



河南省首届地热产业高质量发展学术交流专题报告会

以科技创新驱动我省地热产业高质量发展

## 报告题目：预测地热甜点的人工智能方法



李克文，博士生导师，石油天然气、地热能开发以及热光伏发电国际知名专家，历任美国斯坦福大学能源资源工程系资深研究员、北京大学教授、长江大学“楚天学者”特聘教授、现任中国地质大学（北京）教授。曾任SPE REE杂志（为石油工程领域国际权威杂志）副主编、国际地热协会教育委员会委员能源行业地热能标准化技术委员会委员以及SPEJ杂志等多个国际专业期刊的技术编辑等职务。公开发表SCI论文90多篇，专著3本，拥有10多项发明专利。李克文教授是国际上从事气润湿反转技术提高采收率的学术创始人、建立了基于流体流动规律的产量预测新方法，有关公式以李克文教授的名字命名。

# 预测地热甜点的人工智能方法

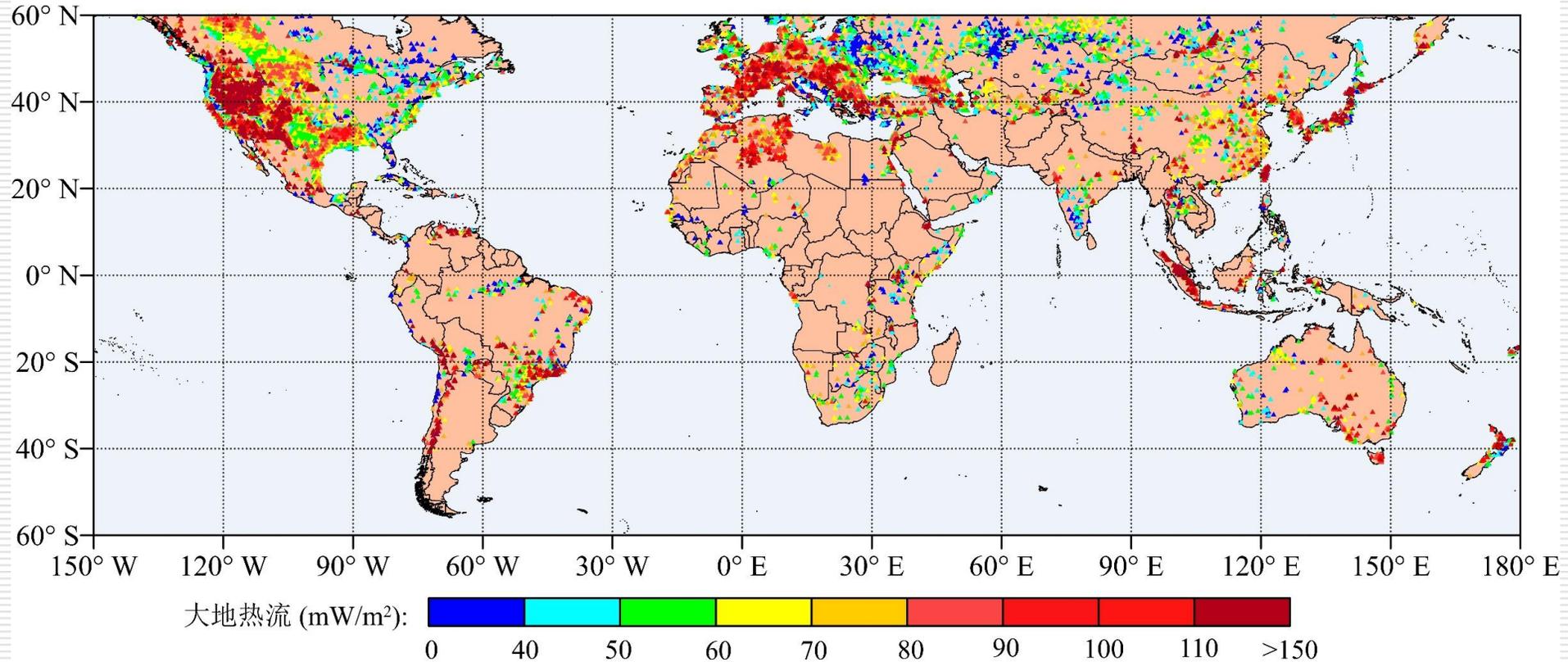
李克文

中国地质大学（北京）

2024年12月10日

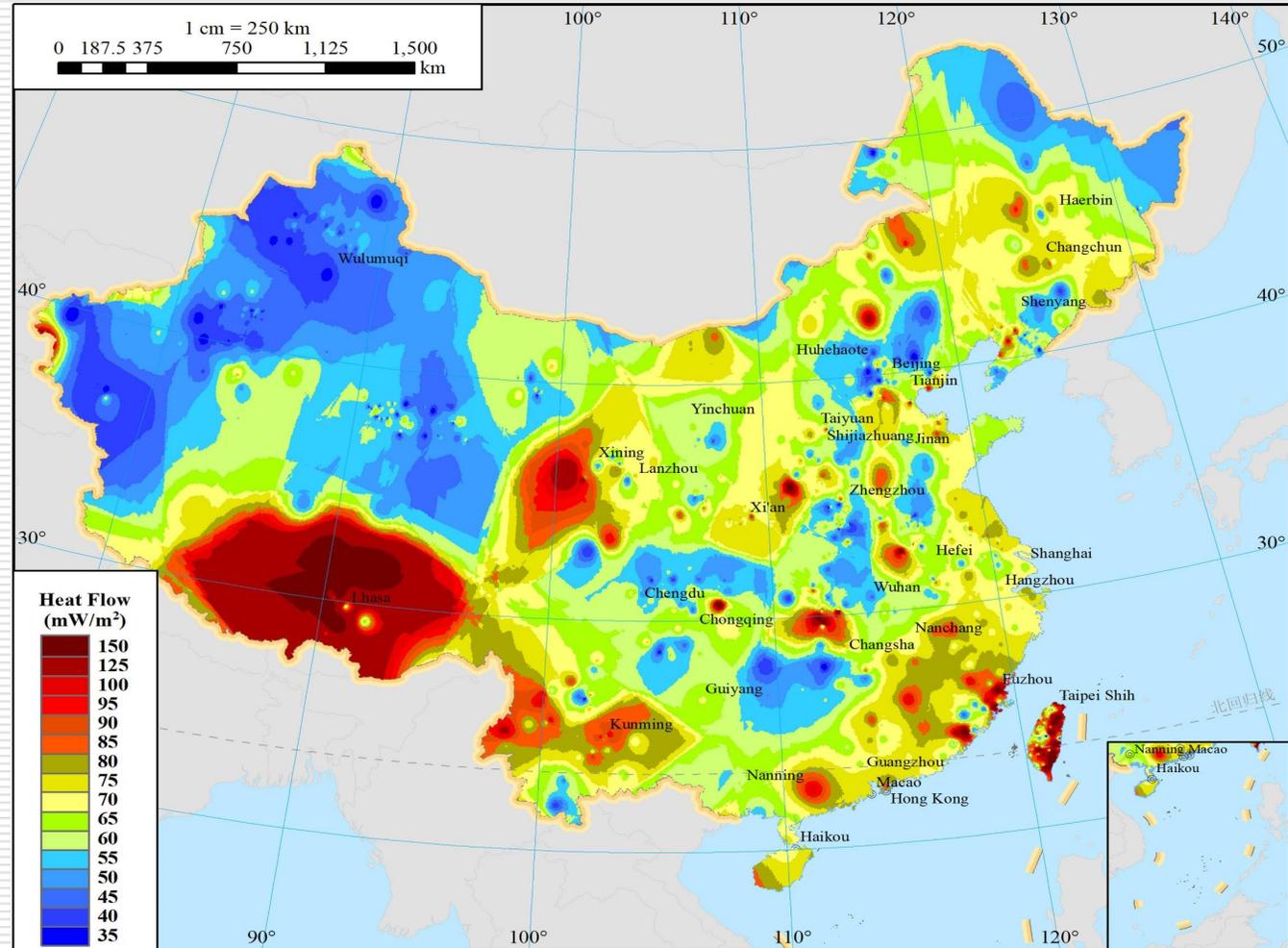
# 背景及意义

- **大地热流**是地热勘探开发的**关键指标**之一；
- 实测热流**数据难度大，时间和经济成本高**；
- 包含中国在内的全球很多地区**实测热流数据稀疏**；
- 探索经济有效的大地热流预测方法是非常必要的。



# 简单插值法

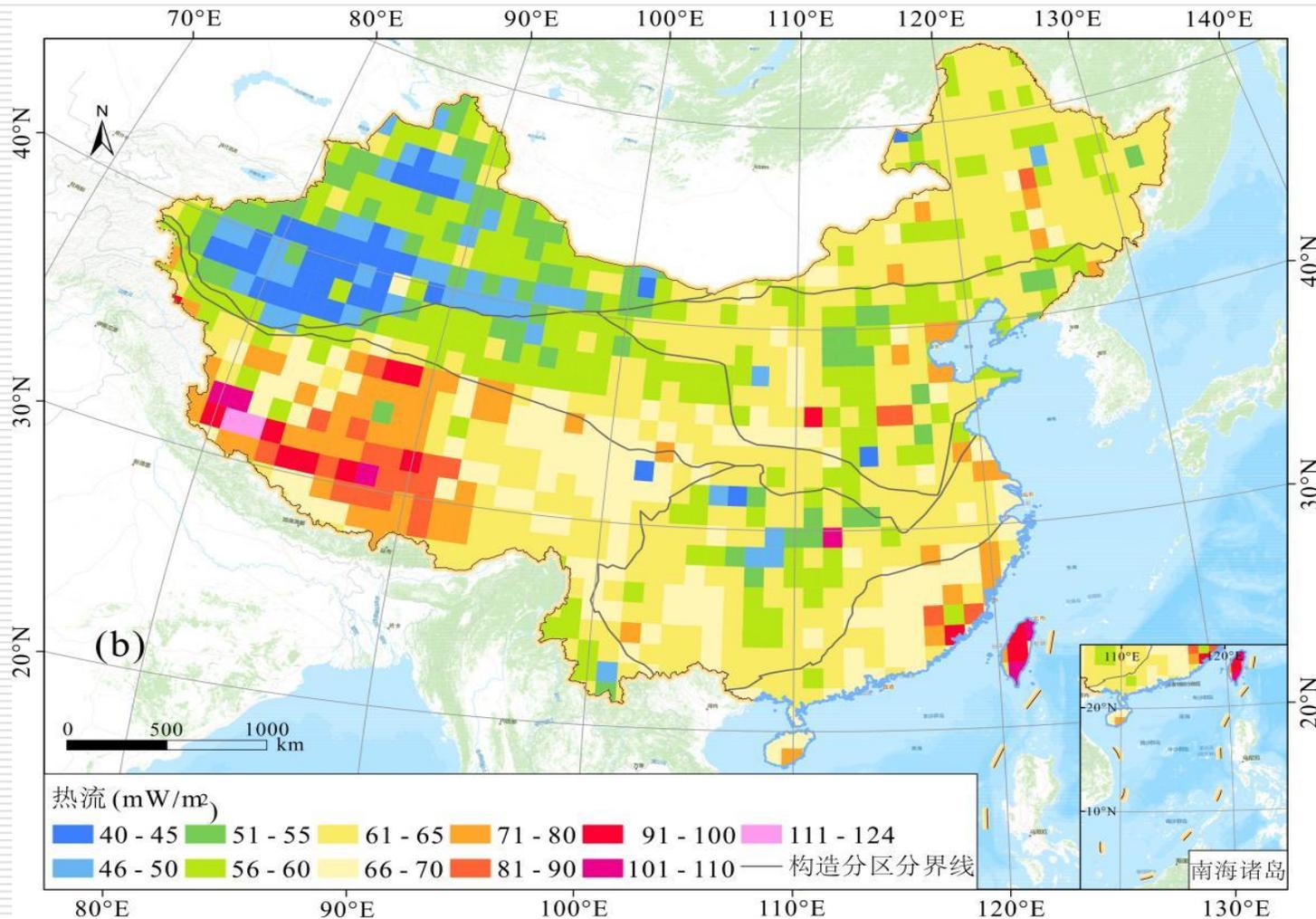
- 适合于数据点相对密集的区域，缺乏足够实测数据时，无法有效捕捉到地质变异性。



(马永生等, 2023)

# 经验法

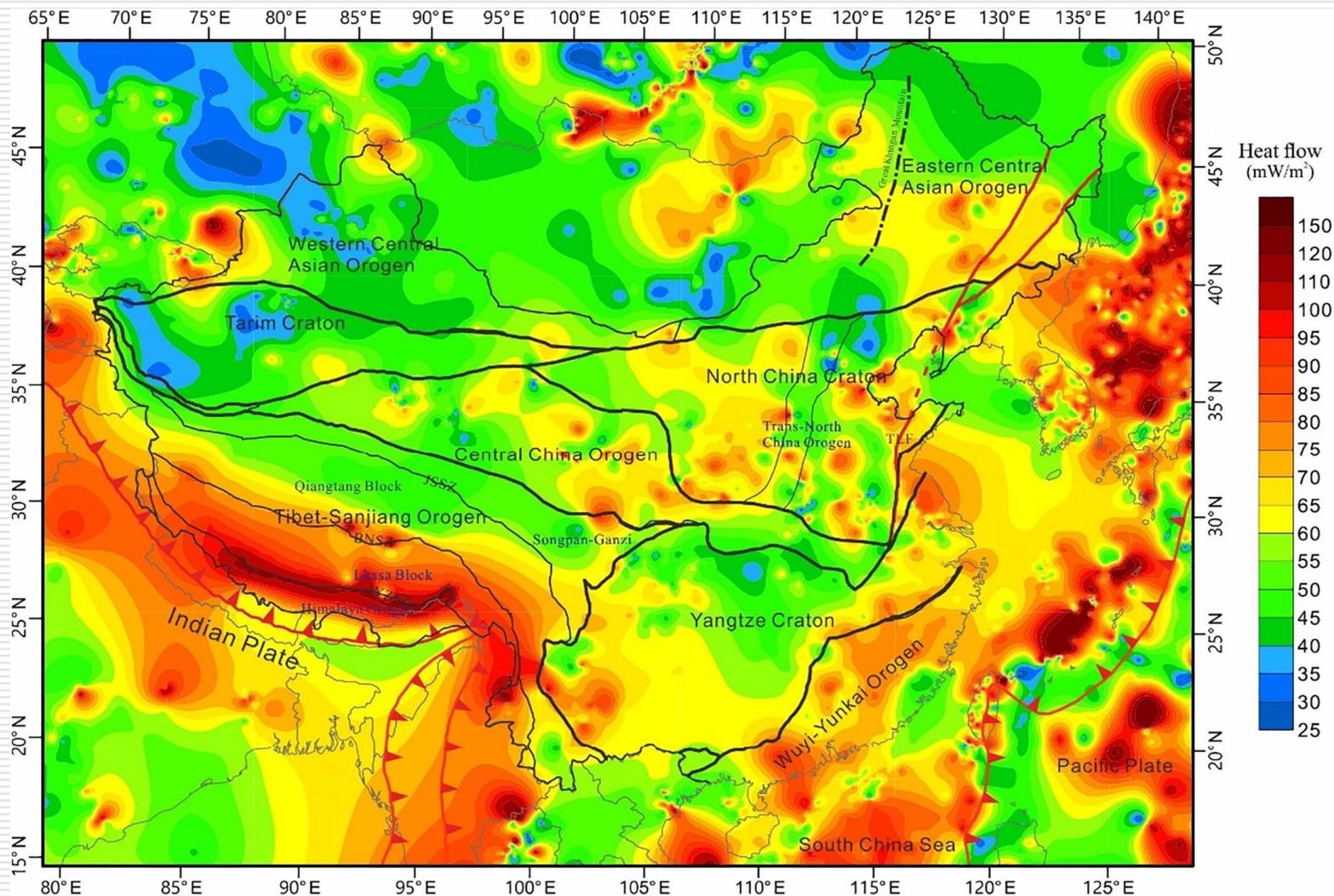
- 通常只考虑了影响热流的部分因素，操作复杂、灵活性差、准确性低。



(王亚茹等, 2022)

# 简单插值法+经验法

- 在藏南人工控制热流点，之后插值生成热流图。

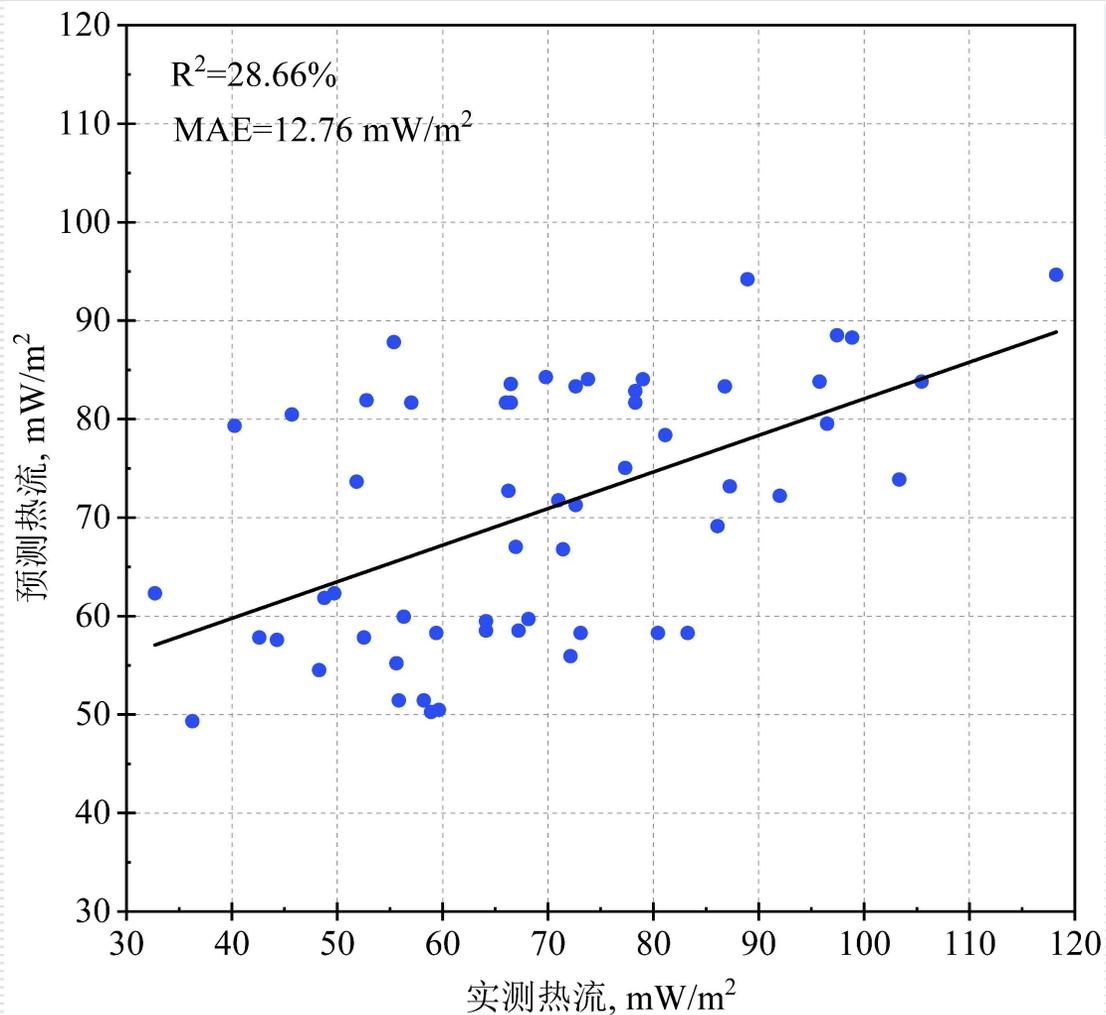


(Jiang等, 2019)

# 解析方法

➤ 只考虑了影响热流的部分因素，仅依靠几类地质数据的解析法鲁棒性较差。

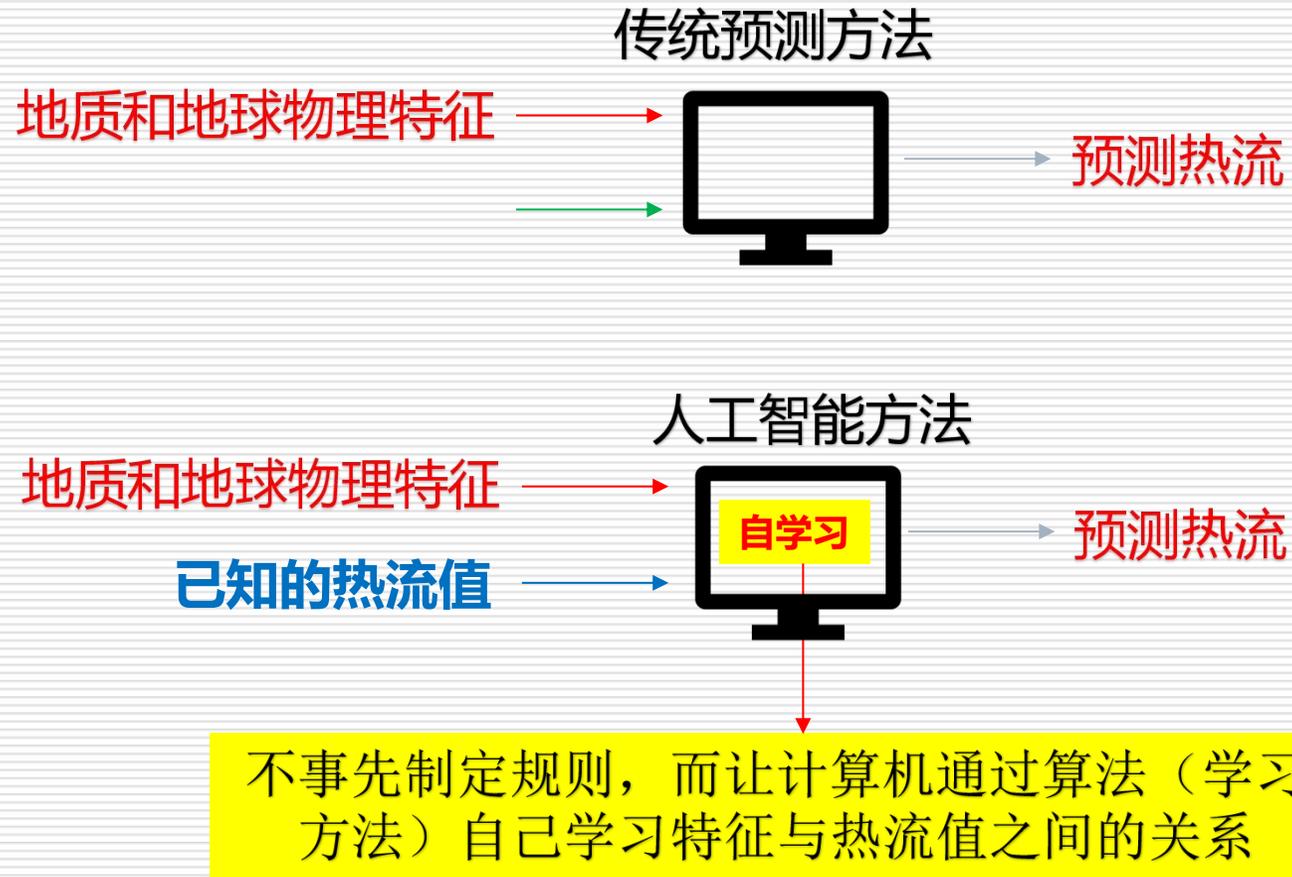
解析法预测青藏高原东南缘实测热流（数据来自唐晗晗等，2020）



$$q_s = -\frac{K(T_c - T_0)}{Z_b - Z_0} - H_0 h_r + \frac{H_0 h_r^2}{Z_b - Z_0} (1 - e^{-(Z_b - Z_0)/h_r})$$

- $q_s$  地表热流, mW/m<sup>2</sup>;
- $K$  地层热导率, W/m/K;
- $Z_b$  居里点深度, m;
- $Z_0$  地表高程, m;
- $H_0$  地表产热,  $\mu\text{W}/\text{m}^3$ ;
- $T_0$  地表温度,  $^{\circ}\text{C}$ ;
- $h_r$  放射性元素层厚度, m;
- $T_c$  居里点温度,  $^{\circ}\text{C}$ ;

# 人工智能方法与传统方法的区别

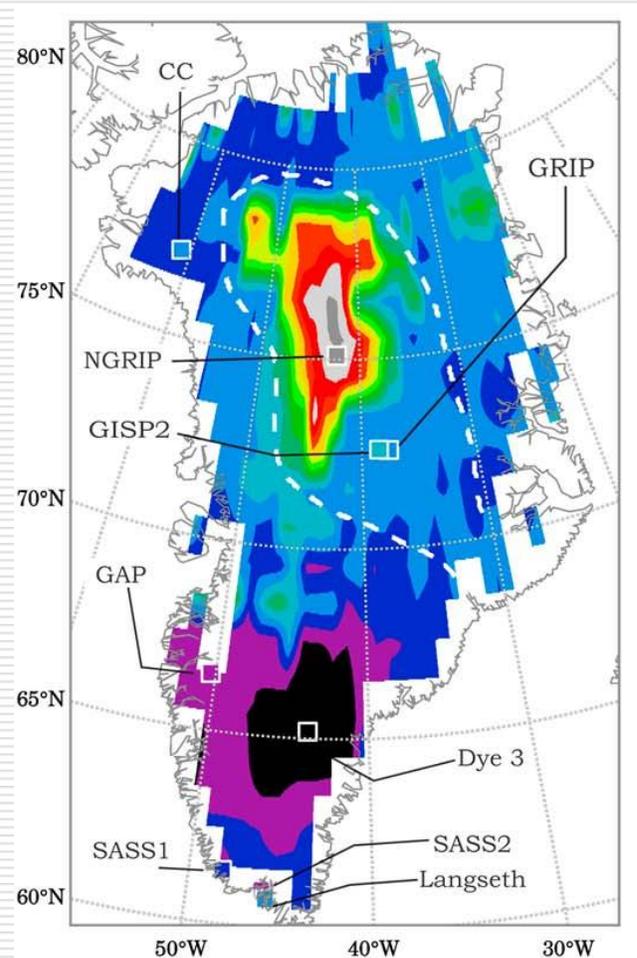
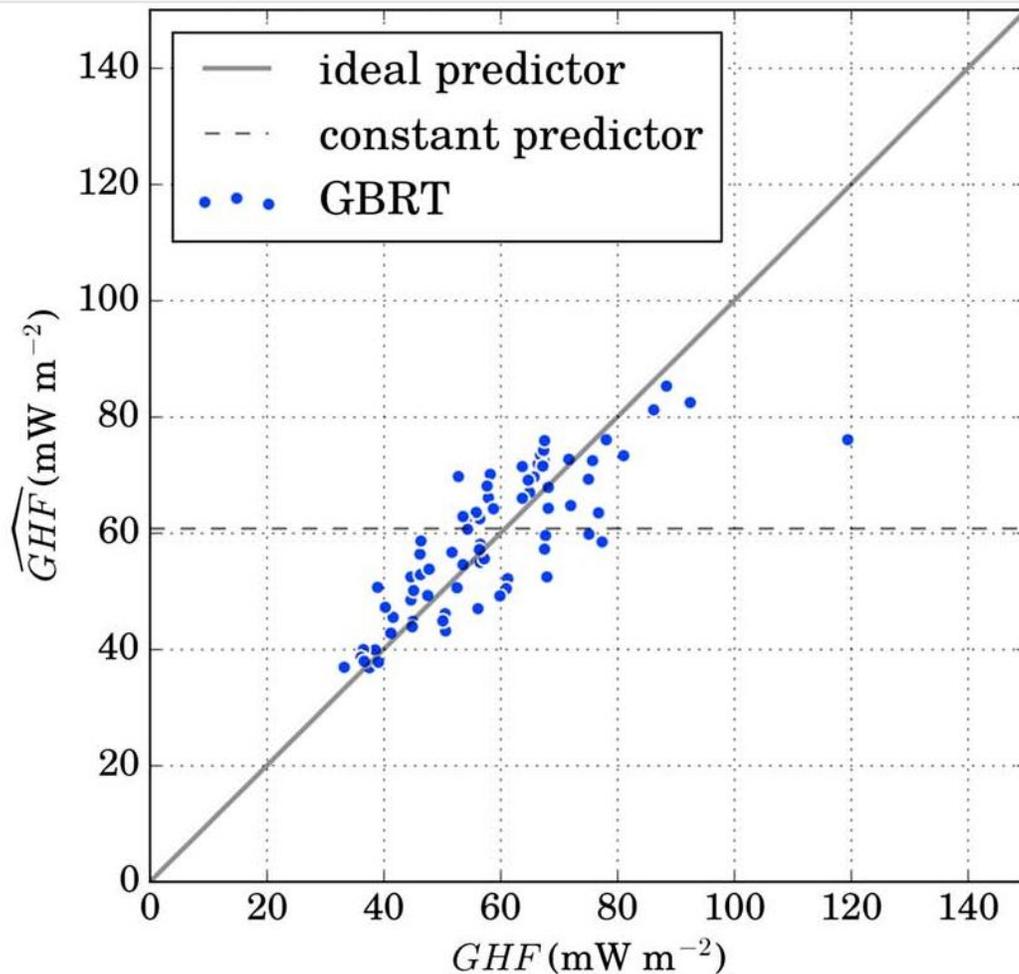


# 人工智能预测方法的优势



# 早期人工智能热流预测模型

- Rezvanbehbahani等人（2017）提出使用梯度提升回归树算法预测格陵兰热流；
- 预测精度： $R^2=67\%$ ， $N\text{-RMSE}=0.15$ 。



# 存在问题

缺少预处理方法优选

目前的研究基于不同的数据预处理方法，但尚缺乏优选研究。深入探究这一问题可以为构建更准确的预测模型提供指导；

缺少对全球和局部数据集的对比

目前缺乏使用不同数据集在同一地区预测的对比分析，这一研究可以帮助理解不同数据集对预测性能的影响，从而选择最合适的数据集来构建模型；

仅依赖于单一算法和模型

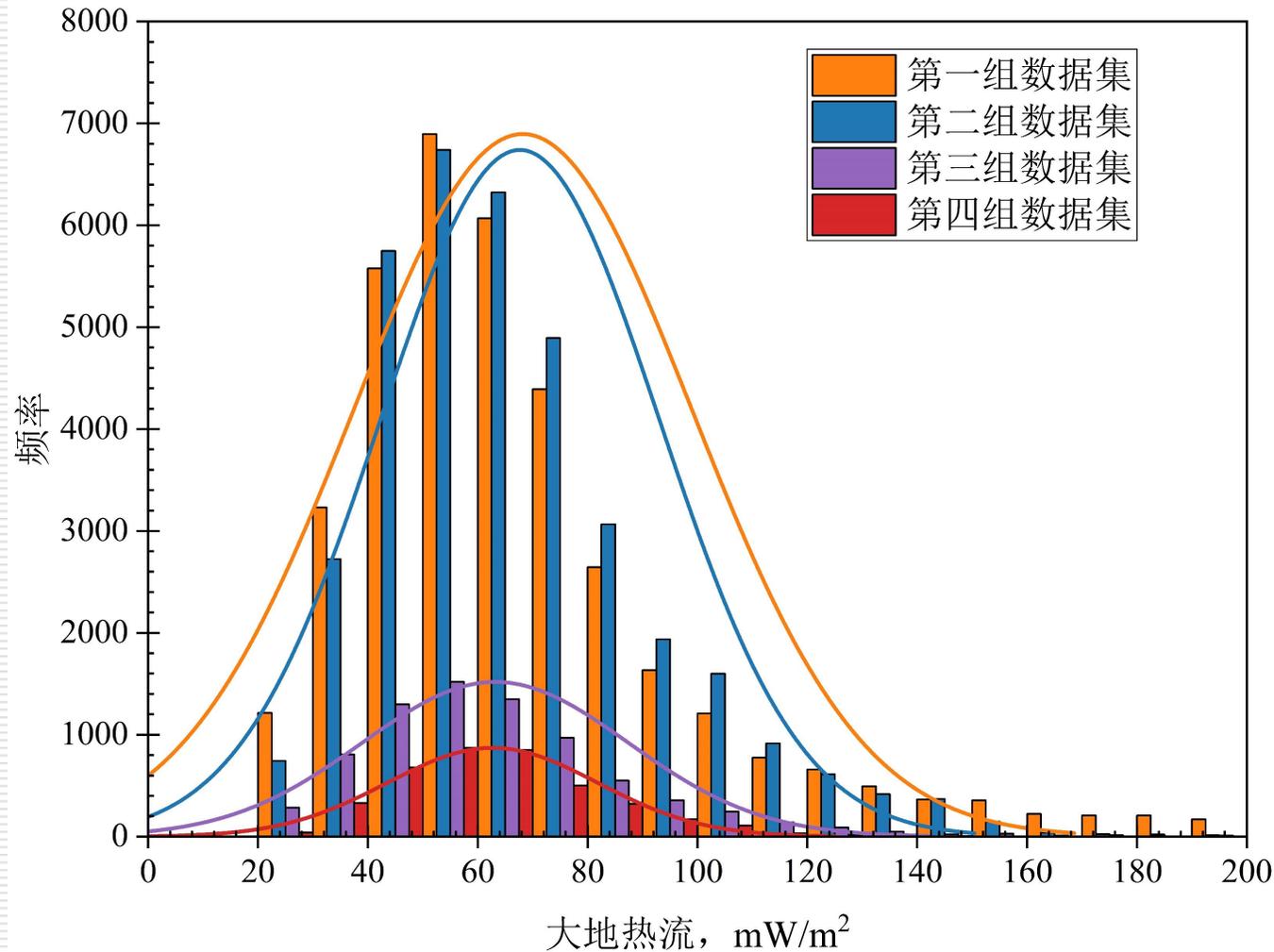
目前研究主要依赖于单一算法，但并未有算法在预测准确性上具有明显的优势。如何进一步提高模型的性能仍是有待解决的问题；

# 构建用于热流预测的全球数据集

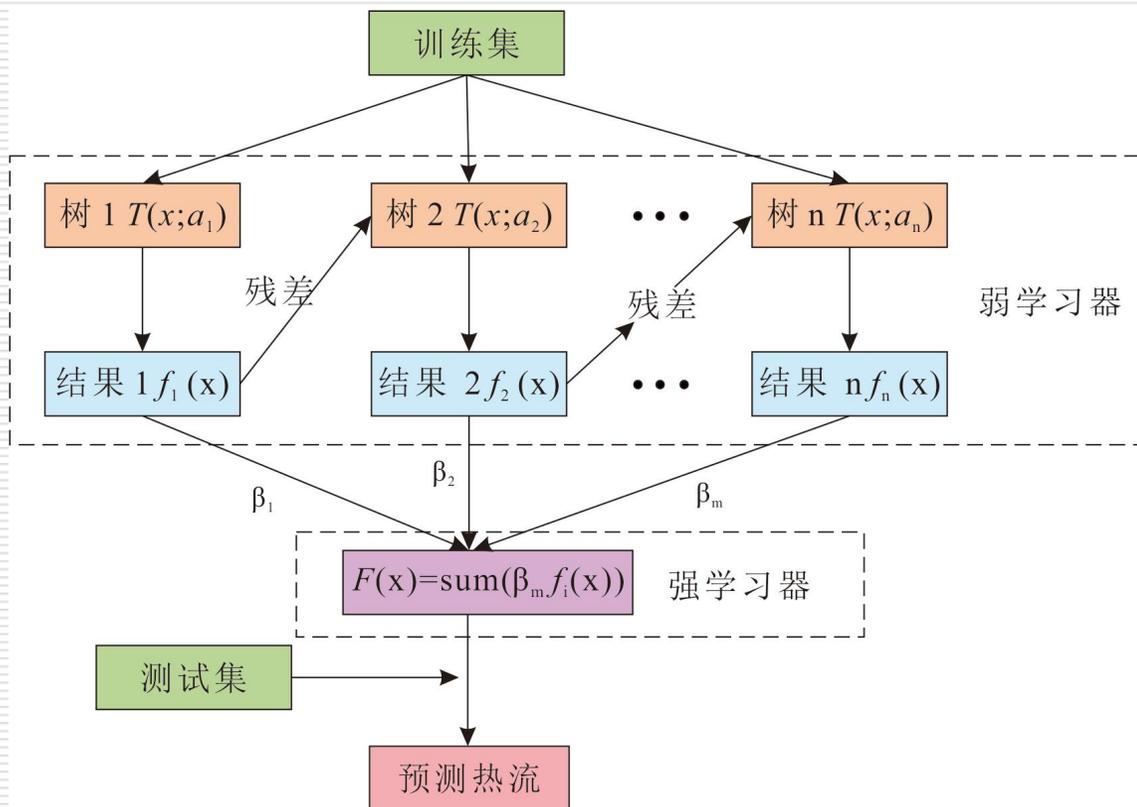
	特征	参考文献
1	莫霍面埋深	Szwilius 等人 <sup>[121]</sup>
2	地壳厚度	Laske 等人 <sup>[127]</sup>
3	沉积厚度	Laske 等人 <sup>[127]</sup>
4	磁异常	Maus 等人 <sup>[122]</sup>
5	地形	Amante & Eakins <sup>[128]</sup>
6	布格重力异常	Balmino 等人 <sup>[123]</sup>
7	均衡重力异常	Balmino 等人 <sup>[123]</sup>
8	岩石圈-软流圈边界深度	Pasyanos等人 <sup>[124]</sup>
9	产热	Goutorbe等人 <sup>[39]</sup>
10	到洋脊的距离	Coffin等人 <sup>[126]</sup>
11	到海沟的距离	Coffin等人 <sup>[126]</sup>
12	到转换边界的距离	Coffin等人 <sup>[126]</sup>
13	到年轻裂谷的距离	Goutorbe等人 <sup>[39]</sup>
14	到火山的距离	Goutorbe等人 <sup>[39]</sup>
15	上地幔密度异常	Kaban等人 <sup>[125]</sup>
16	活动断层密度	根据 Styron & Pagani <sup>[129]</sup> 计算
17	居里面深度	Li 等人 <sup>[71]</sup>

# 确定最优的数据预处理方法

- 第一组：36325个数据，原始数据；
- 第二组：36325个数据，低通滤波；
- 第三组：7740个数据， $0.5^\circ$  网格平均；
- 第四组：3939个数据， $1^\circ$  网格平均后低通滤波。



# GBRT算法和评价指标



- 在处理较大数据集时，速度较快；
- 调参过程相较于深度学习算法简单；
- 在之前热流预测的研究中表现出了较好的性能；

- 以80%和20%的比例划分为训练集和测试集；
- 在训练过程中还采用了4折交叉验证；
- 网格搜索方法进行优化，确定回归树的数量、树的深度和学习率最优组合。

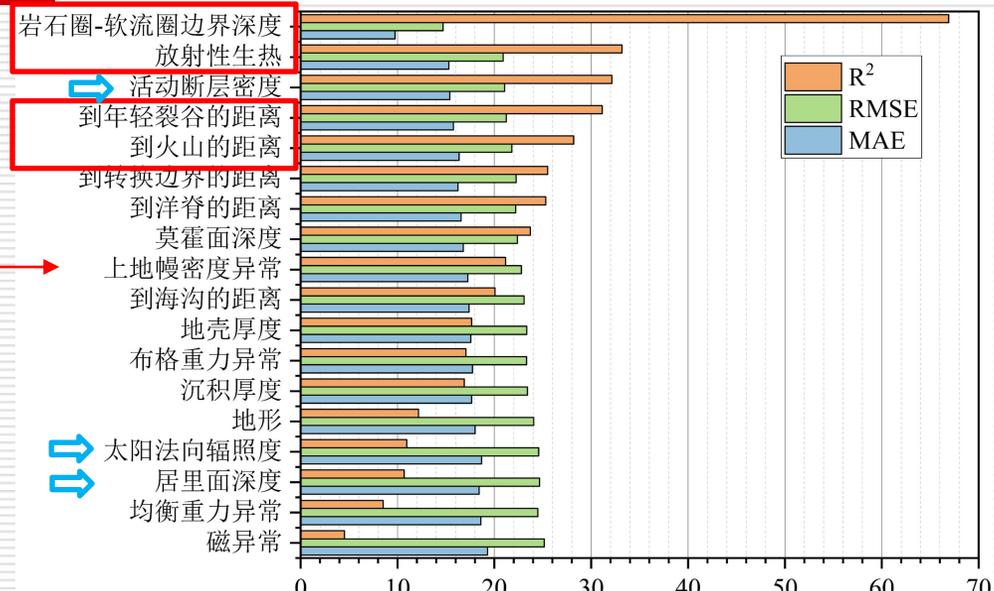
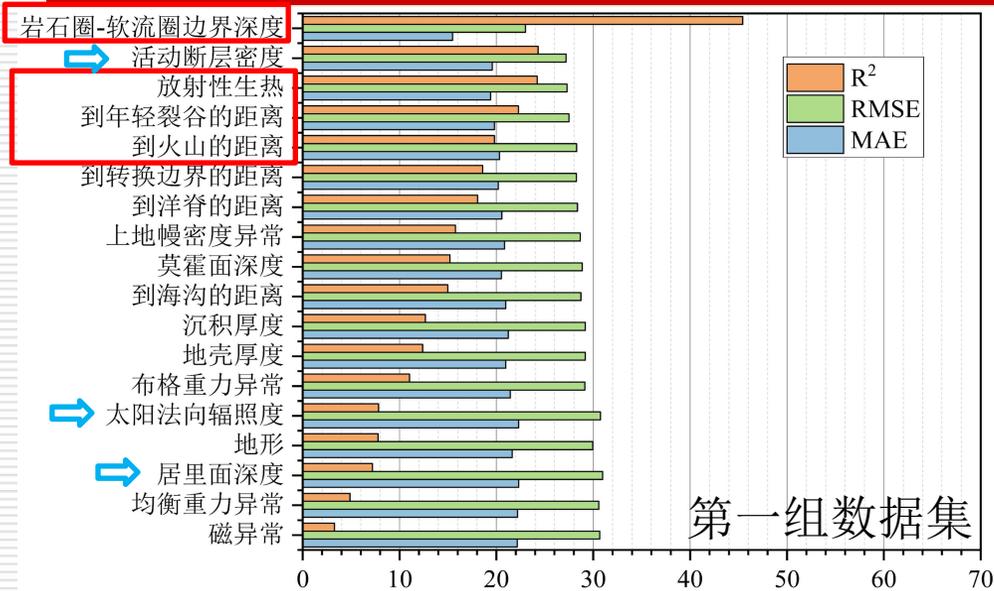
评价指标

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

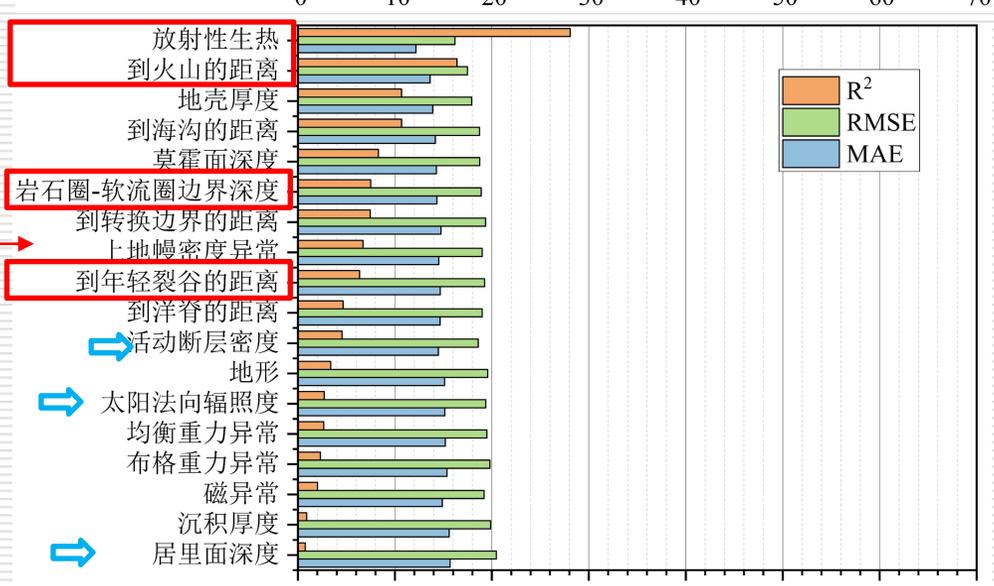
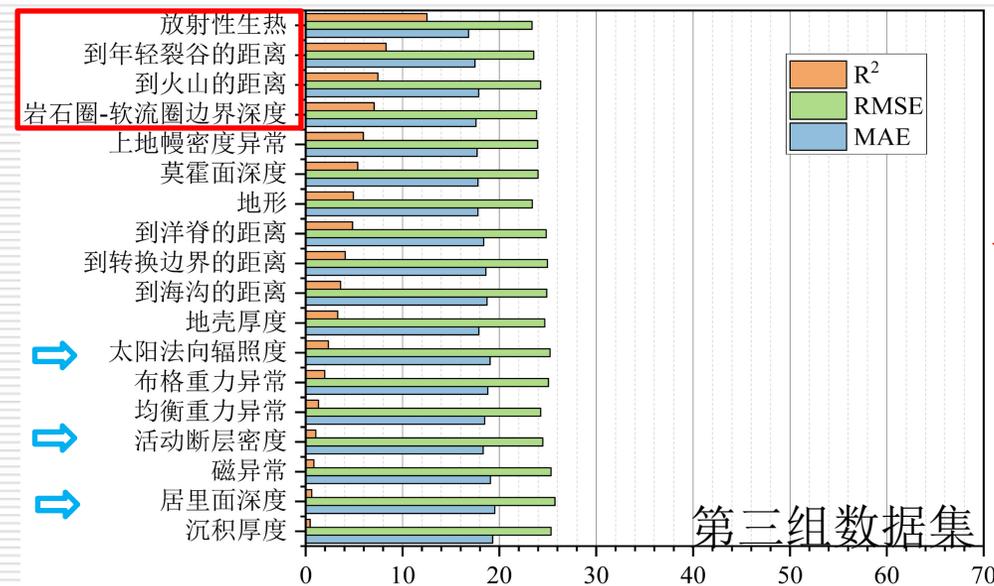
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - Y_i)^2}$$

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

# 单特征预测热流能力测试



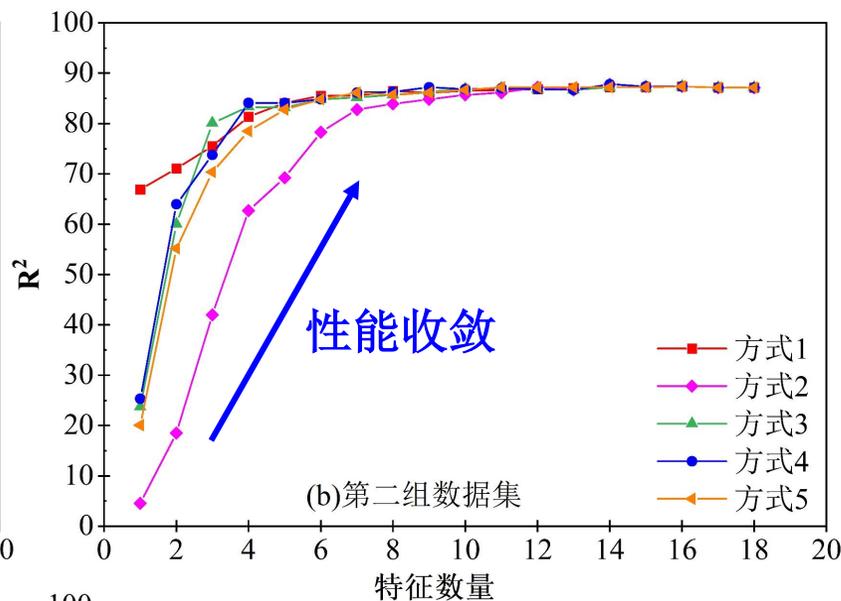
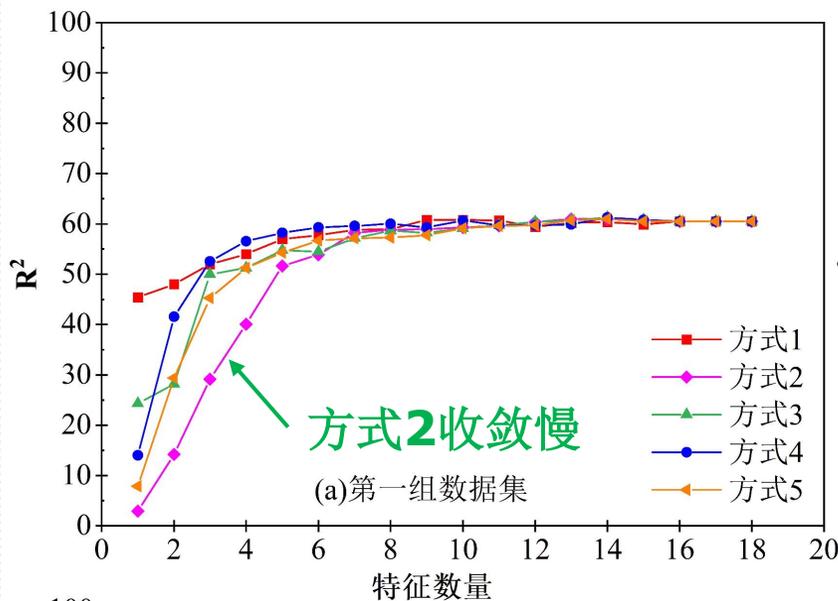
排名相似



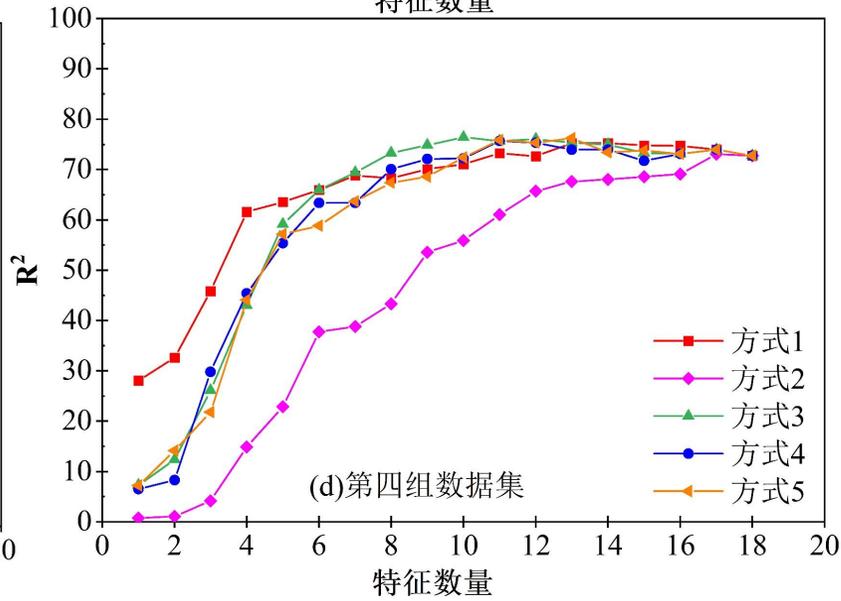
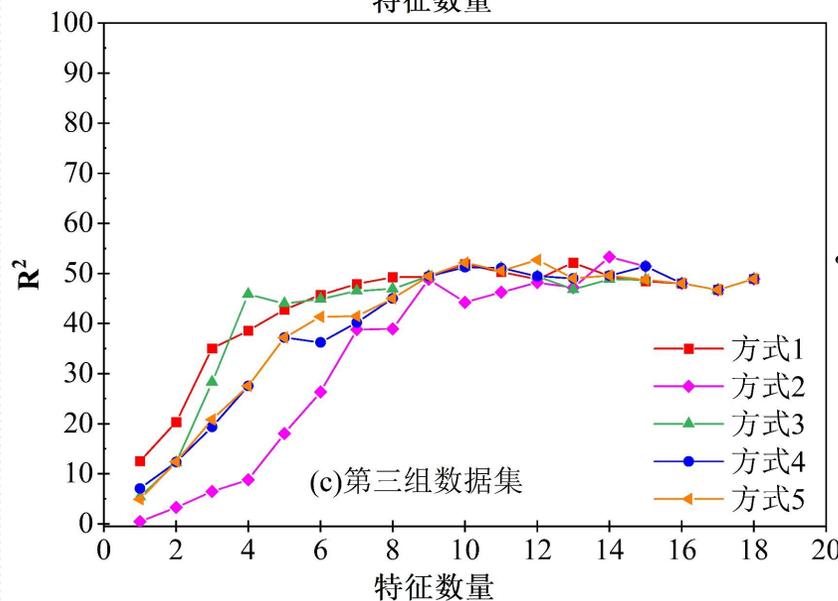
排名不同

# 特征数量和组合对模型性能的影响

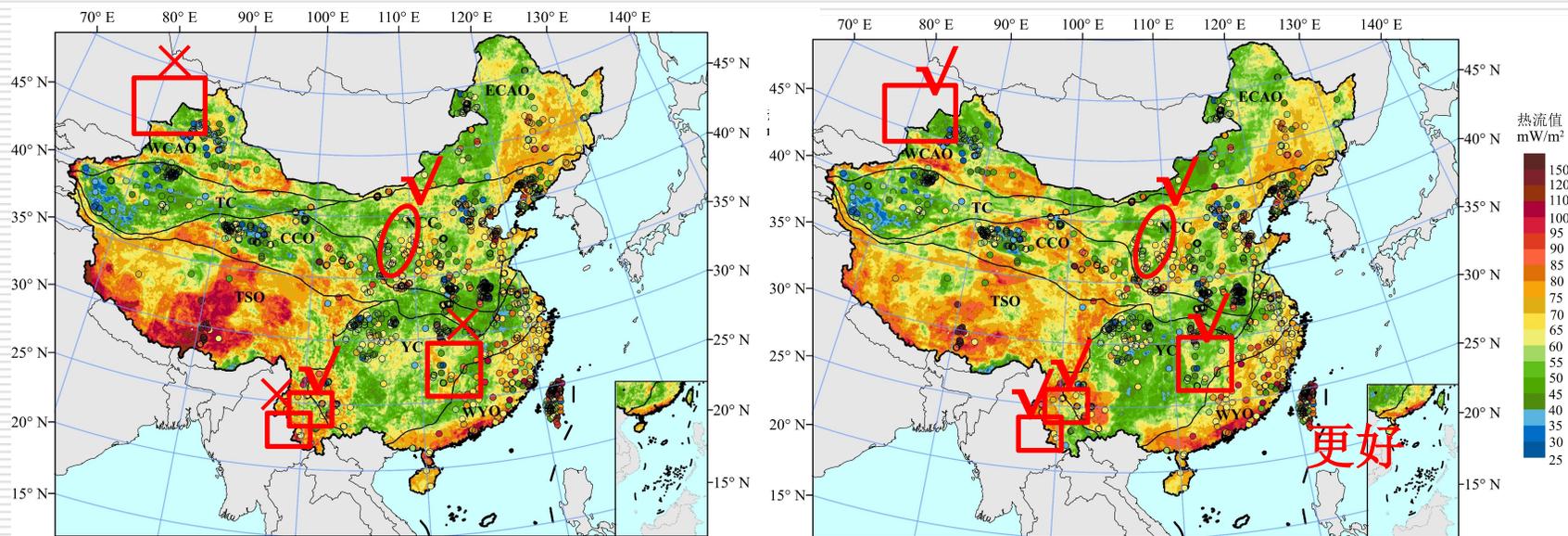
收敛快 →



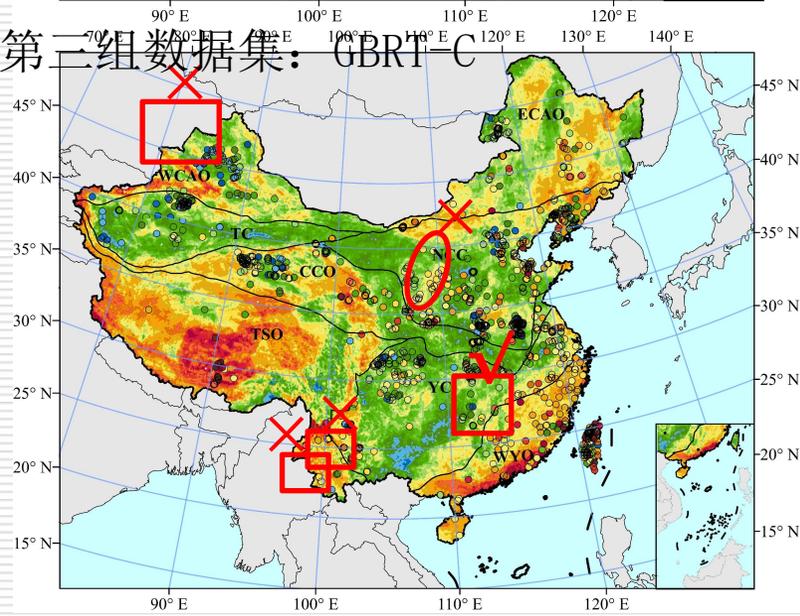
收敛慢 →



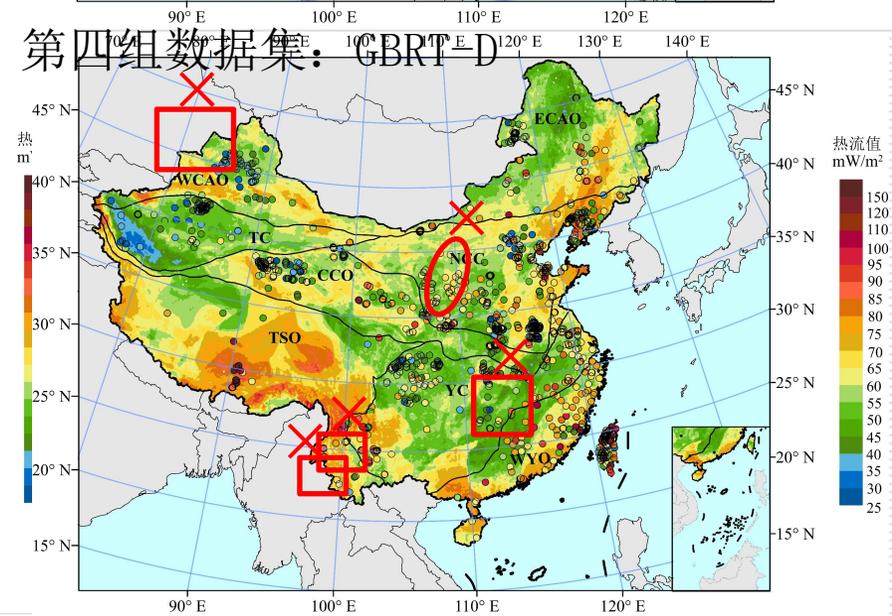
# 不使用中国数据的泛化能力测试



第二组数据集: GBRT-C

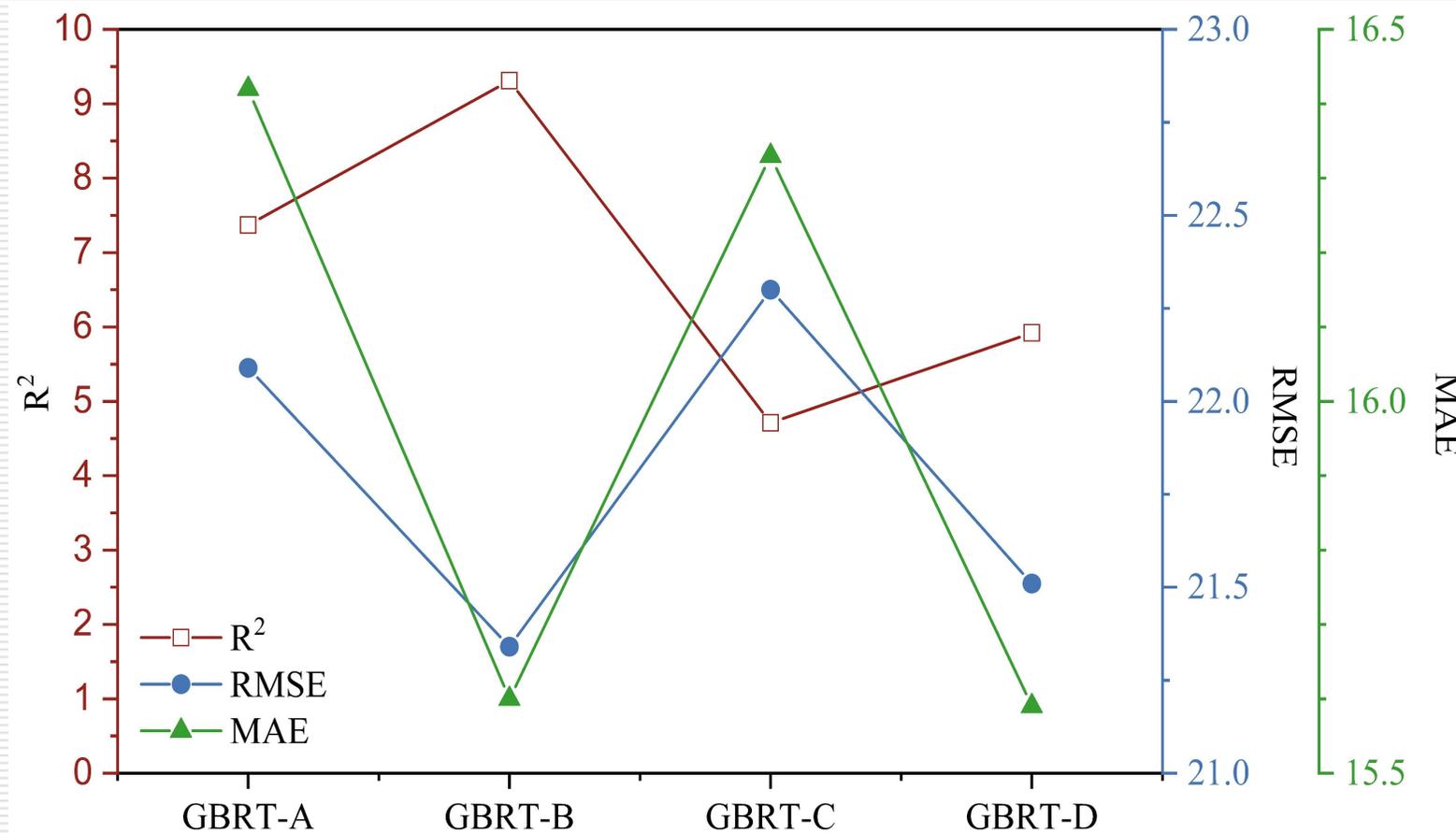


第四组数据集: GBRT-D



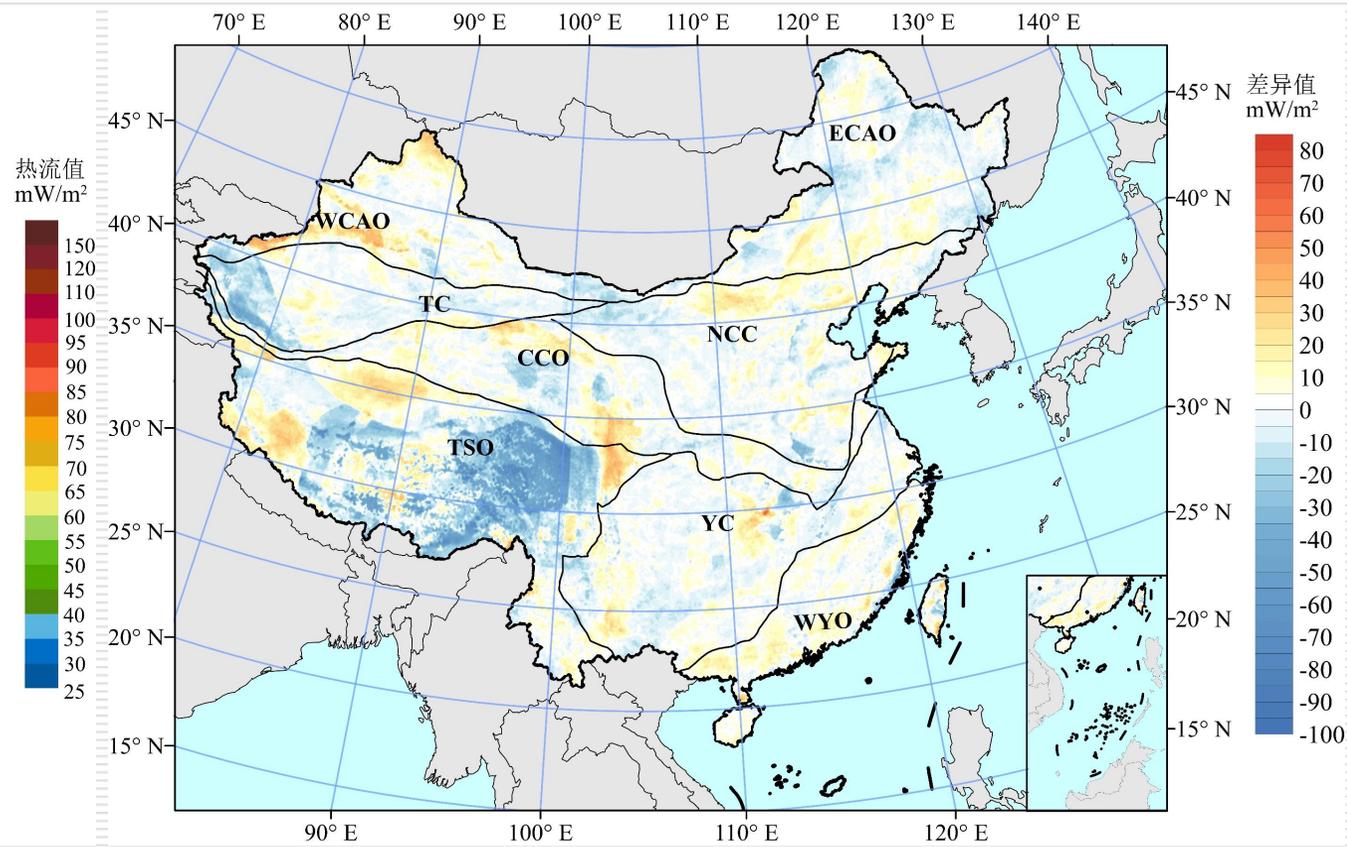
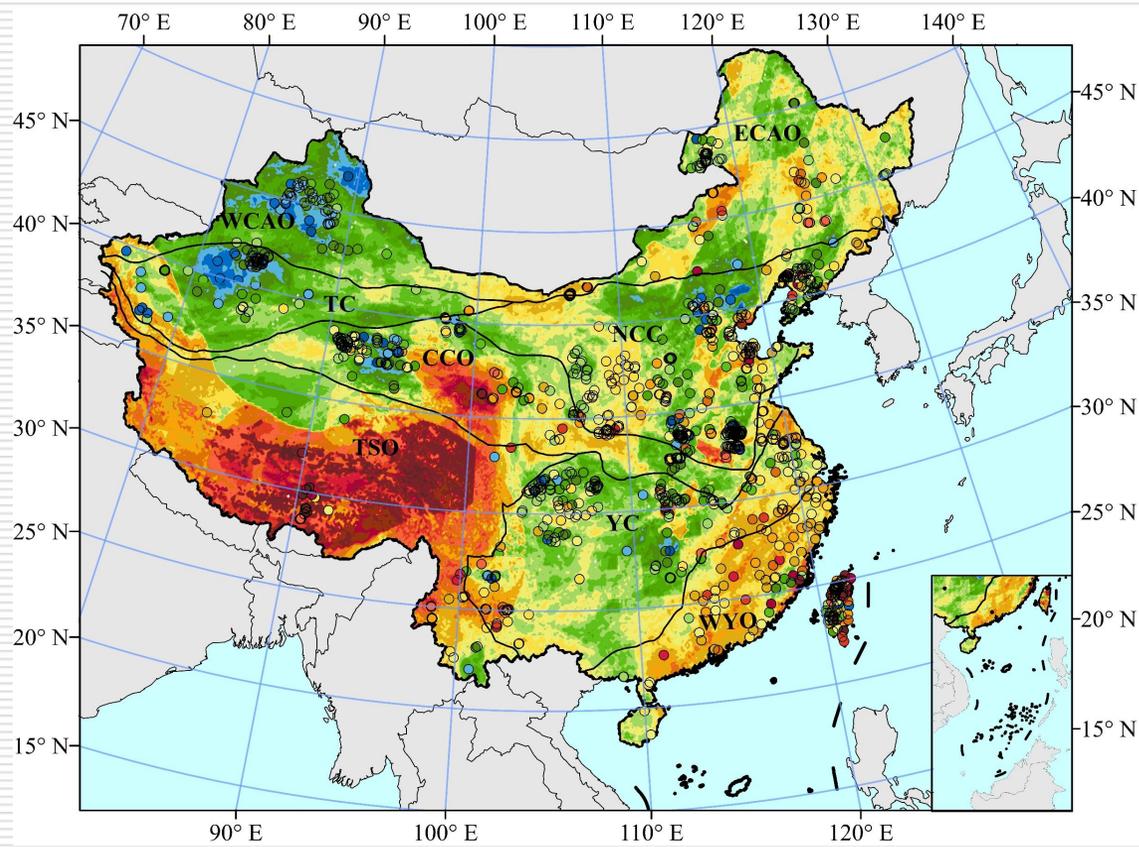
# 基于低通滤波处理的第二组数据集的模型性能更好

- 网格平均值可能无法反映真实的热流水平，将实测点分配到更大单元中有可能掩盖一些细节；
- 网格平均会使模型预测热流值范围收窄，不能很好的突出高值区和低值区；
- 在对中国大地热流预测中，基于仅采用低通滤波处理的模型展现了更好的预测性能。

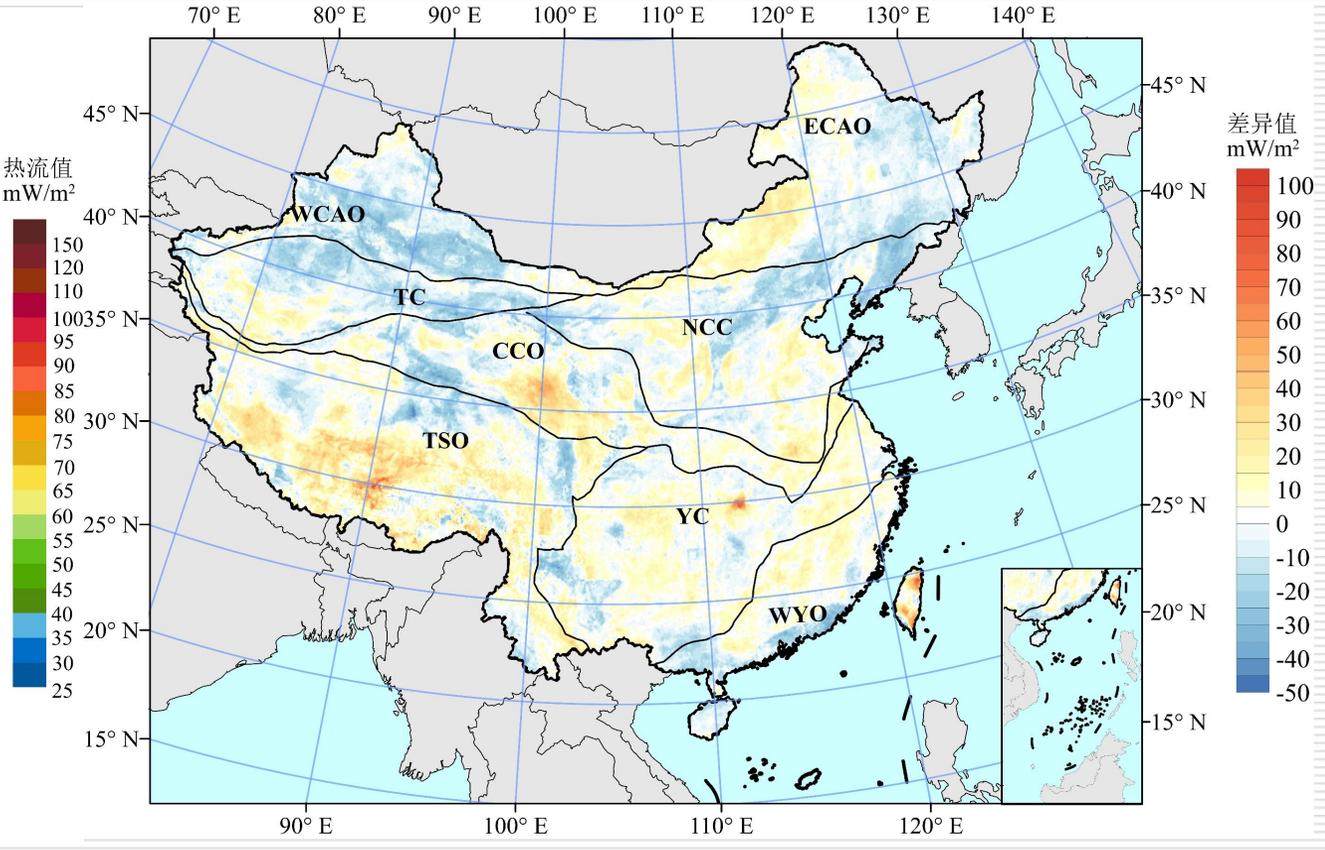
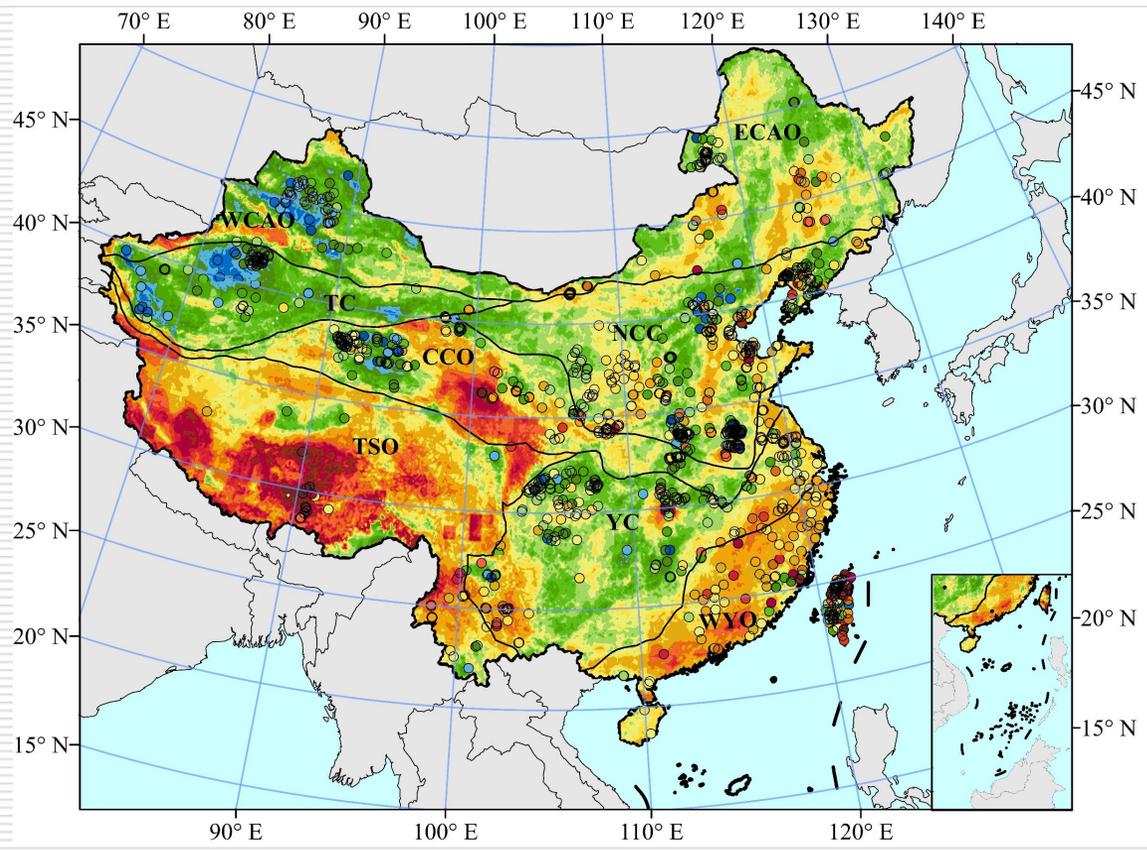


# 仅使用中国内部数据建立模型

- 相比于基于全球数据集的预测，西部差异大于东部差异；
- 考虑到东西部的构造差异，引入外部数据以提高对稀疏地区的泛化能力是必要的。

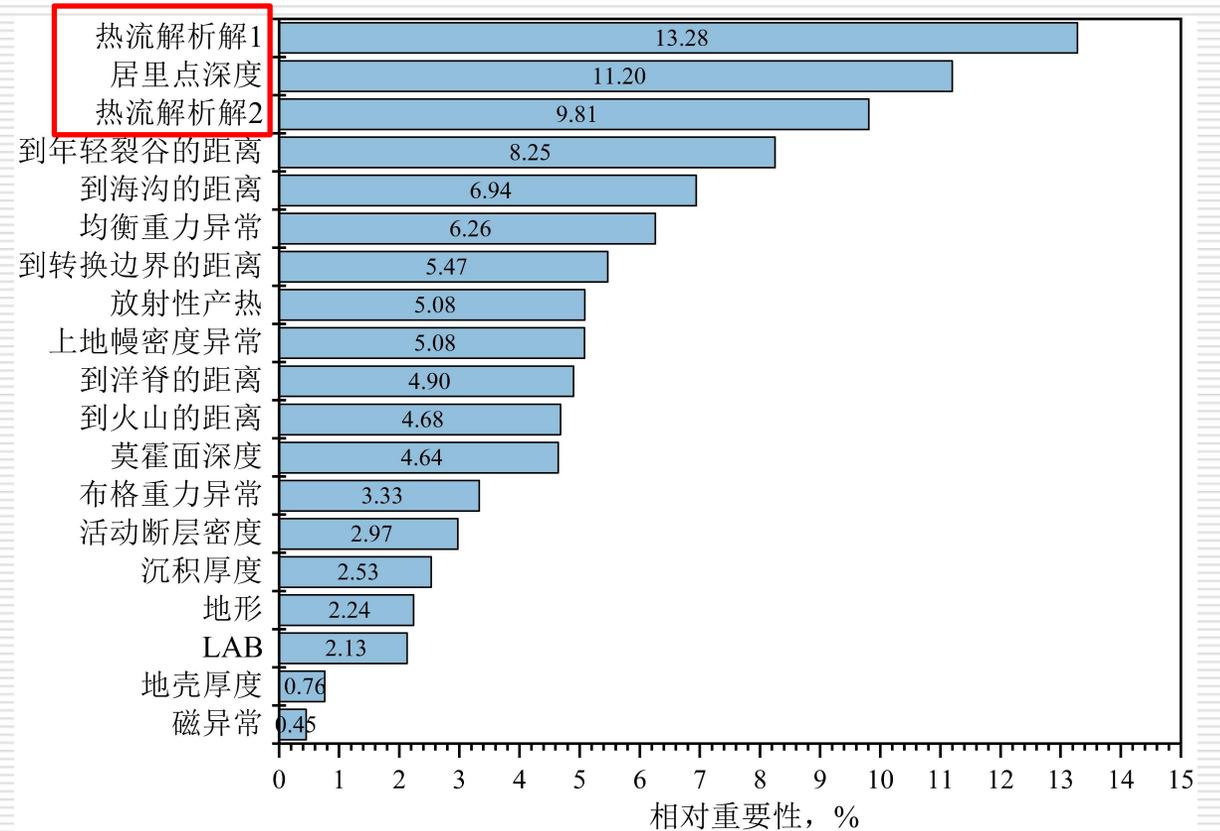
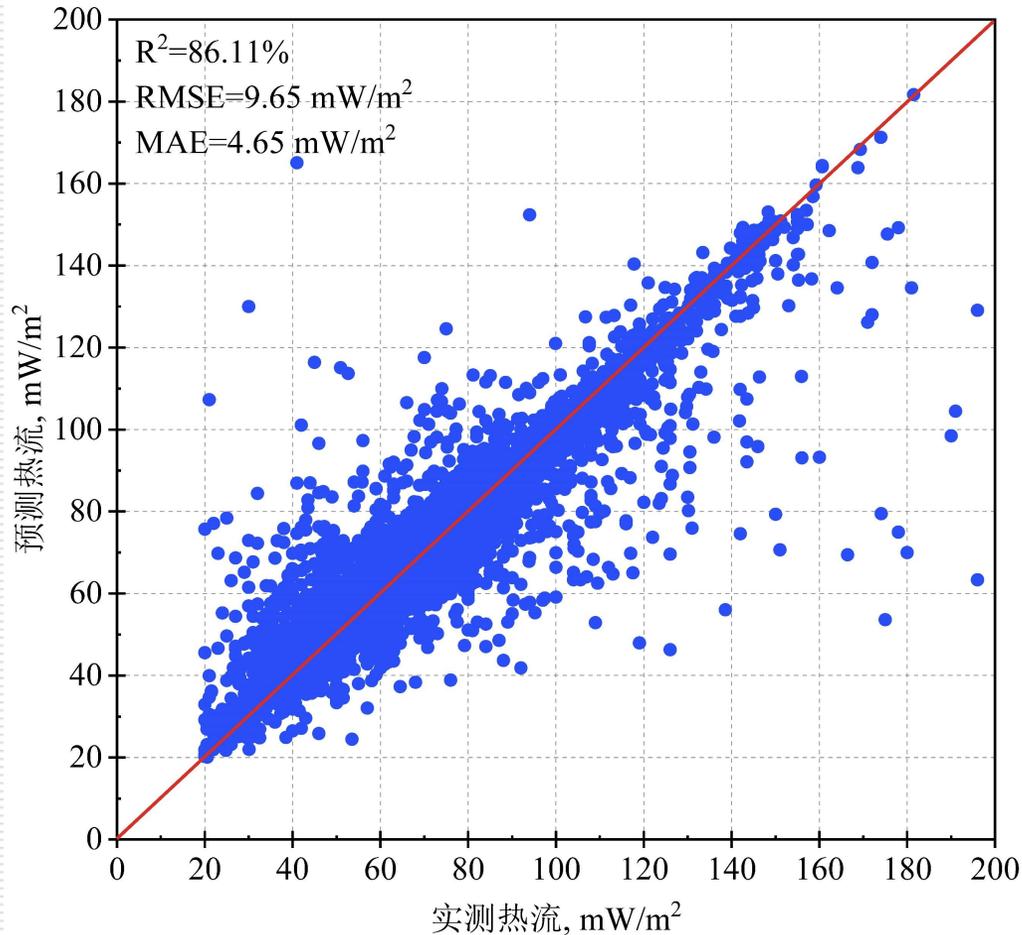


# 基于全球数据集的数据驱动的中国大地热流图



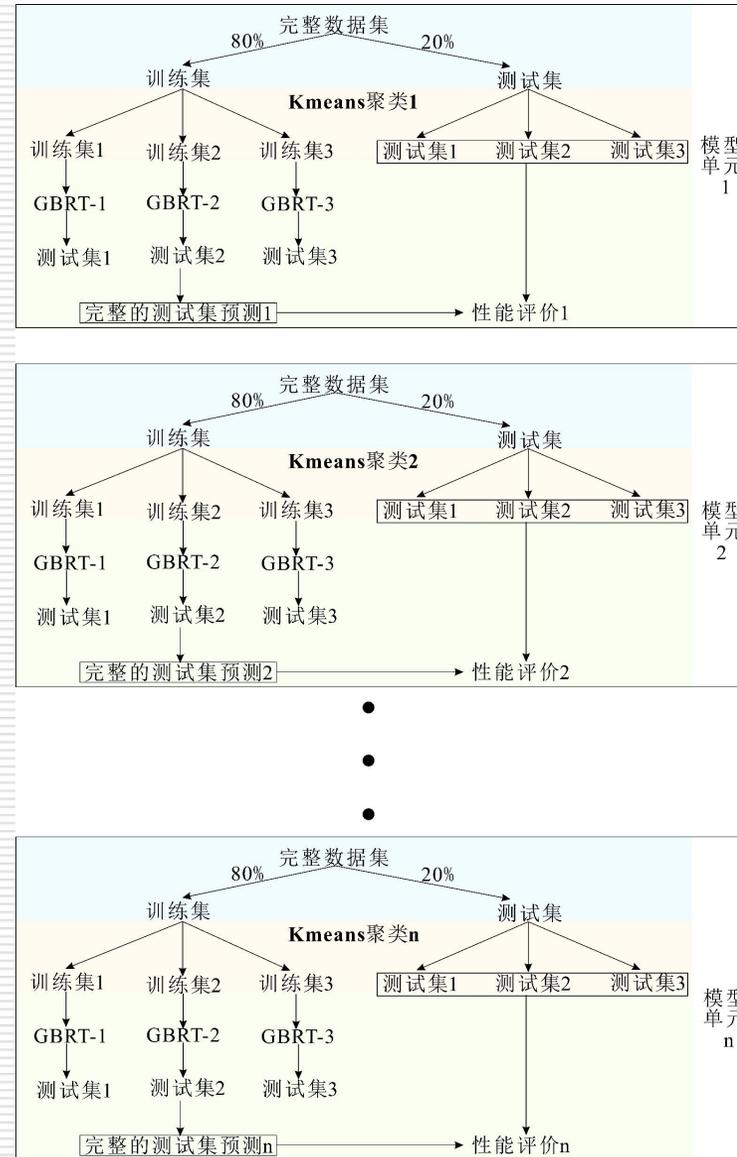
# 模型性能与特征相对重要性

- 在面对数据稀缺的情况时，物理机制的引入为模型赋予了更好的泛化能力和理论上的合理性；
- 模型的解释性和理论基础得到了增强，使得预测结果和分析过程拥有坚实的物理根据。



# 建立大地热流人工智能融合模型

改变聚类策略构建多个模型单元



结合K-means聚类GBRT构建模型单元

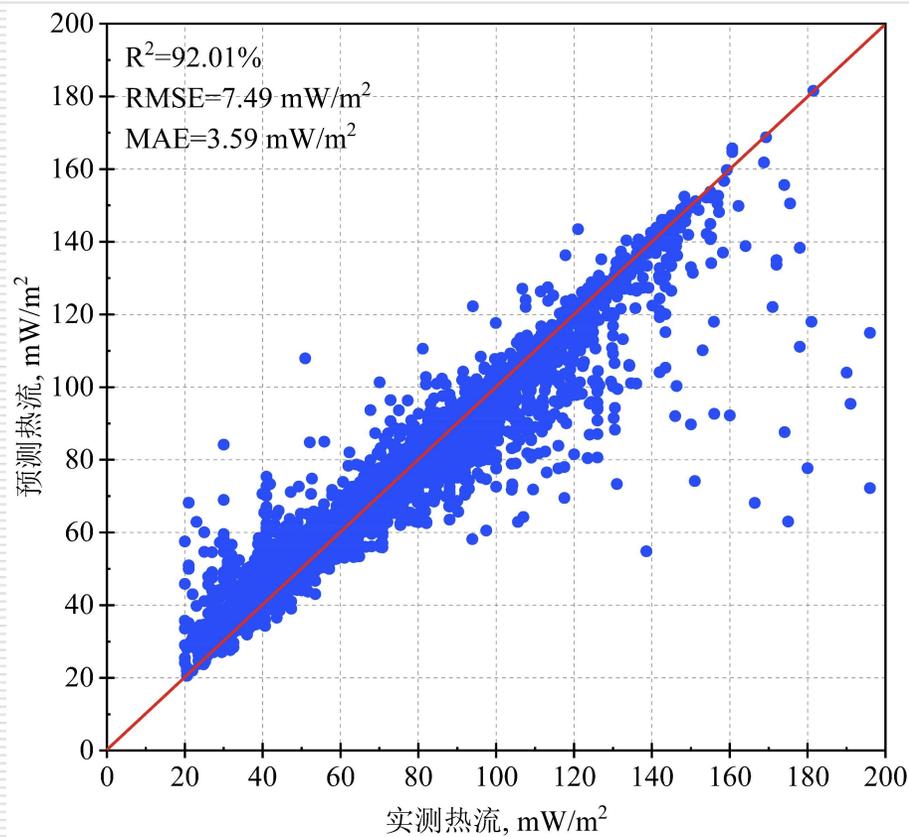
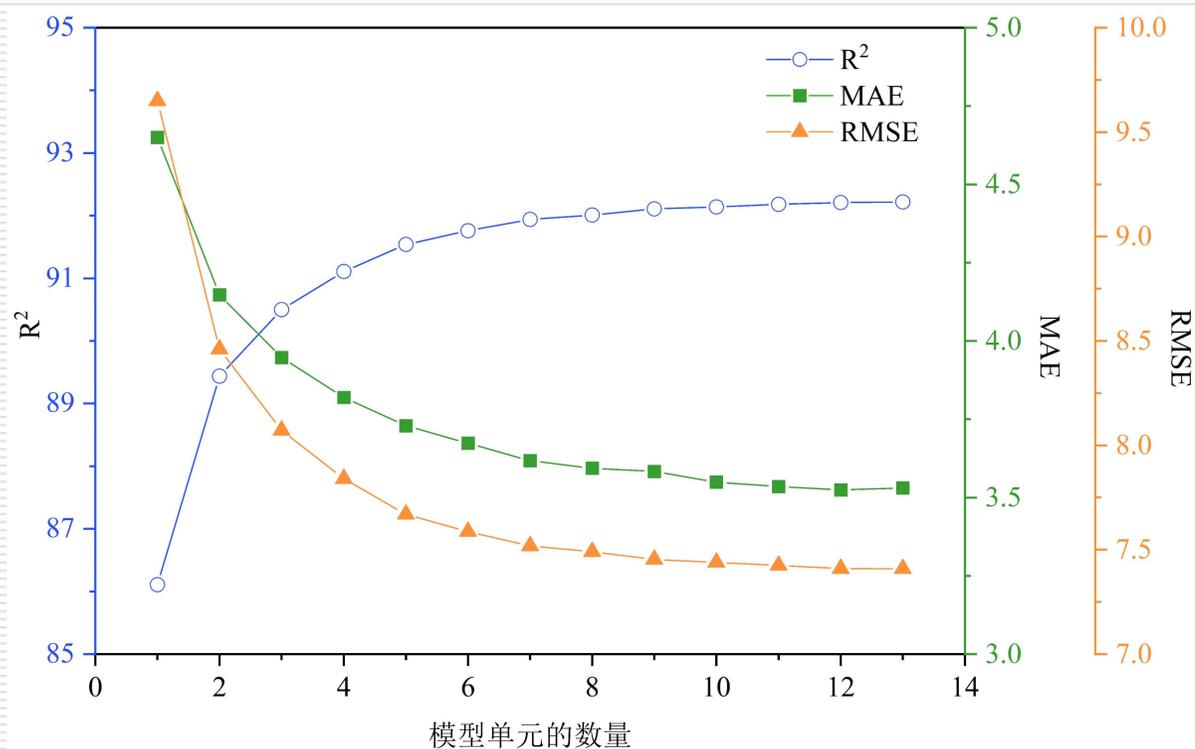
最终结果

$$\frac{\text{预测1} + \text{预测2} + \dots + \text{预测n}}{n}$$

通过平均不同模型单元的预测结果，得到最终的融合模型

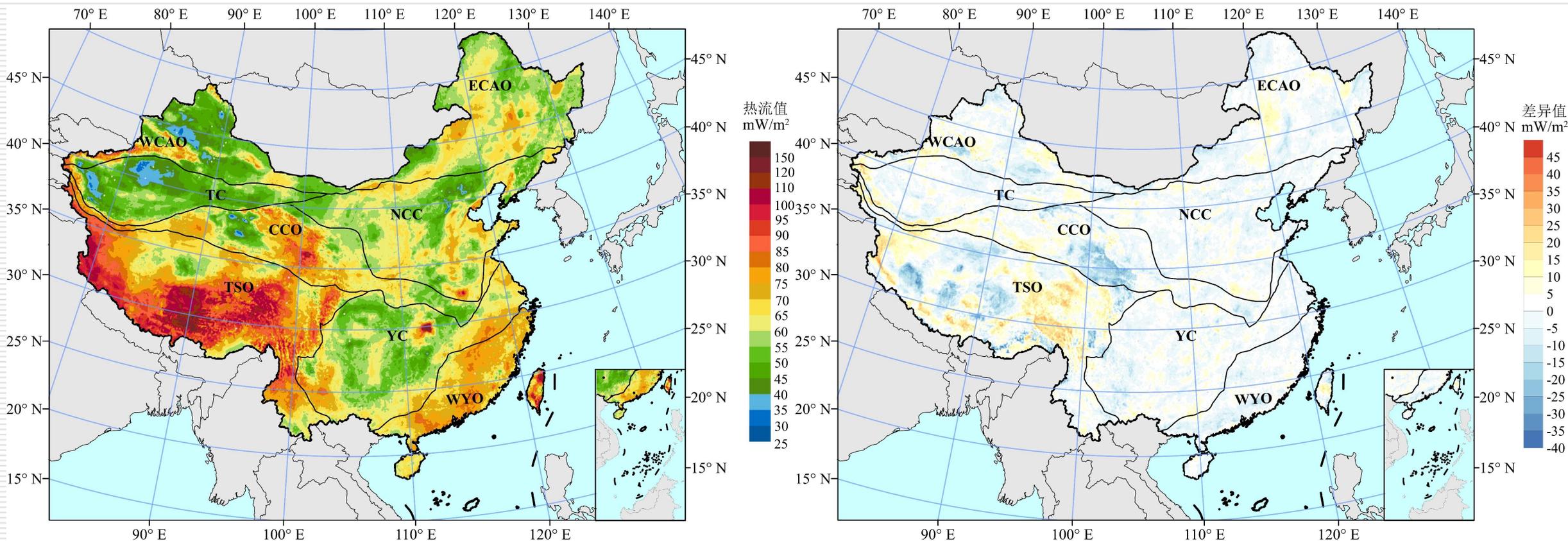
# 模型单元数量对融合模型性能的影响

- 融合模型性能随模型单元数量增加而逐渐收敛；
- 最终确定基于8个模型单元的融合模型；
- 相比于单一GBRT模型，在测试集上的 $R^2$ 提高了6.85%，MAE下降了22.72%，RMSE下降了22.37%。



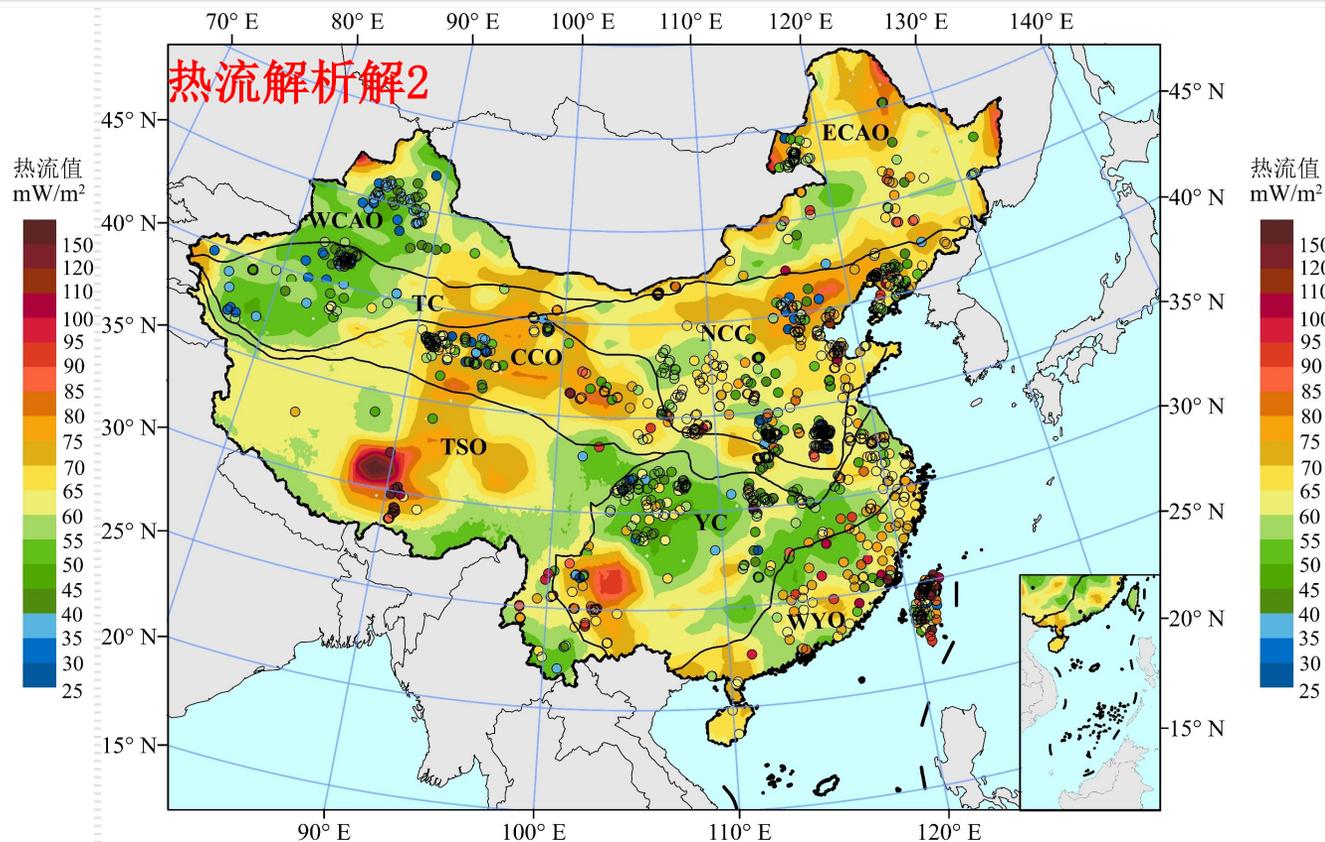
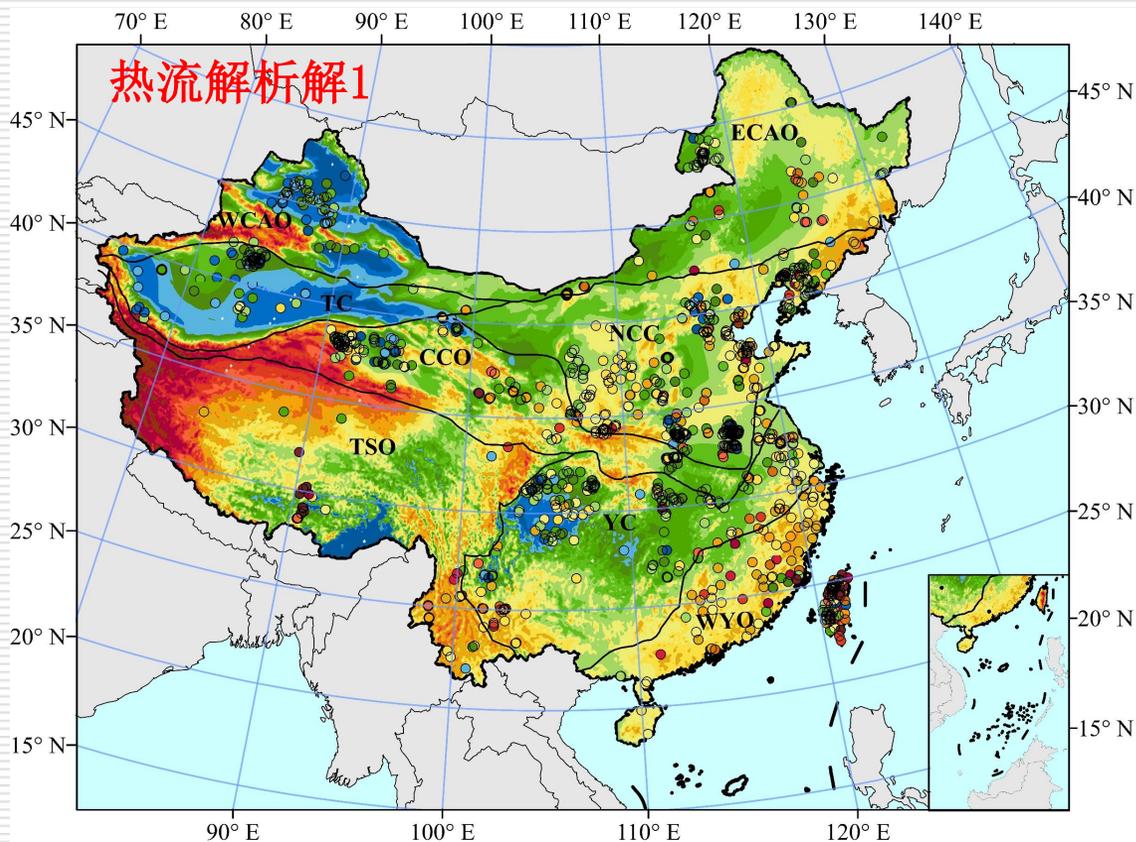
# 绘制了首个基于融合模型的中国大地热流预测图

➤ 与数据驱动模型相比，在西部的差异大，而在东部差异小；

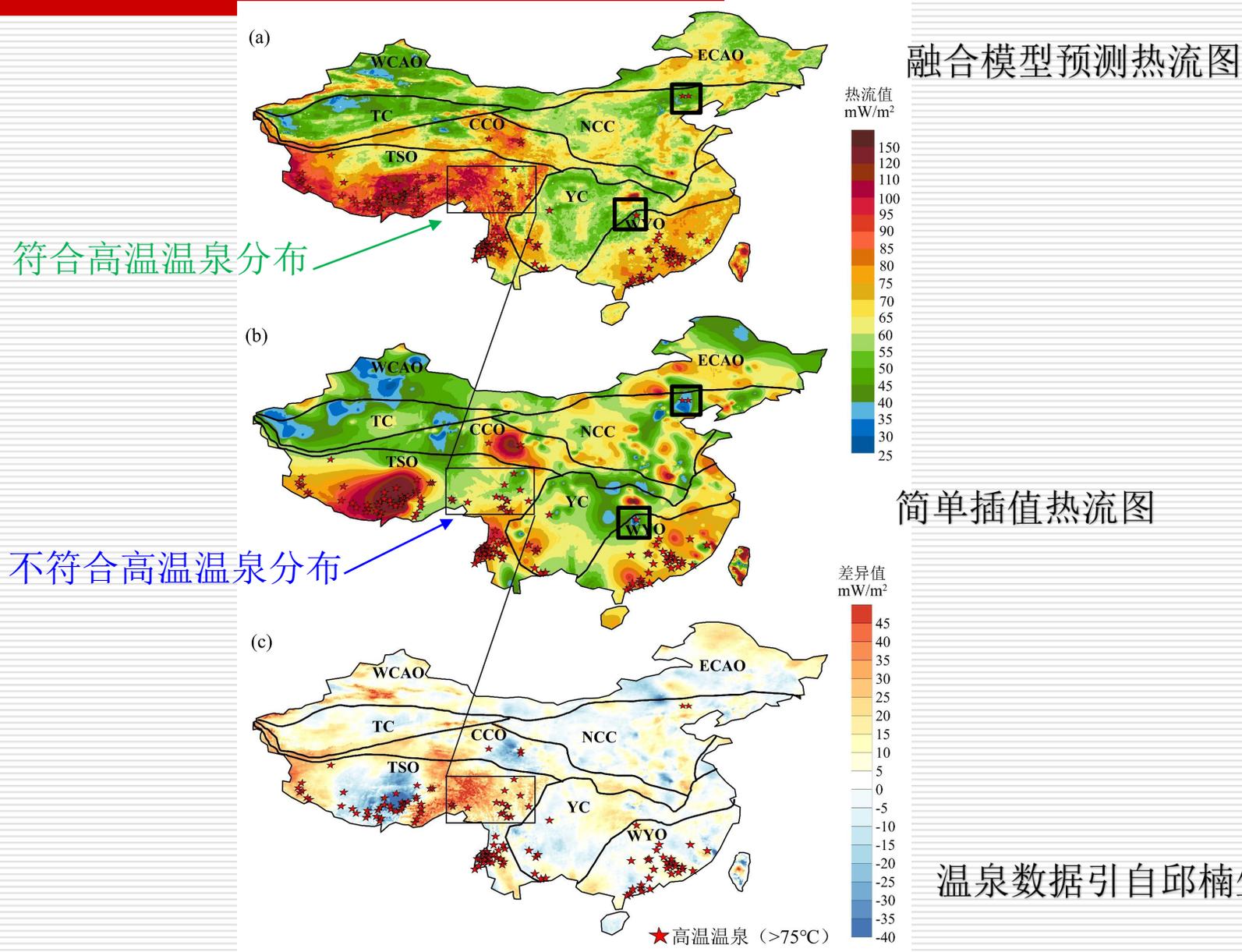


# 与传统解析法的对比

- 单纯解析法计算的热流与实测热流差距较大，尤其是热流解析解2；
- 相比之下，热流解析解1计算的热流分布更符合实测热流；
- 进一步证明了内热对地壳结构的控制作用；



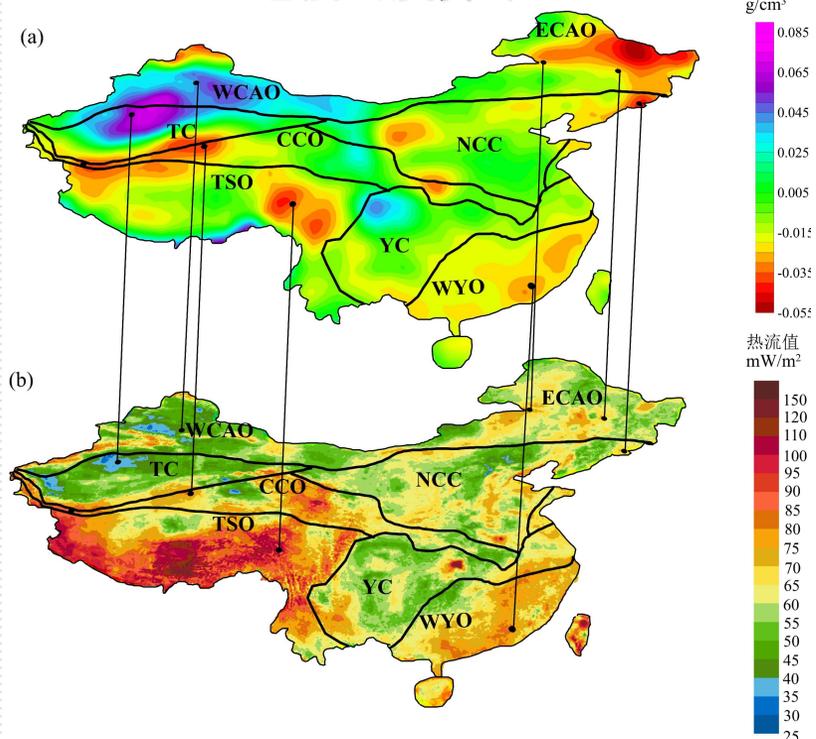
# 融合模型预测热流图更加符合高温温泉分布特征



温泉数据引自邱楠生等, 2021

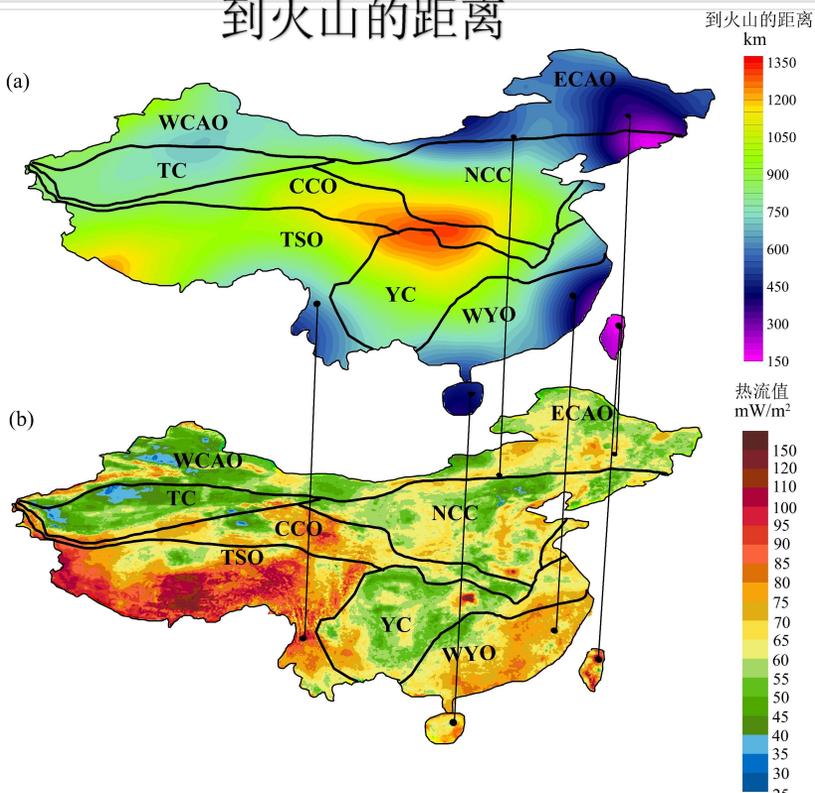
# 对预测结果的地质验证与解释

## 上地幔密度异常



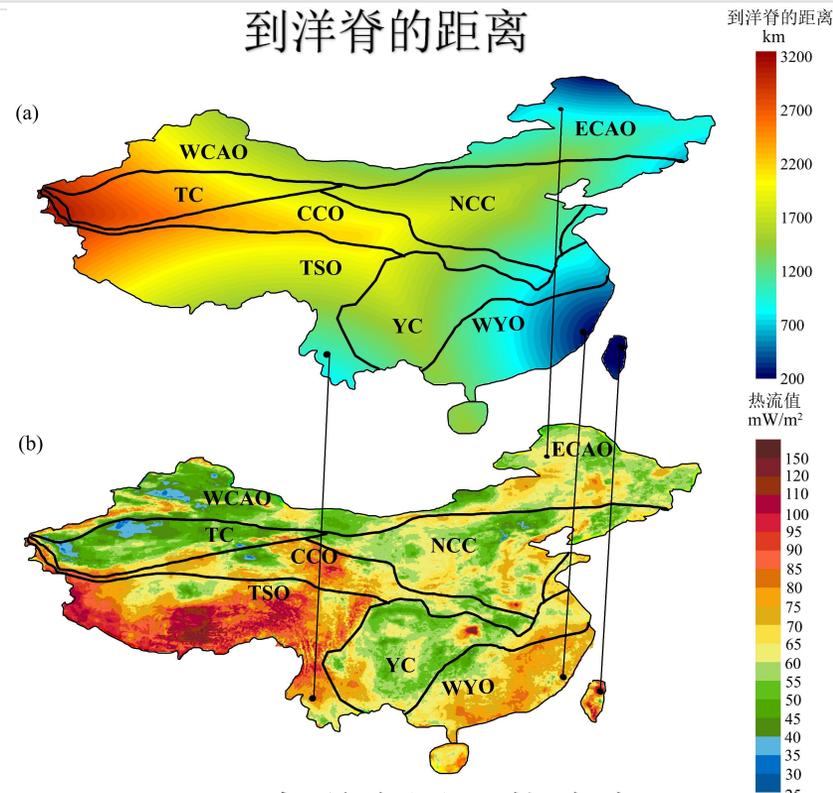
➤ 负异常对应高热流

## 到火山的距离



➤ 距离火山近，热流高

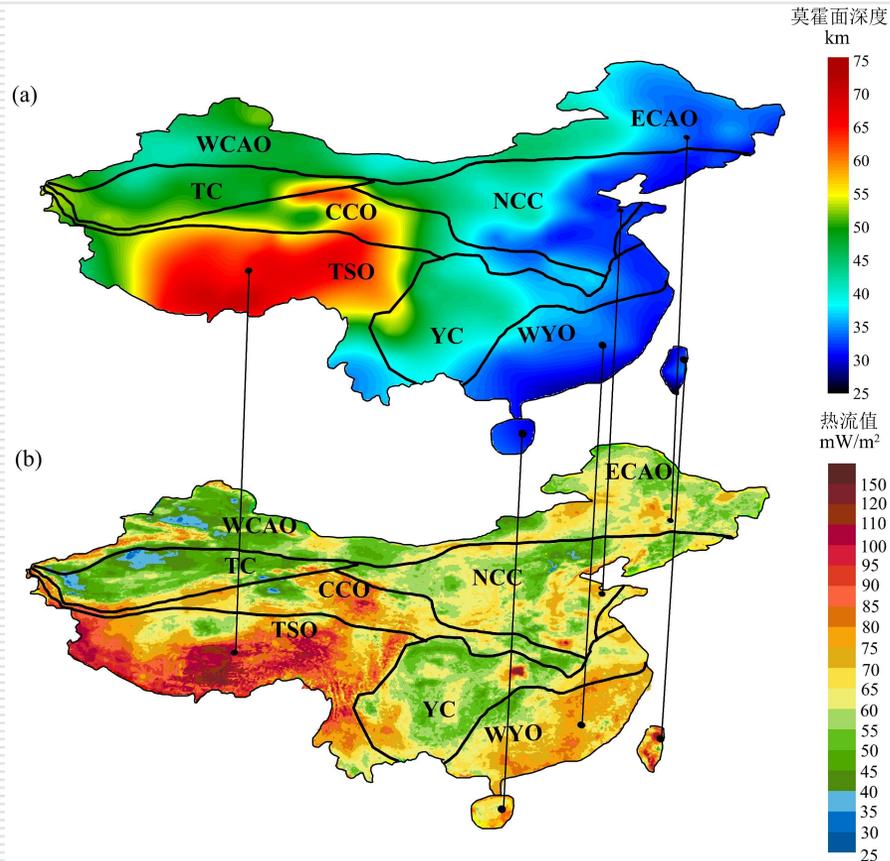
## 到洋脊的距离



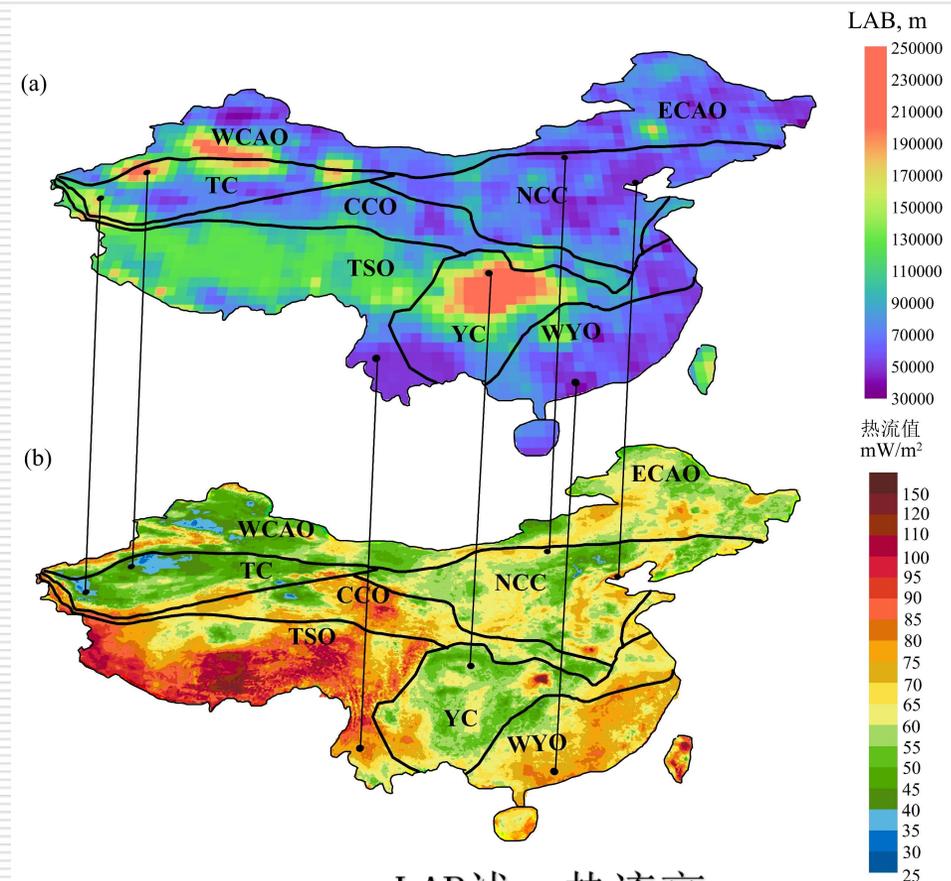
➤ 距离洋脊近，热流高

# 对预测结果的地质验证与解释

- 预测结果与关键地质特征相符；
- 任何单一特征都无法完全解释热流的变化，预测结果是综合各种特征的结果。



莫霍面埋深与热流间具有V形关系



LAB浅，热流高

# 基于最新热流图的中国深部地温场

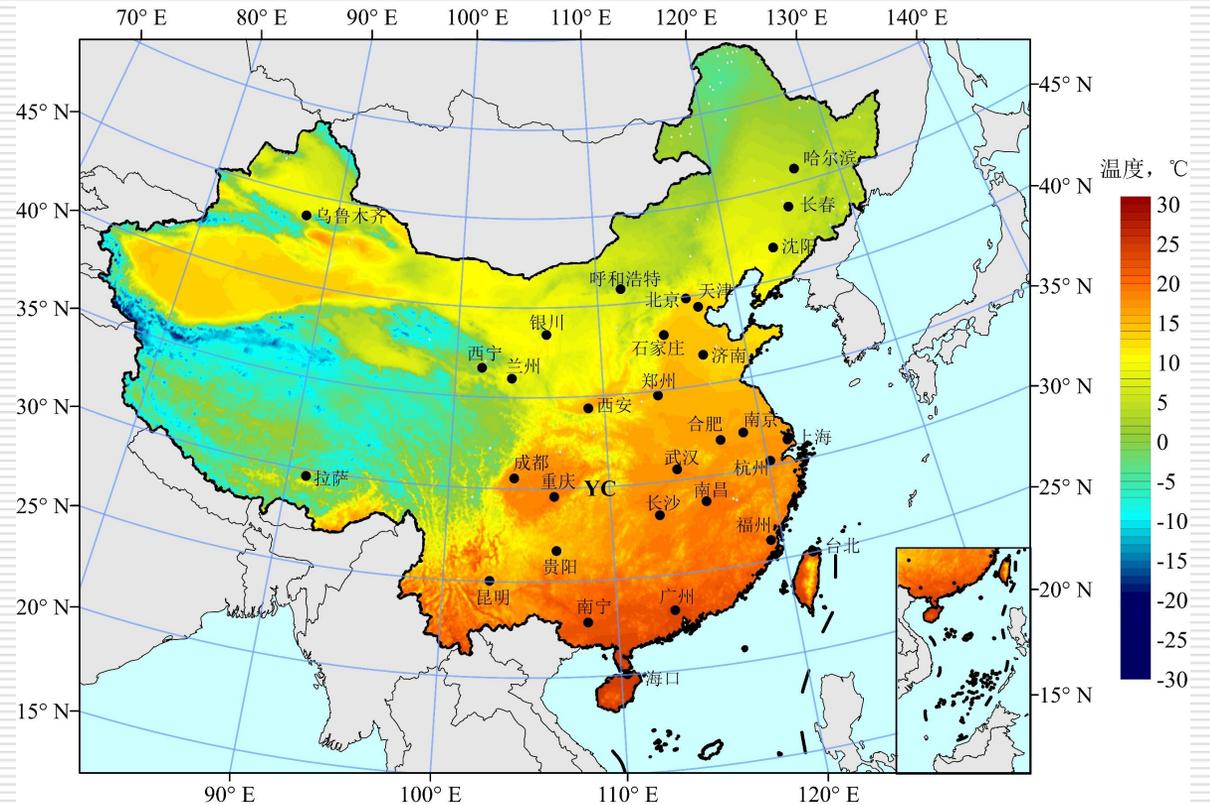
小于3km: 
$$T(z) = T_0 + \frac{q_0 z}{K} - \frac{Az^2}{2K}$$

A为生热率, K为热导率, z深度,  $q_0$ 为热流,  $T_0$ 为地表温度

大于3km: 
$$A(z) = A_0 \exp\left(-\frac{z}{D}\right)$$

$$T(z) = T_0 + \frac{z}{K}(q_0 - A_0 D) + \frac{A_0 D^2}{K} \left[ 1 - \exp\left(\frac{-z}{D}\right) \right]$$

本论文针对3km之上和之下, 均采用了最新的生热率分层填图成果

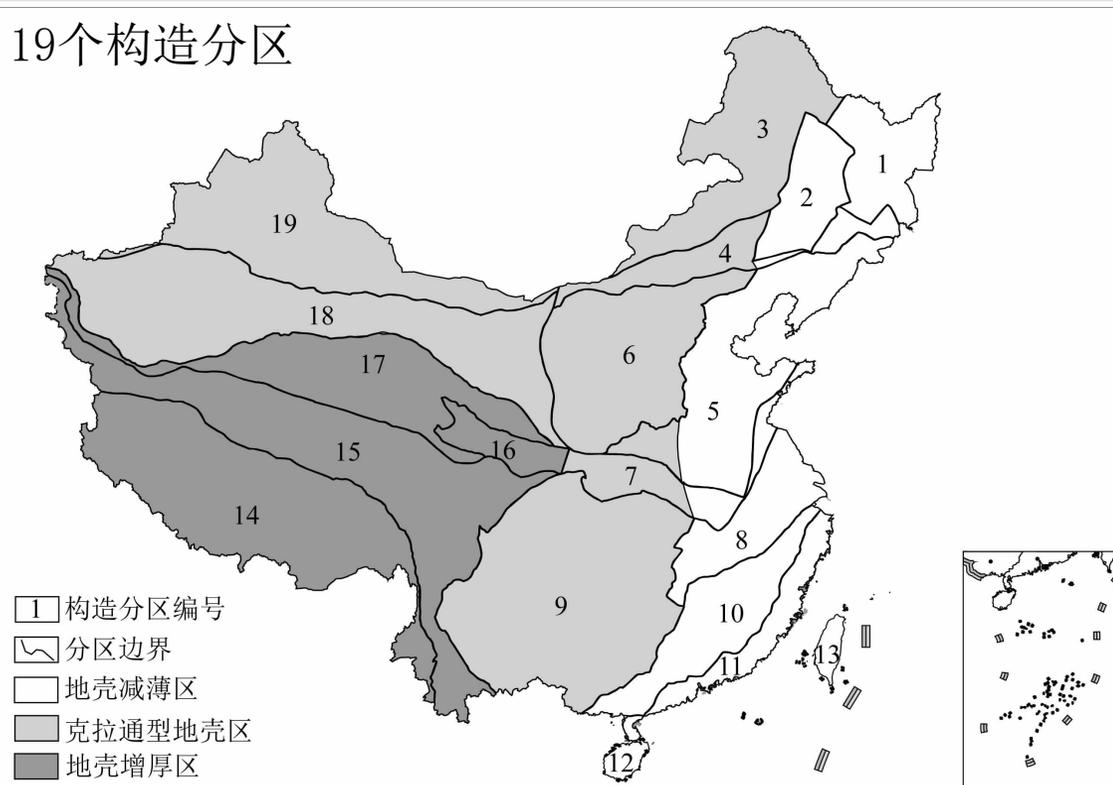


Global Solar Atlas

1994-2023年地表2m之上年平均气温数据

# 采用了最新的中国生热率填图成果

19个构造分区



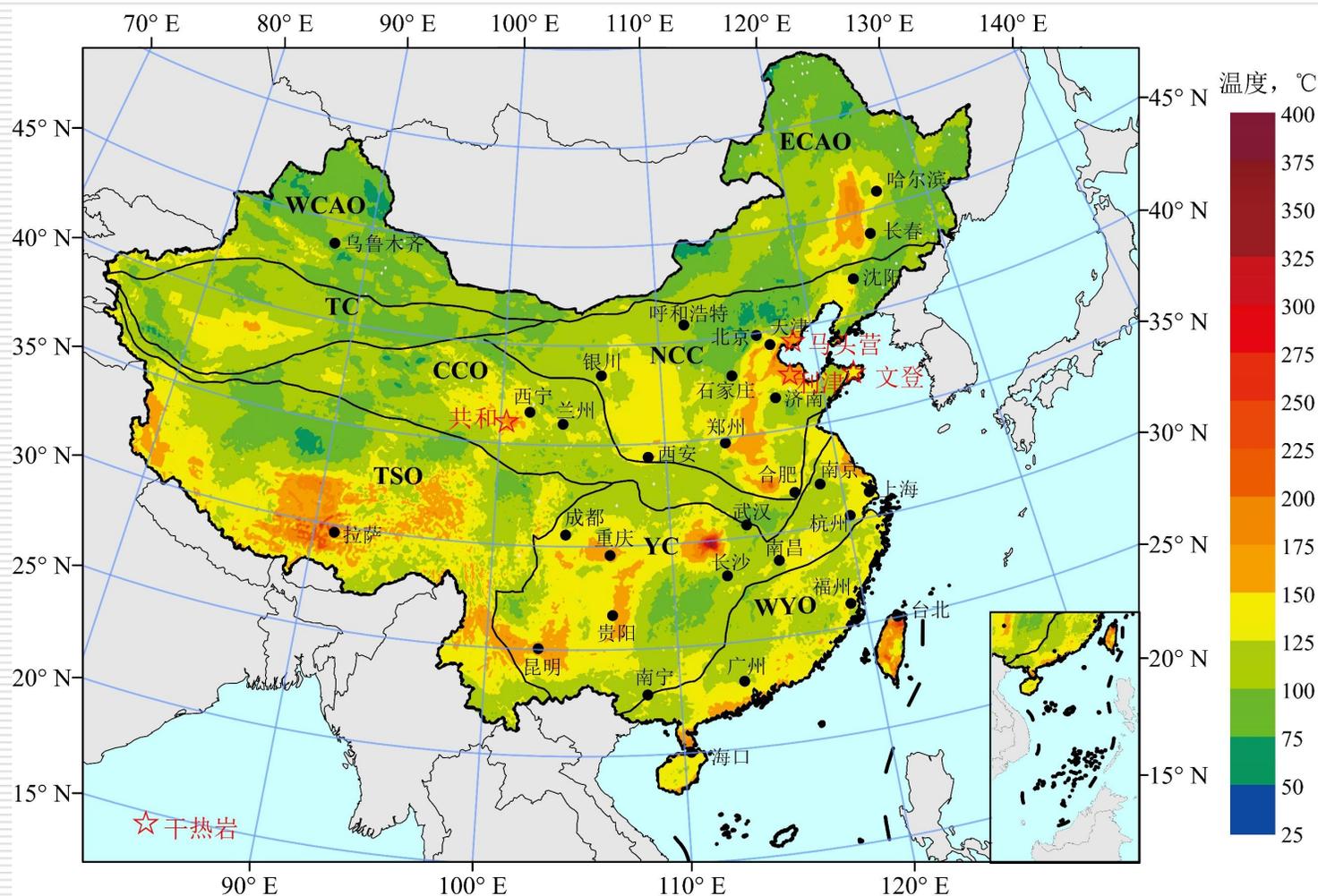
王贵玲等, 2023

分区序号	生热率, $\mu\text{W}/\text{m}^3$					
	第四系岩层	新生界岩层	中生界岩层	基底层	上地壳	中地壳
1	0.75	1.14	2.97	1.57	1.20	0.60
2	0.75	1.14	1.64	2.89	1.50	0.70
3	0.75	1.14	2.12	1.63	1.40	0.60
4	0.72	1.59	2.80	2.72	1.52	0.97
5	1.33	1.49	1.50	1.72	1.01	0.38
6	1.43	1.82	1.60	2.05	1.26	0.60
7	1.02	1.43	1.39	1.60	0.91	0.55
8	0.86	2.38	2.14	1.96	1.10	0.83
9	1.02	2.38	2.14	1.96	1.26	0.60
10	-	-	-	5.18	1.88	0.60
11	-	-	-	3.01	1.88	0.60
12	-	-	-	2.85	1.67	0.60
13	-	-	-	3.01	1.67	0.60
14	0.90	1.23	2.39	2.39	1.57	0.43
15	0.90	2.20	2.12	2.33	1.57	0.72
16	0.90	3.80	3.96	4.03	0.94	0.38
17	2.21	2.21	1.62	0.72	0.94	0.38
18	1.28	1.18	1.16	1.06	0.95	0.38
19	1.28	1.18	0.85	1.31	0.95	0.38



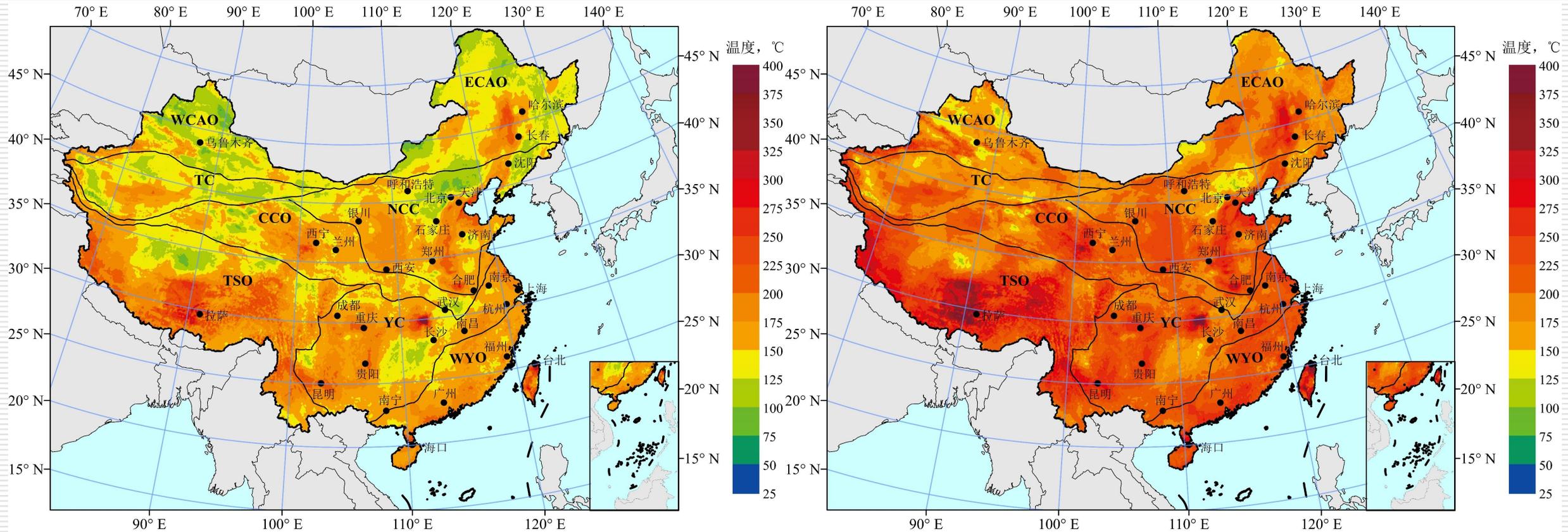
# 中国5km地温分布图

➤ 些已发现和推测的干热岩在本研究推算的地温分布图中都有体现，验证了该地温分布图的准确性



# 中国7km和10km地温分布图

- 西藏-三江造山带的南部、扬子克拉通成都-重庆-贵阳-昆明一线、华北克拉通的东部以及东中亚造山带的东部等地区，这些地区可能是高温地热资源的有利区。



# 结论

01

大地热流与太阳直接法相辐照度之间存在V字形关系，转折点在直接法相辐照度为4 kWh/m<sup>2</sup>左右。这种关系在全球范围内普遍存在，且不受海拔的影响，在不同的海拔范围内均可观察到这种关系；

02

低通滤波数据预处理方法优于其他方法。基于该方法的模型对中国实测热流预测的准确性更高。而网格平均处理会减少数据量并可能导致模型不稳定和细节丢失，网格平均值可能无法反映实际热流水平，影响模型训练与预测效果；

03

人工智能大地热流预测模型的性能随着特征数量增加而逐步提高，最终趋于稳定。这一收敛过程受到数据量及特征组合的影响。通过使用更多的学习样本和优先考虑重要性更高的特征，能显著加快模型的收敛速度；

04

本研究建立的GBRT模型能够在不使用中国实测大地热流数据的情况下，识别出中国相对高和低热流区，而插值方法在此情况下无法做出类似预测。此外，通过将中国数据纳入训练集，可以进一步提高模型对中国实测数据预测的准确性；

# 结论

05

由于中国实测数据集中在东部地区，导致利用全球数据与中国区域内部数据建立的模型对中国西部的预测差异显著超过东部。鉴于中国数据分布的不均衡性，结合国内外数据构建泛化能力更强的预测模型是必要的；

06

结合K-means聚类、GBRT以及对预测结果进行平均的方式，构建了首个大地热流人工智能融合预测模型。该模型的性能随模型单元数量增加具有收敛现象。集成8个模型单元的融合模型比单一模型在测试集上的 $R^2$ 提高了6.85%，MAE下降了22.72%，RMSE下降了22.37%，显著提升了模型的精度和可靠性；

07

绘制了首张基于融合模型预测的中国最新大地热流图，经过了高温温泉数据和地质数据的验证，不仅展现了比传统预测方法更高的准确性，而且更贴合地质及地球物理特征；

08

基于最新的大地热流图和更精细的生热率数据，计算并绘制了中国3~10km 深部地温分布图。该图比基于传统简单插值热流图推算的深部地温更加符合实际测温数据。新的地温分布图指示，西藏-三江造山带的南部、中央造山带的中部、扬子克拉通成都-重庆-贵阳-昆明一线、武夷-云开造山带的东南部、华北克拉通的东部以及东中亚造山带的东部等地区，可能是高温地热资源勘探开发的有利区；



中國地質大學

CHINA UNIVERSITY OF GEOSCIENCES

北京 BEIJING

THANKS

请您批评指正