2024_“ShuWei Cup”

Problem B：Research on the Collaborative Estimation Method for Spatial Variables

In mathematical statistics, simple sampling often assumes that sampling points from the same population are independent of each other. In contrast to mathematical statistics, spatial statistics assumes that sampling points of spatial variables are dependent and exhibit certain trends in their values. Therefore, the values of a spatial variable and its trend can be used to estimate its values at unsampled locations. The most common method for spatial estimation is the Kriging algorithm.

In practical engineering, the same spatial variable can be measured using different methods. Although these methods measure the same physical quantity, the differences in the measurement principles lead to discrepancies in the measured values. However, these values still exhibit significant spatial correlation. Additionally, some spatial variables may have different physical meanings but show certain correlations and similarities. These spatially correlated variables can either have the same or different units of measurement.

Some spatial variables are costly and difficult to measure. Despite their high accuracy, only a limited number of samples are obtained in practical work. On the other hand, some spatial variables are easy to measure and inexpensive, leading to extensive sampling over a large area.

In engineering research, when studying the spatial variation patterns of spatial variables with insufficient sampling, researchers often refer to other spatial variables that are strongly correlated with the target variable and use collaborative estimation. In geological and mining research, the main method to address such problems is the Co-Kriging method. While the Co-Kriging algorithm is well-understood in theory, its implementation is difficult due to the challenge of calculating the cross-covariance or cross-variogram functions between the spatial variables, making it difficult to apply this algorithm in practice.

With the development of artificial intelligence and machine learning methods,

these techniques have been successfully applied in many industries. The problem provided offers two datasets, Attachment 1 and Attachment 2, which contain measurements of four spatial attributes from the same region, with the target variable having a higher measurement cost. The problem asks participants to research appropriate methods for the collaborative estimation of spatial attribute data given in the attachments:

Problem 1: Use the data in Attachment 1 to study the variation patterns of one of the spatial variables (F1_target variable).

（1）Randomly and uniformly resample the target variable, and estimate the spatial variable values at unsampled locations using the resampled values. Present the results as a contouring map.

（2）Vary the sample size and explore the relationship between sample size and estimation error.

Problem 2: Study the correlation between the target variable and collaborative variables using the data in Attachment 1. Select two collaborative variables as the estimating collaborative variables for the target variable.

Problem 3: Using the data in Attachment 1 and the findings from Problem 2, select one or two collaborative variables and study the variation patterns of the spatial variable (F1_target variable).

（1）Randomly and uniformly resample the target and collaborative variables, and estimate the spatial variable values at unsampled locations using the resampled values. Present the results as a contouring map.

（2）Vary the sample size and explore the relationship between sample size and estimation error.

（3）Select at least two methods and compare them.

Problem 4: The target variable (F2_target variable) in Attachment 2 has insufficient sampling data. Select the optimal method from Problem 3 to estimate the trend of the target variable and present the results as a contouring map.

Data Description:

1． The data in the attachments are sampled from a rectangular area. The rectangular structure is as follows:

（1） X-coordinate range: Column spans range from 51250.0000 meters to 64500.0000 meters;

（2） Y-coordinate range: Row spans range from 78750.0000 meters to 92000.0000 meters;

（3）The study area is divided into small grids of 50 meters by 50 meters, with a total of 266×266 grid points. Sampling values for the spatial variables are provided at these grid points.

2. The file "F2_target_variable_sampling.txt" in Attachment 2 contains 1000 sample points uniformly sampled from the grid points in the study area. For research purposes, samples can be extracted in sequence as needed. For example, to select 100 sample points, extract the first 100 points from the data; for 200 sample points, extract the first 200 points, and so on.