Food and Agriculture Organization of the United Nations

GLOBAL SYMPOSIUM ON
SOIL INFORMATION AND DATA

MEASURE
MONITOR
MANAGE

Mapping SOC In Farmland Of Southern china Using A Bayesian Spatial Model

*Hanjie Ni, Bifeng Hu**

September 25-28, 2024
Nanjing, China

GLOBAL SOIL PARTNERSHIP

Outline

1. Introduction
2. Main objectives
3. Materials and methods
4. Results
5. Discussion & Conclusion

1. Introduction

- **The Critical Role of Soil Organic Carbon in Climate and Agriculture**

SOC significantly impacts climate change and soil fertility, making its spatial prediction essential for sustainable land management.

- **Advances and Limitations in Predicting Soil Properties**

Various methods, including OK, GWR, and machine learning, have been developed to predict soil properties. OK is widely used but limited by sampling density and environmental factors. GWR accounts for spatial heterogeneity, while machine learning models like RF handle non-linear relationships effectively but lack spatial autocorrelation integration. A more comprehensive approach is needed to overcome these limitations.

- **INLA-SPDE: A Bayesian Approach for Spatial Modeling**

INLA-SPDE is a flexible and efficient Bayesian spatial model widely used in fields like air pollution and disease mapping. It enhances model explainability by providing posterior distributions and has been successfully applied to digital soil mapping, especially for SOC prediction. However, its performance in complex terrains like southern China remains to be validated.

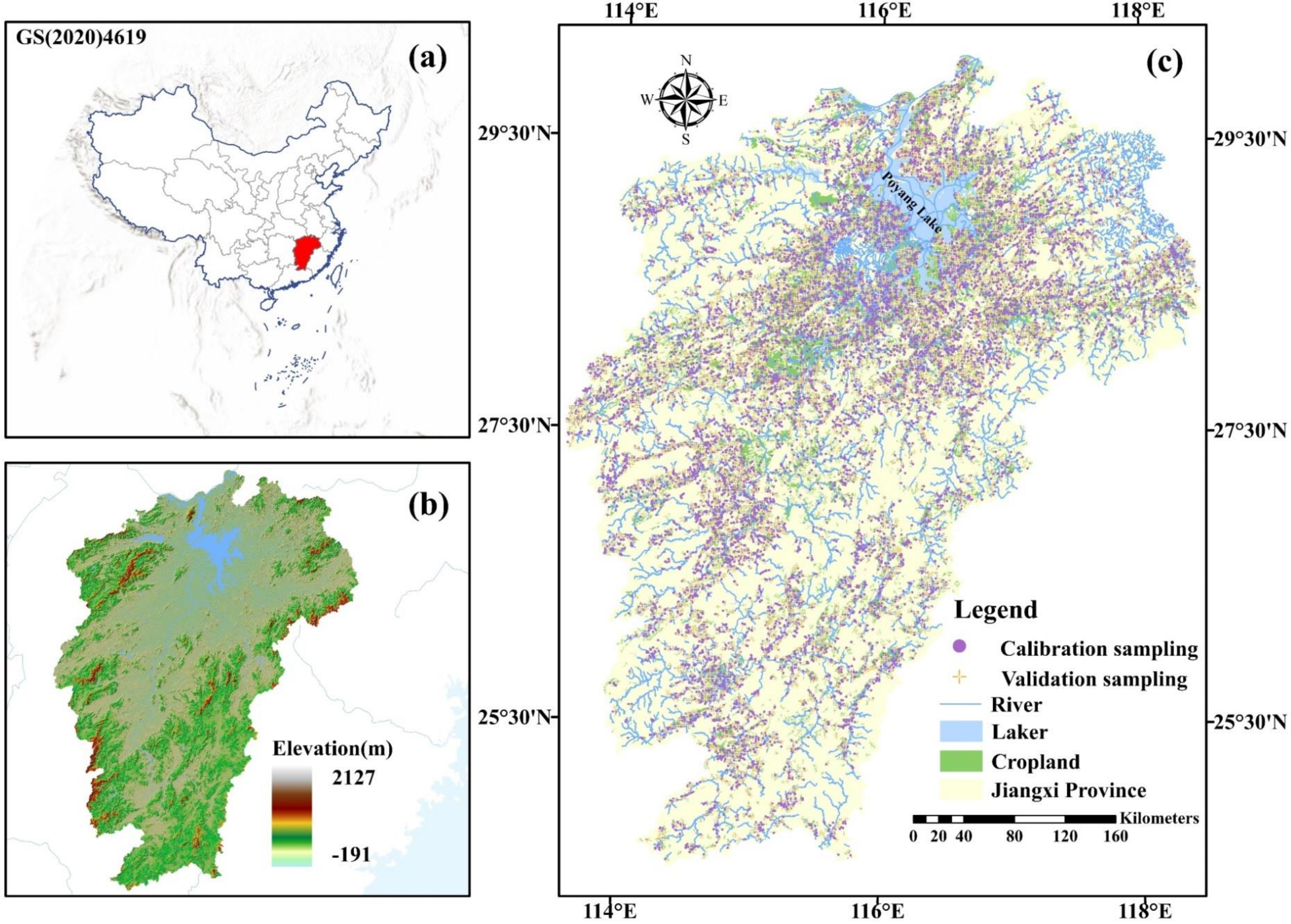
Crops in Farmlands in Southern China

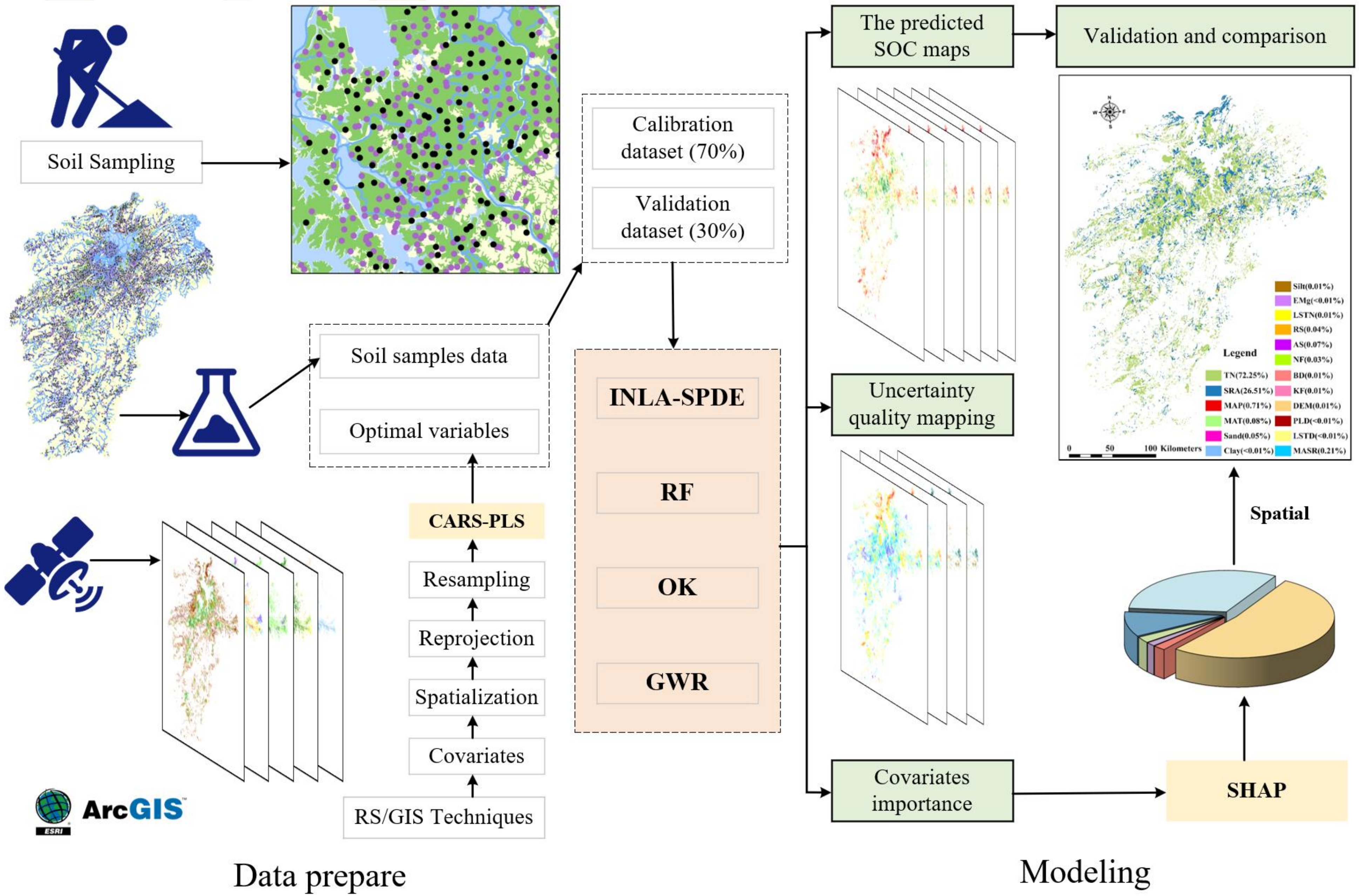
2. Main objectives

- ➢ To compare the performance of INLA-SPDE to classic different spatial predictive models on mapping SOC in farmland of Jiangxi Province in southern China which is a hilly regions featured by very complex terrain condition
- ➢ To analyze uncertainty of the results produced by INLA-SPDE and RF models
- ➢ To quantify the overall importance of different covariates and map the spatially varying primary covariates for mapping SOC at the pixel scale by using INLA-SPDE and an explainable machine learning model termed Shapley Additive explanations (SHAP)

3. Materials and methods

- A total of 16,050 soil samples
- Surface layer (0-20 cm)
- Farmland in Jiangxi Province





Key steps

- **Data Preparation**
- **Select optimal covariate**
- **Mapping SOC**
- **Calculating uncertainty**
- **Spatial SHAP**

Covariates selection

Environmental factor	Variable	Acronym	Spatial resolution	Unit	Source
Climate	Mean annual temperature	MAT	1 km	◦C	http://www.resdc.cn/
	Mean annual precipitation	MAP	1 km	mm	http://www.resdc.cn/
	Evapotranspiration	ET	1 km	-	https://lpdaac.usgs.gov/
	Mean annual solar radiation	MASR	1 km	MJ/m²	http://www.geodata.cn
	Land surface temperature, day time	LSTD	1 km	◦C	https://lpdaac.usgs.gov/
	Land surface temperature, night time	LSTN	1 km	◦C	https://lpdaac.usgs.gov/
Terrain	Elevation	DEM	30 m	m	USGS ASTGTM
	Topographic wetness index	TWI	30 m	-	Calculated from Elevation
	Topographic position index	TPI	30 m	-	Calculated from Elevation
	Multi-resolution valley bottom flatness	MRVBF	30 m	-	Calculated from Elevation
	Slope	Slope	30 m	◦	Calculated from Elevation
	Aspect	Aspect	30 m	◦	Calculated from Elevation
	Valley depth	VD	30 m	m	Calculated from Elevation
	Slope length	SL	30 m	m	Calculated from Elevation
	Catchment Slope	CAS	30 m	◦	Calculated from Elevation
	Convergence Index	CI	30 m	-	Calculated from Elevation
	Length-Slope Factor	LSF	30 m	-	Calculated from Elevation
	Catchment Area	CA	30 m	-	Calculated from Elevation

Environmental factor	Variable	Acronym	Spatial resolution	Unit	Source
Soil properties	Bulk density	BD	30 m	-	This study
	Cation exchange capacity	CEC	30 m	cmol/kg(+)	This study
	Soil pH	pH	30 m	-	This study
	Total nitrogen	TN	30 m	g/kg	This study
	Total phosphorus	TP	30 m	g/kg	This study
	Total potassium	TK	30 m	g/kg	This study
	Exchangeable Magnesium	EMg	30 m	cmol/kg(+)	This study
	Available silicon	ASi	30 m	mg kg^{-1}	This study
	Available sulphur	AS	30 m	mg kg^{-1}	This study
	Plough layer depth	PLD	30 m	cm	This study
	Clay	Clay	250 m	-	SoilGrids250m 2.0
	Sand	Sand	250 m	-	SoilGrids250m 2.0
	Silt	Silt	250 m	-	SoilGrids250m 2.0
	Percentage of soil coarse fragment (> 2 mm) (%)	CF	250 m	-	SoilGrids250m 2.0
Biota	Normalized Difference Vegetation Index	NDVI	1 km	-	http://www.resdc.cn/
	Net primary production	NPP	1 km	-	http://www.resdc.cn/
	Rotation system	RS	-	-	This study
Lithology	Soil class	SC	1 km	-	The Second National Soil Survey
	Parental material	PM	1 km	-	The Second National Soil Survey
Soil management	Input of nitrogen fertiliser	NF	-	kg ha^{-1}	This study
	Input of phosphate fertiliser	PF	-	kg ha^{-1}	This study
	Input of potash fertiliser	KF	-	kg ha^{-1}	This study
	Amount of straw return	SRA	-	kg ha^{-1}	This study
	Irrigation capacity	IC	-	-	This study
	Drainage capacity	DC	-	-	This study
	Soil erosion degree	SED	-	-	This study

Environmental covariates selection

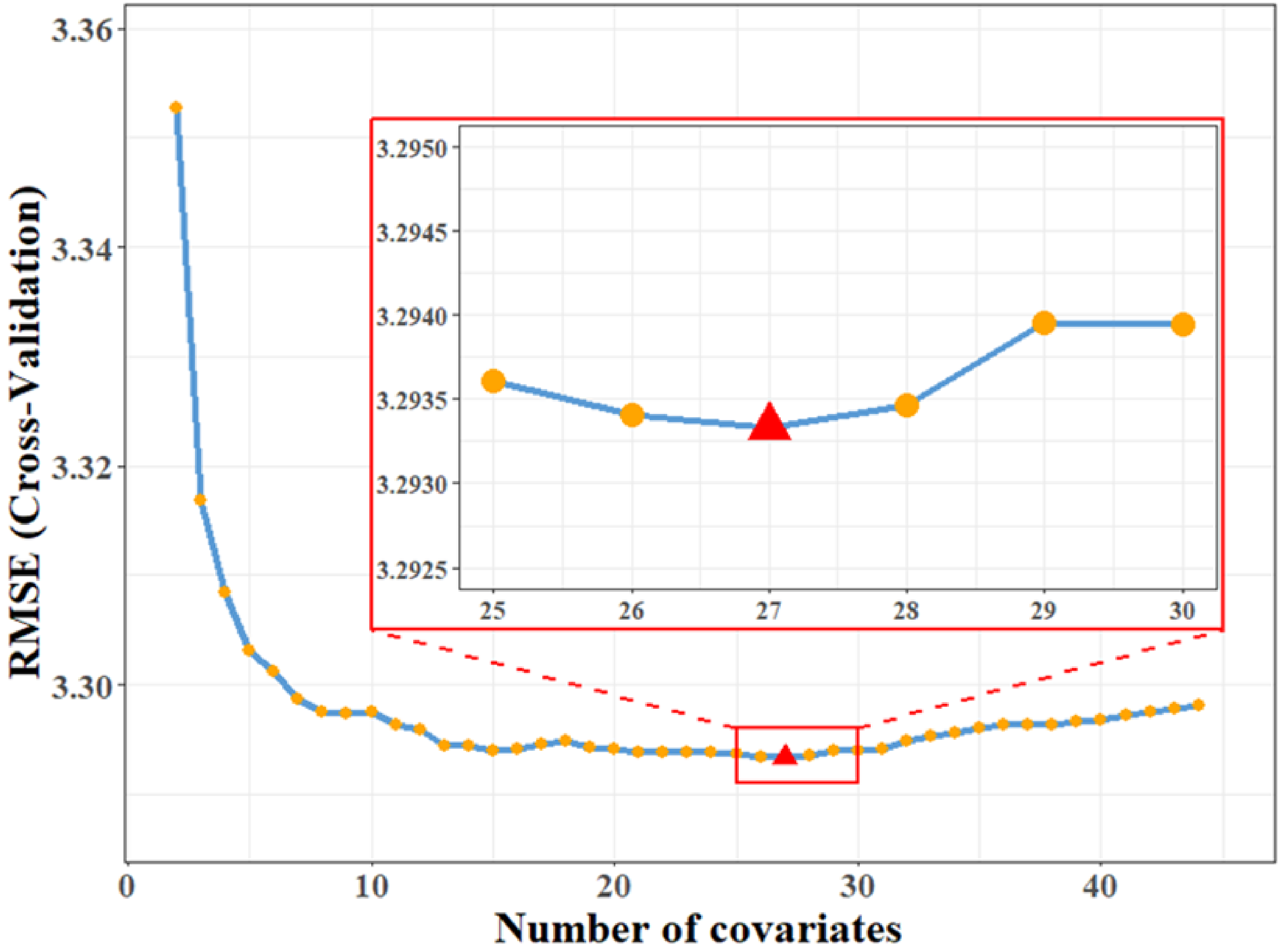
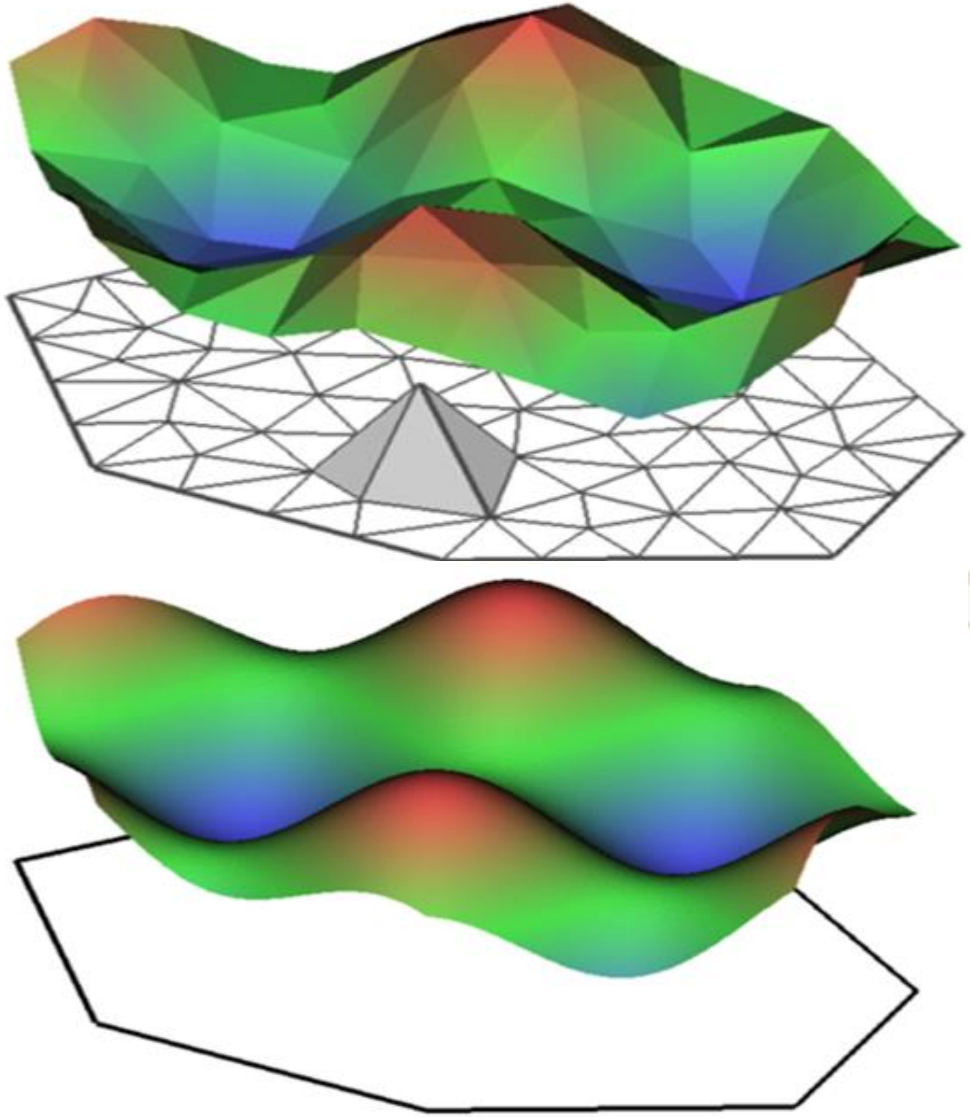
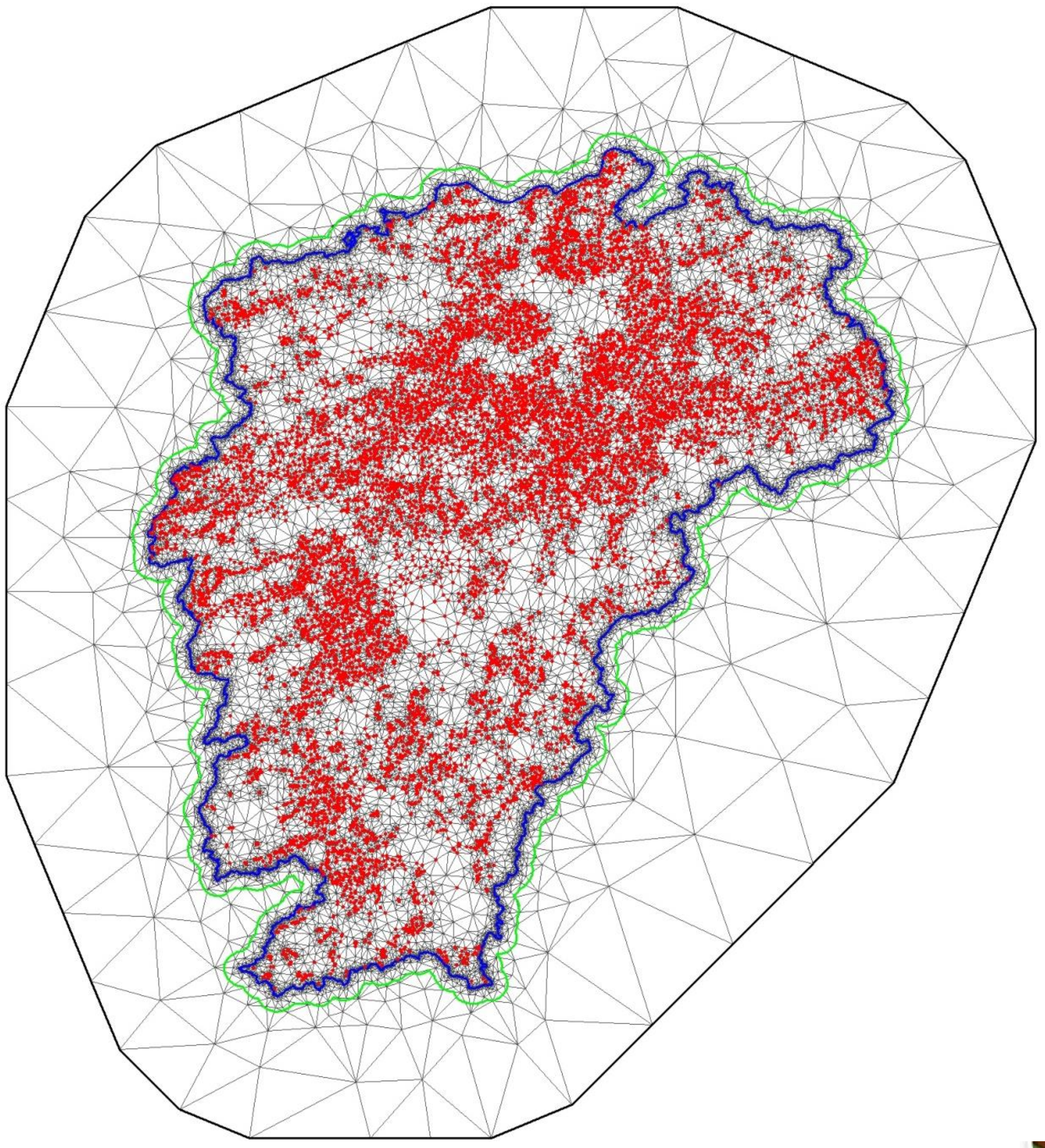


Table
Selected optimal variables for spatial prediction modeling by CARS-PLS algorithm.

Environmental covariates	CARS-PLS（N=27）
Climate	MAT，MAP，MASR，LSTD，LSTN
Terrain	DEM，TWI，TPI
Soil properties	BD，CEC，pH，TN，TP，EMg，AS，PLD，Clay，Sand，Silt
Biota	NDVI，RS
Lithology	SC，PM
Soil management	NF，KF，SRA，SED

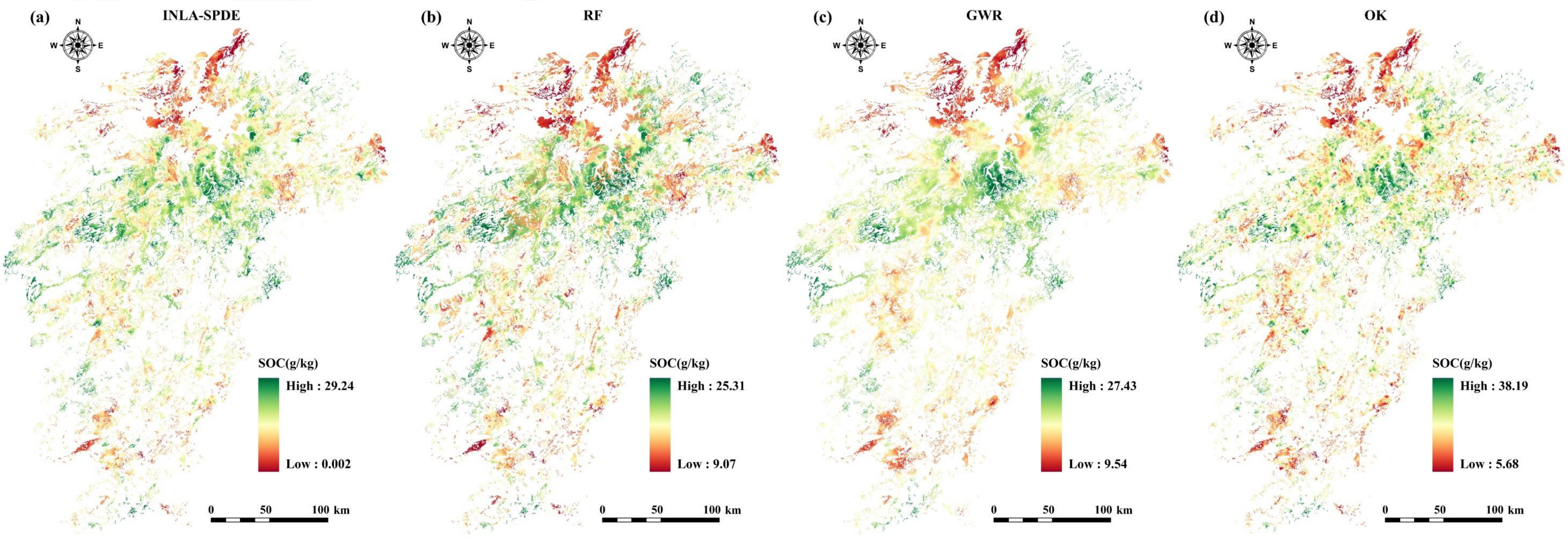
Mesh construction





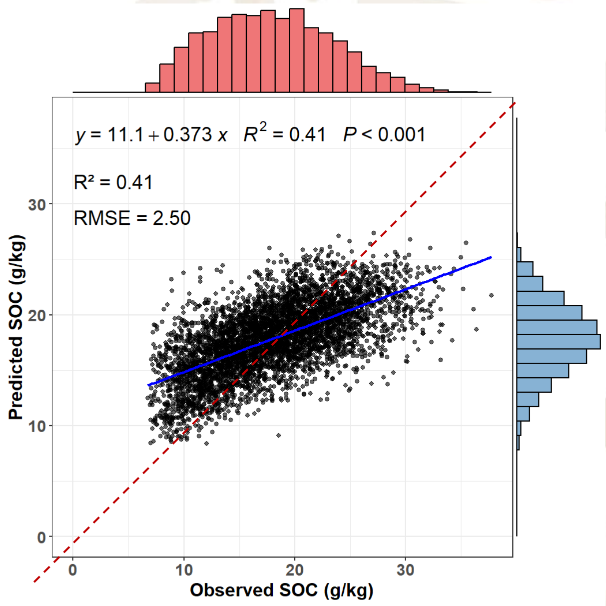
4.Results

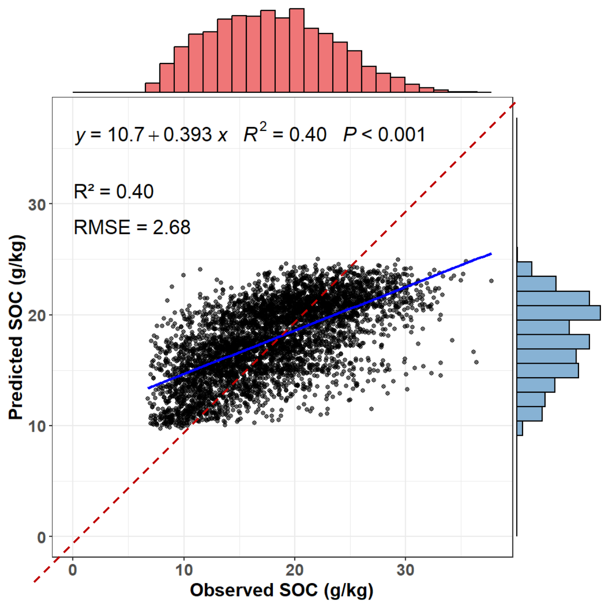
The predicted SOC maps

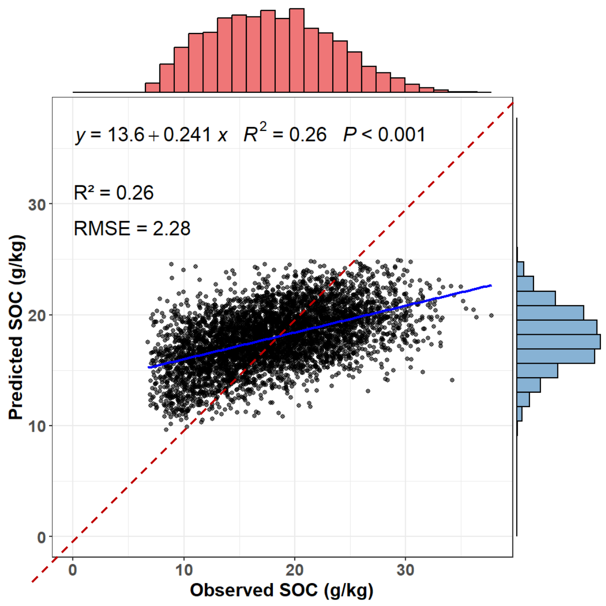


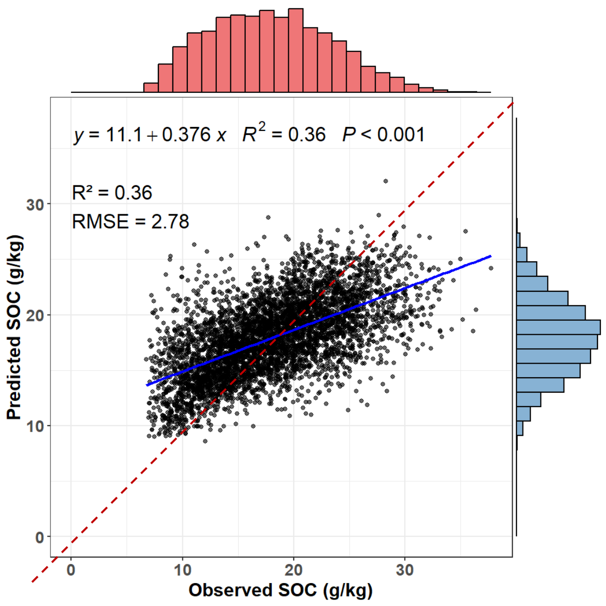
INLA-SPDE **RF** **GWR** **OK**

Prediction accuracy



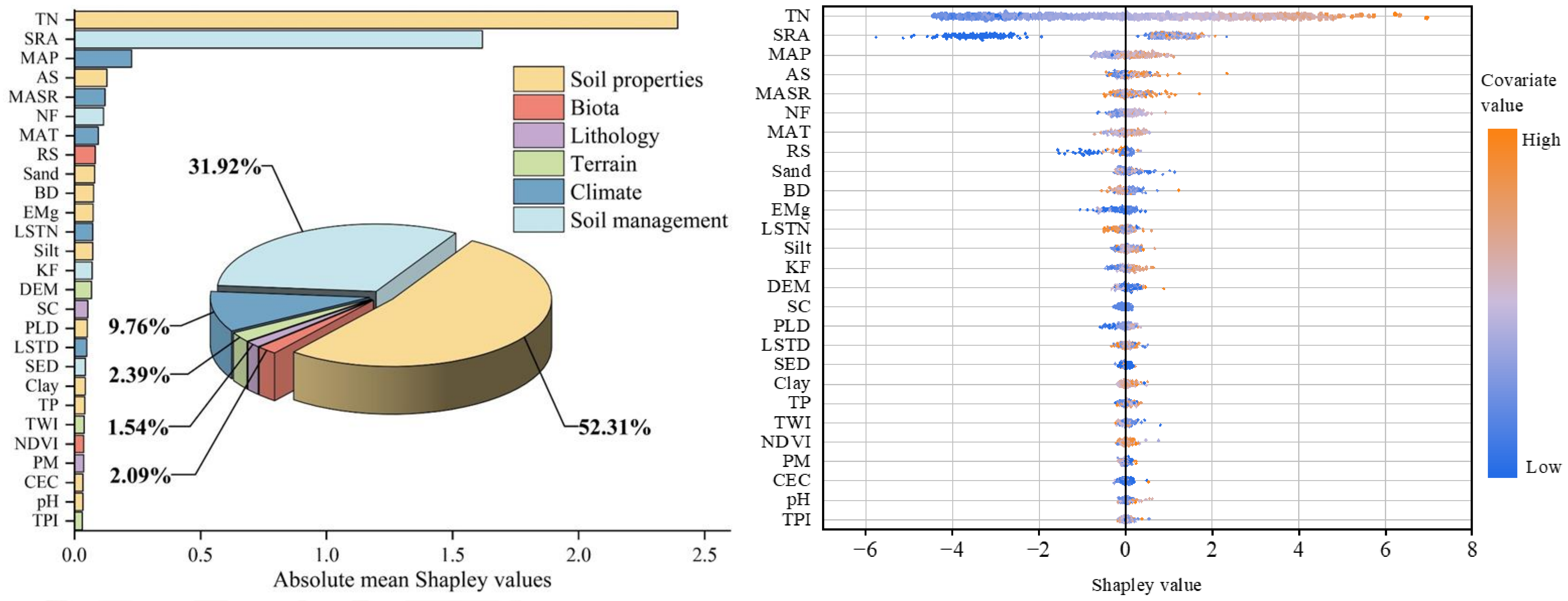


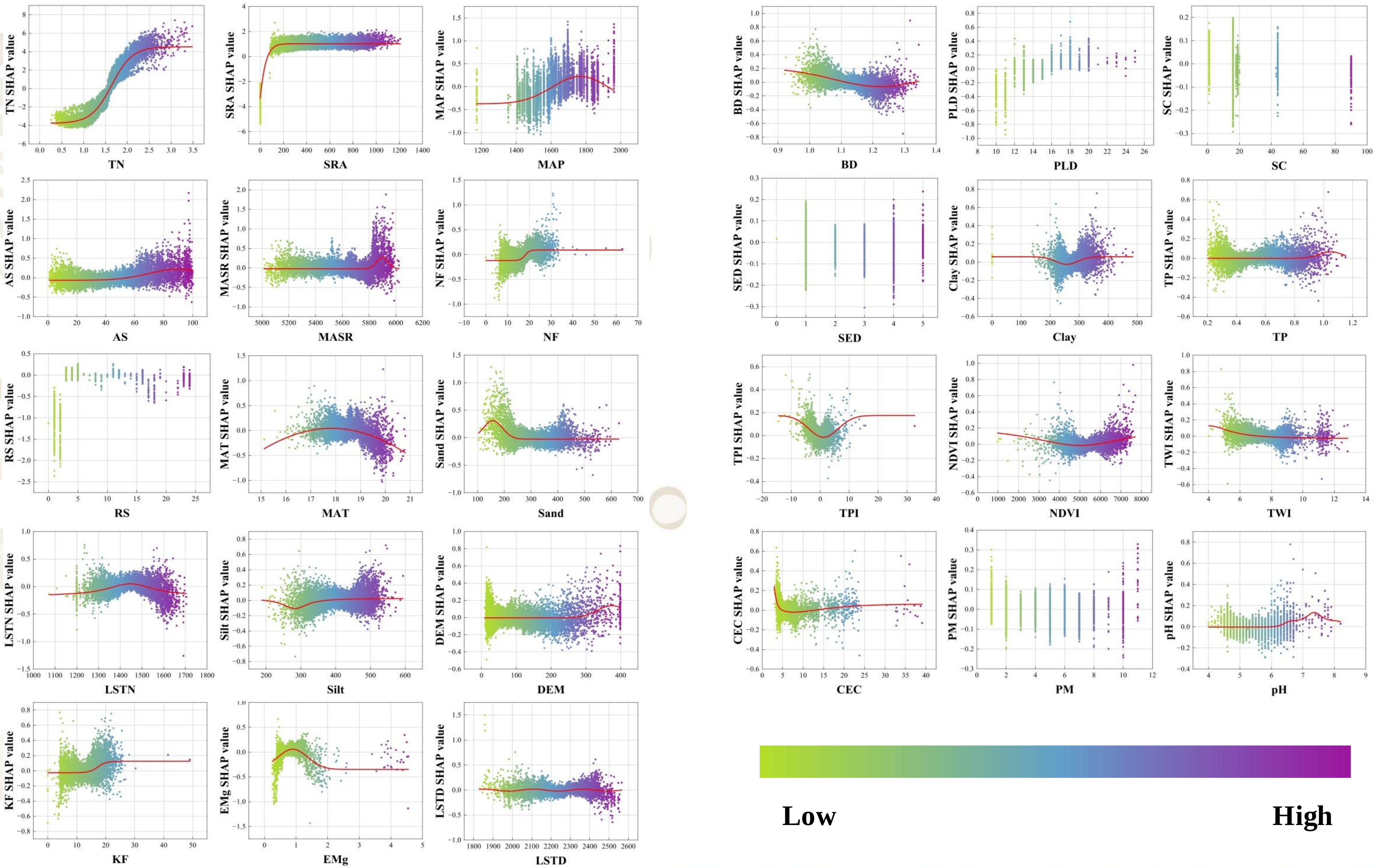




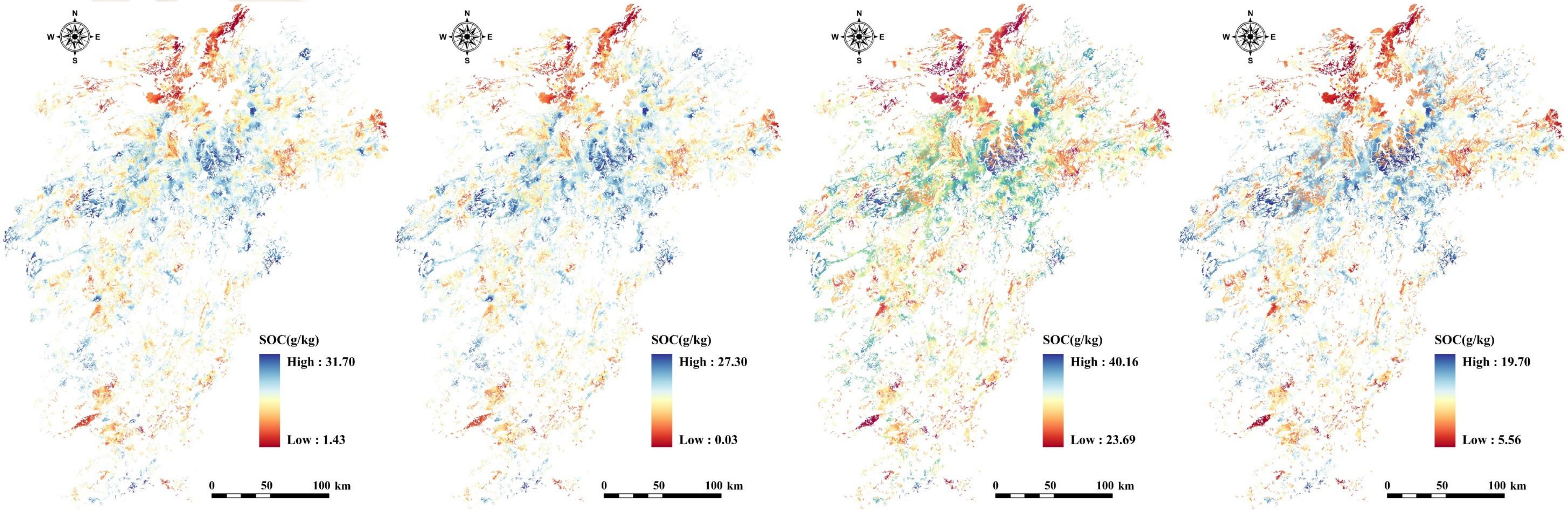
INLA-SPDE	RF	GWR	OK
R^2=0.41	R^2=0.40	R^2=0.26	R^2=0.36
RMSE=2.50	RMSE=2.68	RMSE=2.28	RMSE=2.78

Relative importance of covariates




TN SHAP value
TN
SRA SHAP value
SRA
MAP SHAP value
MAP
BD SHAP value
BD
PLD SHAP value
PLD
SC SHAP value
SC
AS SHAP value
AS
MASR SHAP value
MASR
NF SHAP value
NF
SED SHAP value
SED
Clay SHAP value
Clay
TP SHAP value
TP
RS SHAP value
RS
MAT SHAP value
MAT
Sand SHAP value
Sand
TPI SHAP value
TPI
NDVI SHAP value
NDVI
TWI SHAP value
TWI
LSTN SHAP value
LSTN
Silt SHAP value
Silt
DEM SHAP value
DEM
CEC SHAP value
CEC
PM SHAP value
PM
pH SHAP value
pH
KF SHAP value
KF
EMg SHAP value
EMg
LSTD SHAP value
LSTD
Low
High

Uncertainty analysis



(a) INLA-SPDE 95% Upper prediction limit

(b) INLA-SPDE 95% Lower prediction limit

(c) RF 95% Upper prediction limit

(d) RF 95% Lower prediction limit

5. Discussion & Conclusion

Limitations and implications

- Computation Time: While INLA-SPDE reduces modeling time compared to MCMC, it still takes significantly more time than RF due to fitting more hyperparameters, with only slight accuracy improvements.
- Scalability: The study focused on regional scale mapping, but further research is needed to explore INLA-SPDE's applicability for high-resolution national or continental scale mapping.
- Future Directions: Future studies should optimize mesh grid settings, explore parallel computing or GPU acceleration, and consider hybrid models like Regression Kriging to enhance performance.

5. Discussion & Conclusion

Conclusion

- INLA-SPDE model shows clearly better performance than classical geostatistical methods (OK and GWR) and slightly outperformed RF.
- INLA-SPDE has advantage on reporting posterior distribution of the model hyperparameters.
- This study demonstrated that INLA-SPDE is able to deal with large data size (16,050 samples), and could including large and various kinds of covariates (27 covariates) in the modelling process.
- By integrating INLA-SPDE and interpretable machine learning model, we can both quantify the overall importance of different variables for mapping soil properties and mapping the spatially varying primary variables for mapping soil properties across the study area at pixel level.

江西财經大學 JIANGXI UNIVERSITY OF FINANCE AND ECONOMICS | 财税与公共管理学院

Food and Agriculture Organization of the United Nations

THANK YOU

GLOBAL SOIL PARTNERSHIP