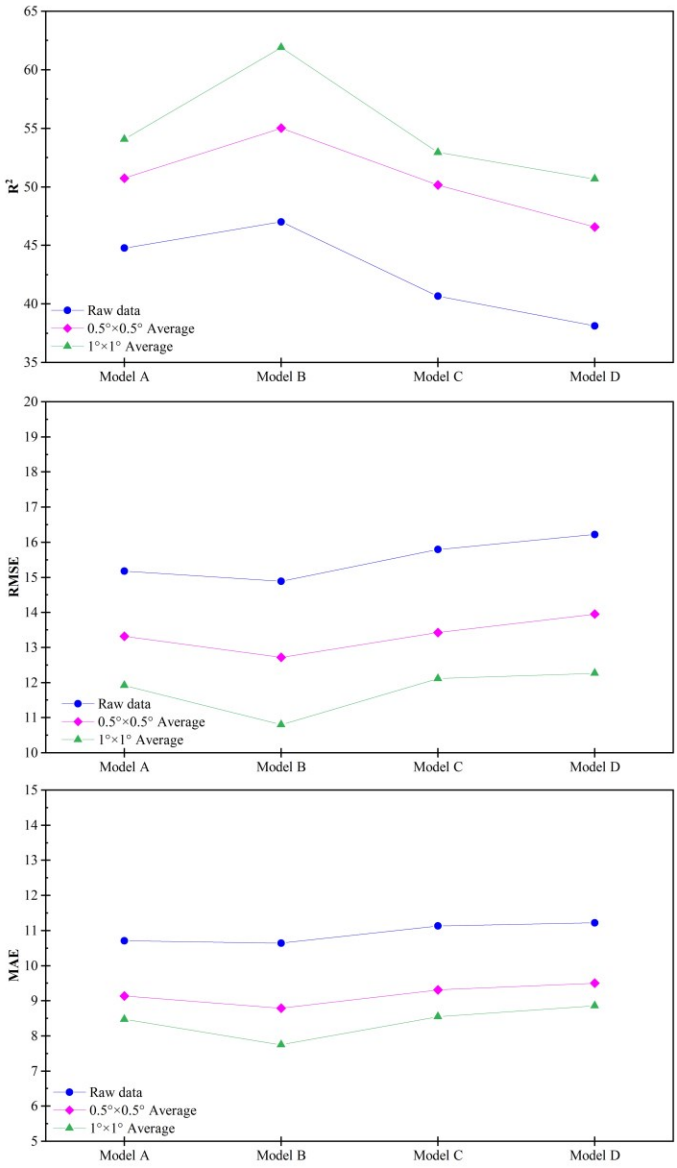


**Figure 11: Performance of different models in prediction GHF of China.**

### 4.4 Uncertainties caused by the feature selections

The performance of the model exhibits convergence behavior as the number of features increases, leading to models based on different numbers and combinations of features having similar $R^2$ scores and RMSE (see section 4.1). However, with the current data, we cannot determine the optimal number and combination of features. Thus, this section briefly discusses the uncertainty they bring to the prediction of GHF in China. We explore this uncertainty using the second dataset, which mainly takes into account that Model B performed best in the previous section. Four different feature sets were constructed, comprising nine, twelve, thirteen, and fourteen features, respectively.

Figure 12 shows the GHF prediction for different feature selections. Overall, they show similar distribution trends as the GHF predictions of Model B. Meanwhile, models based on various feature selections perform similarly in estimating the existing measured GHF in China (Fig. 13). Nevertheless, substantial discrepancies are still observed in heat flow predictions in many localized areas. At the boundary between the Tarim Craton and the Western Central Asian Orogen, the predicted GHF using nine and thirteen features (Fig. 12a and Fig. 12c) was significantly lower than that of Model B. In the northwest of the Tibet - Sanjiang Orogen, the model based on 12 features predicts a more extensive area with low GHF (Fig. 12b). In Central China Orogen, the model with 14 features predicts more regions with relatively high GHF (Fig. 12d). There are many local differences beyond those mentioned above. However, due to the lack of sufficient measured data, we are still unable to determine which feature selection could provide a more reasonable heat flow prediction. And that is also beyond the scope of the current study and may be the focus of future work.

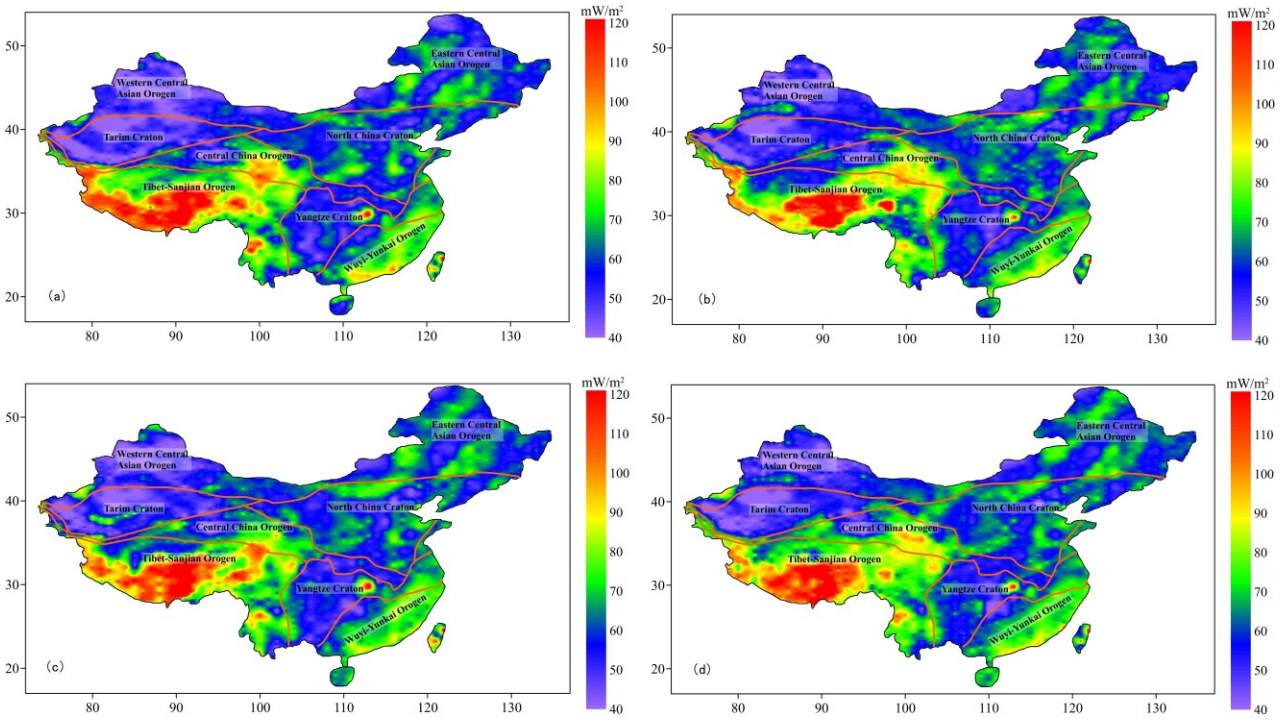


**Figure 12: Predicted heat flow models for different feature selections. (a) Only nine features (distance to ridge, distance to transform, distance to volcanoes, distance to young rift, bouguer gravity anomaly, crustal thickness, distance to trench, sedimentary thickness, and upper mantle density anomaly), (b) Twelve features (distance to young rift, density of active faults, heat production province, and LAB are not included), (c) Thirteen features (topography, isostatic gravity anomaly, and magnetic anomaly are not included), (d) Fourteen features (heat production province, and LAB are not included).**

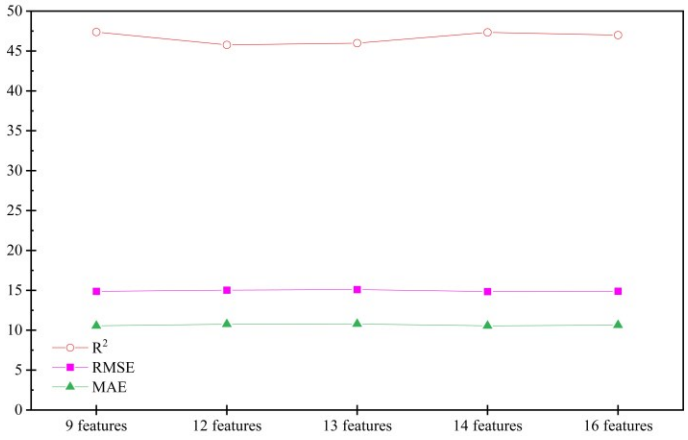


**Figure 13: Comparison of the performance of different feature selection models in predicting heat flow in China.**

### 4.5 Interpretation

The correlation between heat flow and different geological and geophysical features is the basis of predicting heat flow using the GBRT algorithm. We calculated the relative importance of the individual features on different models to interpret the GBRT predictions (Fig. 14). It is visible that the ranking of feature importance for the final GHF prediction (Fig. 14) is not consistent with the ranking of individual scores (Fig. 4). This indicates that the GBRT models establish a complex nonlinear relationship among the features rather than a simple linear regression. Model A and Model B predicted similar relative importance of features, and the same is true between Model C and Model D, which means that the averaging of heat flow will affect the judgment of the model on the importance of features.

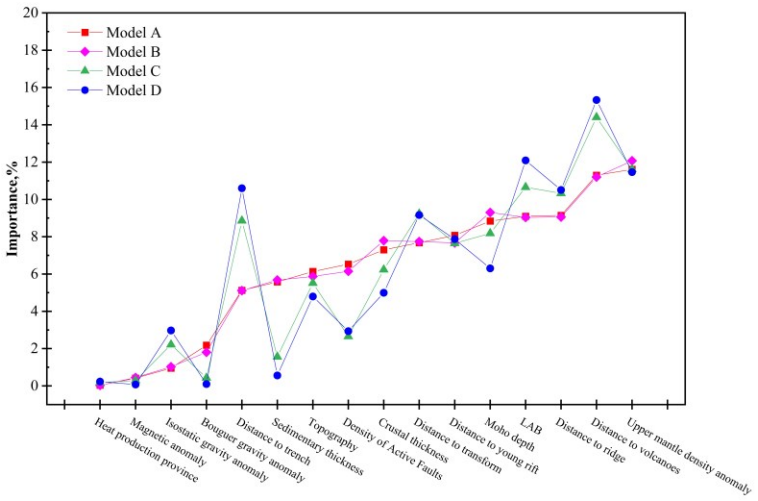


**Figure 14: Relative importance of the individual features (the sum of the relative importance is 1).**

Nonetheless, almost all of the models predict the distance to volcanoes and upper mantle density anomaly as the top two important features (Fig. 14). Figs. 15a and 15b show the variation characteristics of these two features in China. Most Eastern Central Asian Orogen regions show negative upper mantle density anomalies and small distances to volcanoes, indicating a high thermal state at depth in these areas. Meanwhile, seismic evidence also suggests the presence of mantle plumes, rifting and magmatic underplating (He et al., 2014), which may have contributed to the relatively high GHF in the area. However, the distribution characteristics of the high GHF area are inconsistent with those of the above two features, which may result from the combined effect of other features.

The Tarim Craton and the Western Central Asia Orogen have high positive upper mantle density anomalies (Fig. 15a), indicating a cold mantle thermal state. The Early Permian basalt eruption at about 277 Ma is the most recent major tectonic-thermal event in the Tarim craton (Yang et al., 2007). Moreover, it has been maintained in a stable tectonic setting since the Mesozoic (Liu et al., 2015), resulting in a relatively cold lithosphere. Furthermore, the relatively large Moho depth (Fig. 15c) and low elevation (Fig. 15d) of these two tectonic units also indicate a cold lithosphere, according to thermal isostasy (Hasterok & Chapman, 2007; Hyndman & Lewis, 1999). Combining these factors, our models predict low GHF for these areas, which is consistent with the measurements. In addition, our model predicts relatively high GHF regions at the boundary between the Tarim Craton and the Western Central Asian Orogen, which may result from the model considering higher elevations in these regions (Fig. 15d).

The Wuyi-Yunkai Orogen has a negative upper mantle density anomaly, a small distance to the volcano, a shallow Moho depth, and a relatively large elevation, all of which indicate a high deep thermal state in this area. And our models predict a high GHF region spreading northeast which is consistent with measured heat flow. In addition, it can be found that the GHF decreases in the NW direction, which is consistent with the upper mantle density anomaly, the distance to the volcano, and the Moho depth.

The Tibet-Sanjiang Orogen and the Central China Orogen are characterized by a thick crust and high elevation (Fig 15c and 15d). The large amount of radioactive heat generated by the extremely thick crust directly enhances the crustal heat flow in these areas (Jiang et al., 2019). It is visible that the predicted GHF in the Central China Orogen has a similar distribution trend to the Moho depth and elevation, and the high GHF corresponds to the thick crust. In addition, our model predicts a region with relatively high GHF at the junction of the Tarim Craton and the Central China Orogen, where no GHF measurements exist (Fig 10 a~c). This prediction may be due to the model's consideration of higher negative upper mantle density anomalies in the region (Fig 15a). However, for most areas, the GHF does not show a clear correspondence with the upper mantle density anomaly and the distance to the volcanoes, which may be the result of the model's comprehensive analysis of various features.

The Yangtze craton has an Archean-Proterozoic crystalline basement (Guo et al., 2014) and can be divided into three parts from west to east, including the upper Yangtze, middle Yangtze, and lower Yangtze (Lu et al., 2013). The upper mantle density anomaly is small in most of the Yangtze Craton. The Moho depth and the distances to volcanoes of the Yangtze Craton tend to decrease from the upper to the lower Yangtze, which leads to a higher GHF estimation for the lower Yangtze (Fig. 10). Meanwhile, it corresponds to the significant deformations in lower Yangtze (Lu et al., 2014), implying that these areas are no longer stable blocks(Li et al., 2015).

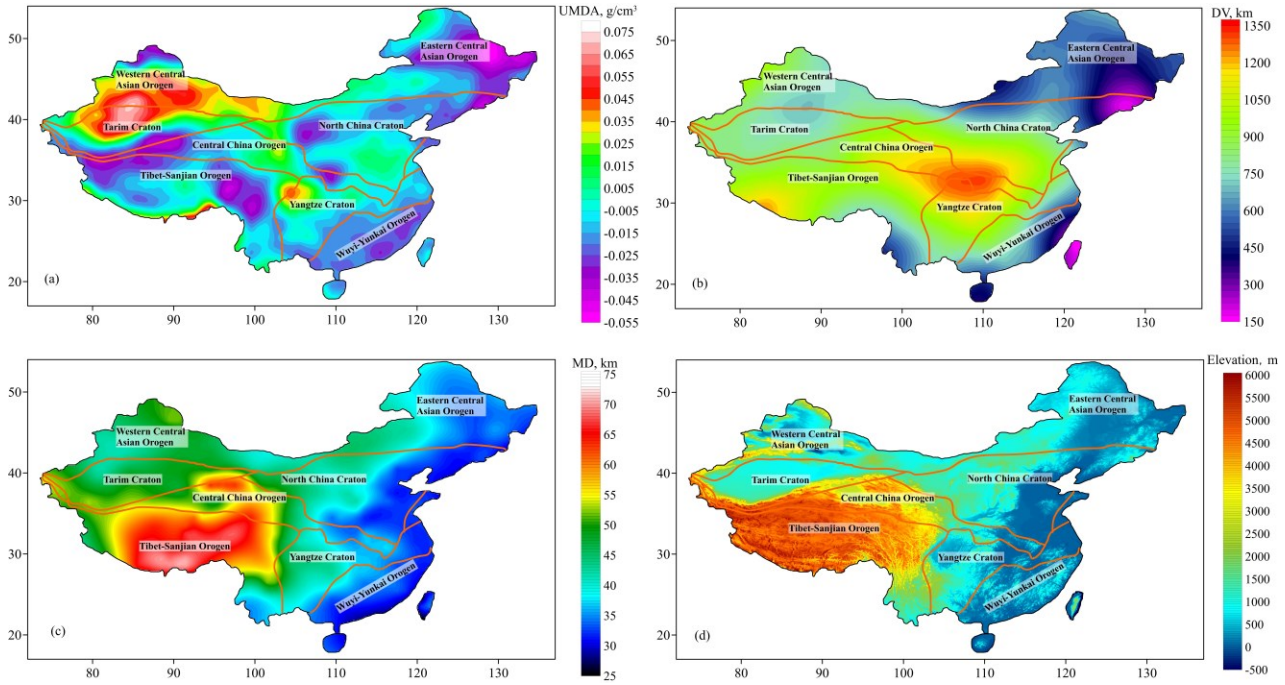


**Figure 15: Maps of (a) upper mantle density anomaly (UMDA), (b) distance to volcanoes (DV), (c) Moho depth (MD), and (d) elevation.**

## 5. DISCUSSION

In this study, we primarily investigated the influence of four distinct data preprocessing techniques on the prediction accuracy of GHF in China. Utilizing the GBRT algorithm, we found that the low-pass filtered dataset exhibited the highest predictive performance, while the dataset subjected to spatial averaging exhibited the poorest performance. In this section, we will discuss these results in greater detail, analyze the limitations of the study, and suggest directions for future research.

We observed an advantage in utilizing the dataset processed with a low-pass filter for predicting GHF in China. This may be attributed to the capacity of low-pass filters to accentuate low-frequency information within the data while mitigating the impact of high-frequency information. Low-frequency information likely reflects large-scale characteristics of geological and geophysical processes, whereas high-frequency information may represent localized variations. In estimating GHF, low-frequency information has a more significant influence on prediction outcomes, while high-frequency information is relatively less crucial. By emphasizing low-frequency information, the model can more readily identify and learn the primary features closely associated with GHF, thereby enhancing prediction accuracy.

Moreover, we observed that employing a dataset subjected merely to spatial averaging results in subpar performance in predicting GHF in China. This may be due to the considerable reduction in data volume caused by the spatial averaging operation, leading to the loss of important local information closely related to GHF. As this local information is vital for accurately estimating GHF, the model may struggle to fully capture these key details when handling the averaged dataset, resulting in diminished performance. Additionally, the averaged data may obscure some variations present in actual observations. This may render it challenging for the model to discern which features are genuinely useful during prediction, consequently affecting the model's performance. As the spatial averaging operation renders the relationships between features in the dataset more ambiguous, the model may struggle to accurately assess the importance of each feature for GHF prediction. This may cause the model to assign unduly low weights to certain crucial features, ultimately reducing prediction accuracy.

Although we explored four preprocessing techniques, it remains possible that other methods better suited to the objectives of this study exist. For instance, band-pass filtering, principal component analysis, and empirical mode decomposition may all yield varying degrees of improvement for GHF data. Future research could attempt to apply these techniques to GHF prediction, further enhancing prediction accuracy.

In conclusion, our study demonstrates the superiority of low-pass filtering in predicting GHF in China. Despite certain limitations, this method provides valuable insights for geothermal resource exploration and assessment. By further optimizing preprocessing methods, refining prediction models, and adequately considering a range of influential factors, future research may achieve more accurate and reliable results in the realm of GHF prediction.

## 6. CONCLUSIONS

In this study, we explore the impact of various data preprocessing techniques on the prediction of heat flow in China using GBRT models. Four distinct datasets, each derived from a different preprocessing method, were created: (i) unprocessed raw data, (ii) low-pass filtered data, (iii) spatially averaged data with a 0.5° × 0.5° grid resolution, and (iv) a combination of 1° × 1° grid averaging and low-pass filtering. Employing these datasets, we developed four GBRT models and generated eight new heat flow maps, comparing and analyzing the discrepancies in predictions. The main conclusions are summarized as follows:

(1) The GBRT model utilizing low-pass filtered data proves to be the most effective among the four models in estimating heat flow in China.

(2) Spatial averaging of datasets alters the assessment of feature relative importance in GBRT models, which may be attributable to the changes in data quantity induced by the averaging operation.

(3) The models constructed using spatially averaged datasets necessitate a greater number of features to achieve performance convergence. Furthermore, the impact of feature combinations on the performance of these models is more pronounced, suggesting that models established with smaller data quantities are more susceptible to the effects of feature combinations.

(4) All four GBRT models are more effective in predicting large-scale heat flow variations (0.5° × 0.5°) than in estimating short-distance heat flow variations.

(5) The heat flow maps predicted using the GBRT model with low-pass filtered data show that the overall heat flow values of Tarim Craton and Western Central Asian Orogen are the lowest, followed by the Yangtze Craton. The heat flow maps predict a large region with abnormally high heat flow over southwestern Tibet-Sangjing Orogen. Further, the maps show that the heat flow is high in the southwest of the Wuyi-Yunkai Orogen. Finally, the models predict several banded regions with relatively high heat flow in North China Craton and Western Central Asian Orogen. All of these predictions are consistent with the actual observations in China.

## REFERENCES

Amante, C., & Eakins, B. W. (2009). ETOPO1 arc-minute global relief model: procedures, data sources and analysis.

Balmino, G., Vales, N., Bonvalot, S., & Briais, A. (2012). Spherical harmonic modelling to ultra-high degree of Bouguer and isostatic anomalies. Journal of Geodesy, 86(7), 499-520.

Bodri, L., & Bodri, B. (1985). On the correlation between heat flow and crustal thickness. Tectonophysics, 120(1-2), 69-81.

Chandrasekharam, D., Baba, A., Ayzit, T., & Singh, H. K. (2022). Geothermal potential of granites: case study-Kaymaz and Sivrihisar (Eskisehir region) Western Anatolia. Renewable Energy, 196, 870-882.

Coffin, M. F., Gahagan, L. M., & Lawver, L. A. (1997). Present-day plate boundary digital data compilation.

Davies, J. H., & Davies, D. R. (2010). Earth's surface heat flux. Solid Earth, 1(1), 5-24.

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X., & Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agricultural and Forest Meteorology, 263, 225-241.

Goutorbe, B., Poort, J., Lucazeau, F., & Raillard, S. (2011). Global heat flow trends resolved from multiple geological and geophysical proxies. Geophysical Journal International, 187(3), 1405-1419.

Guo, J.-L., Gao, S., Wu, Y.-B., Li, M., Chen, K., Hu, Z.-C., Liang, Z.-W., Liu, Y.-S., Zhou, L., & Zong, K.-Q. (2014). 3.45 Ga granitic gneisses from the Yangtze Craton, South China: implications for Early Archean crustal growth. Precambrian Research, 242, 82-95.

Hasterok, D., & Chapman, D. S. (2007). Continental thermal isostasy: 1. Methods and sensitivity. Journal of Geophysical Research: Solid Earth, 112(B6).

He, C., Dong, S., Chen, X., Santosh, M., & Niu, S. (2014). Seismic evidence for plume-induced rifting in the Songliao Basin of Northeast China. Tectonophysics, 627, 171-181.

He, J., Li, K., Wang, X., Gao, N., Mao, X., & Jia, L. (2022). A Machine Learning Methodology for Predicting Geothermal Heat Flow in the Bohai Bay Basin, China. Natural Resources Research, 1-24.

Hyndman, R., & Lewis, T. (1999). Geophysical consequences of the Cordillera–Craton thermal transition in southwestern Canada. Tectonophysics, 306(3-4), 397-422.

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

Laske, G., Masters, G., Ma, Z., & Pasyanos, M. (2013). Update on CRUST1. 0—A 1-degree global model of Earth's crust. Geophys. res. abstr,

Li, C. F., Wang, J., Lin, J., & Wang, T. (2013). Thermal evolution of the North Atlantic lithosphere: new constraints from magnetic anomaly inversion with a fractal magnetization model. Geochemistry, Geophysics, Geosystems, 14(12), 5078-5105.

Li, X., Zhu, P., Kusky, T. M., Gu, Y., Peng, S., Yuan, Y., & Fu, J. (2015). Has the Yangtze craton lost its root? A comparison between the North China and Yangtze cratons. Tectonophysics, 655, 1-14.

Liu, S., Lei, X., & Wang, L. (2015). New heat flow determination in northern Tarim Craton, northwest China. Geophysical Journal International, 200(2), 1196-1206.

Lösing, M., & Ebbing, J. (2021). Predicting geothermal heat flow in Antarctica with a machine learning approach. Journal of Geophysical Research: Solid Earth, 126(6), e2020JB021499.

Lu, G., Zhao, L., Zheng, T., & Kaus, B. J. (2014). Strong intracontinental lithospheric deformation in South China: Implications from seismic observations and geodynamic modeling. Journal of Asian Earth Sciences, 86, 106-116.

Lu, J., Zheng, J., Griffin, W. L., & Yu, C. (2013). Petrology and geochemistry of peridotite xenoliths from the Lianshan region: Nature and evolution of lithospheric mantle beneath the lower Yangtze block. Gondwana Research, 23(1), 161-175.

Lucazeau, F. (2019). Analysis and mapping of an updated terrestrial heat flow data set. Geochemistry, Geophysics, Geosystems, 20(8), 4001-4024.

Mareschal, J.-C., Jaupart, C., & Iarotsky, L. (2017). The Earth heat budget, crustal radioactivity and mantle geoneutrinos. Neutrino geoscience, 4.1-4.46.

Maus, S., Barckhausen, U., Berkenbosch, H., Bournas, N., Brozena, J., Childers, V., Dostaler, F., Fairhead, J., Finn, C., & Von Frese, R. (2009). EMAG2: A 2–arc min resolution Earth Magnetic Anomaly Grid compiled from satellite, airborne, and marine magnetic measurements. Geochemistry, Geophysics, Geosystems, 10(8).

Pasyanos, M. E., Masters, T. G., Laske, G., & Ma, Z. (2014). LITHO1. 0: An updated crust and lithospheric model of the Earth. Journal of Geophysical Research: Solid Earth, 119(3), 2153-2173.

Rezvanbehbahani, S., Stearns, L. A., Kadivar, A., Walker, J. D., & van der Veen, C. J. (2017). Predicting the geothermal heat flux in Greenland: A machine learning approach. Geophysical research letters, 44(24), 12,271-212,279.

Shapiro, N. M., & Ritzwoller, M. H. (2004). Inferring surface heat flux distributions guided by a global seismic model: particular application to Antarctica. Earth and planetary science letters, 223(1-2), 213-224.

Styron, R., & Pagani, M. (2020). The GEM global active faults database. Earthquake Spectra, 36(1_suppl), 160-180.

Szwillus, W., Afonso, J. C., Ebbing, J., & Mooney, W. D. (2019). Global crustal thickness and velocity structure from geostatistical analysis of seismic data. Journal of Geophysical Research: Solid Earth, 124(2), 1626-1652.

Wang, J., & Li, C.-F. (2018). Curie point depths in Northeast China and their geothermal implications for the Songliao Basin. Journal of Asian Earth Sciences, 163, 177-193.

Yang, S.-F., Li, Z., Chen, H., Santosh, M., Dong, C.-W., & Yu, X. (2007). Permian bimodal dyke of Tarim Basin, NW China: geochemical characteristics and tectonic implications. Gondwana Research, 12(1-2), 113-120.