User Guide Book for Geographical Process Regionalization Tools

Overview

The first version of the Process-oriented Geographical Regionalization (PGR) toolbox comprises five tools:

1. Core Tool: "Geographical Process Regionalization Tool"
2. "Spatial Time-series Generation Tool" for creating the spatial time series required by the PGR tool
3. "Spatial Adjacency Matrix Generation Tool" for establishing spatial adjacency relationships
4. "One-to-Many Geographical Process Similarity Mapping Tool" for exploratory spatial time series data analysis
5. "Spatial Time-series Neighborhood Statistical Analysis Tool" for conducting statistical analyses on spatial time series neighborhoods

These tools collectively facilitate the implementation of process-oriented geographical regionalization, from data preparation to exploratory analysis and final regionalization. They provide a comprehensive suite for researchers and practitioners working with geographical processes and their spatial-temporal patterns.

1 Toolbox Overview

1.1 Tool set

The Geographical Process Regionalization Toolbox comprises three toolsets, each designed for a specific function: regionalization, construction of input parameters for regionalization, and exploratory analysis of geographical processes. Figure 1 illustrates these three toolsets contained within the toolbox.

This structure provides a clear and organized approach to process-oriented geographical regionalization, allowing users to efficiently access tools for data preparation, exploratory analysis, and the core regionalization process.

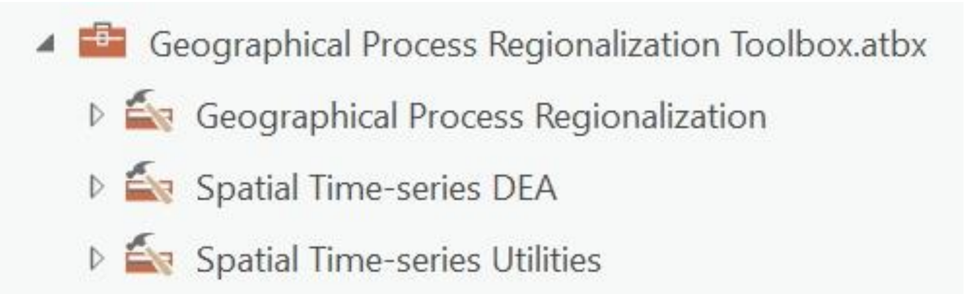


Figure 1. Toolsets of process-oriented geographical regionalization

1.2 Tool

The Geographical Process Regionalization Toolbox contains three toolsets encompassing a total of five tools: one core tool and four auxiliary tools. Figure 2 illustrates these five tools and their respective toolsets.

1. Geographical Process Regionalization Tool: This is the primary tool for executing geographical process regionalization tasks. It takes spatial time series, spatial adjacency matrices, and other necessary parameters as inputs to perform regionalization and output the results.
2. Spatial Time-series Generation Tool: This tool provides standardized spatial time series data for the regionalization tool. It converts raw spatiotemporal data, which often doesn't meet the format requirements for geographical process regionalization, into suitable spatial time series data.
3. Spatial Adjacency Matrix Generation Tool: Geographical process regionalization requires spatial proximity constraints. This tool generates adjacency matrix files that meet the regionalization requirements based on the polygon files corresponding to the spatial time series.
4. One-to-Many Geographical Process Similarity Mapping Tool: This tool assesses the similarity between a selected reference spatial time series unit and all other units. It normalizes the similarity values to a 0-1 range for each spatial unit, reflecting the degree of similarity between each unit and the reference unit. This tool is used to explore spatial dependence in geographical processes, providing a simple measure of the "closer proximity, higher similarity" principle.
5. Spatial Time-series Neighborhood Statistical Analysis Tool: This tool performs statistical analysis and visualization of the mean and variance of process similarities between a spatial location and its neighbors. If a location's geographical process is very similar to its surroundings, the mean similarity with its neighbors will be very small, while the variance will be very large.

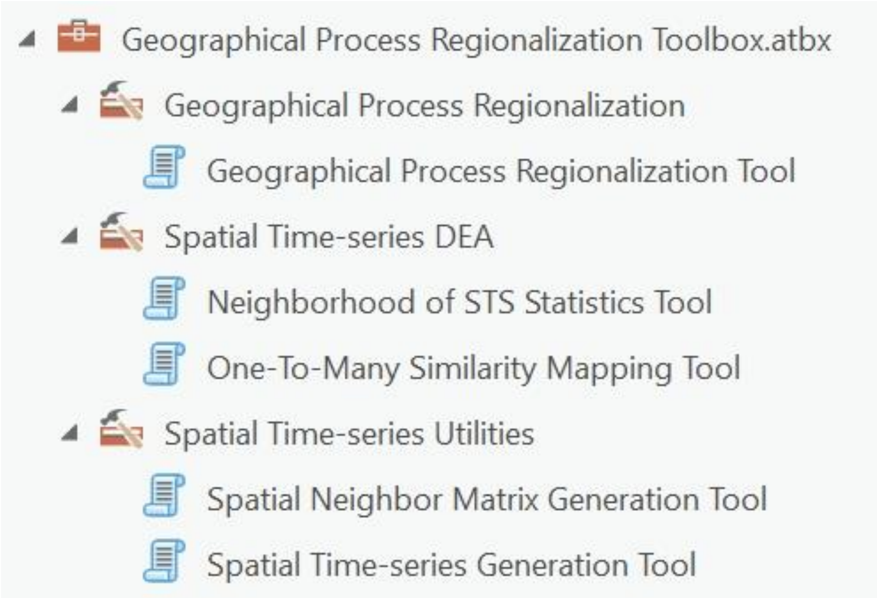


Figure 2. Tools in PGR Toolbox

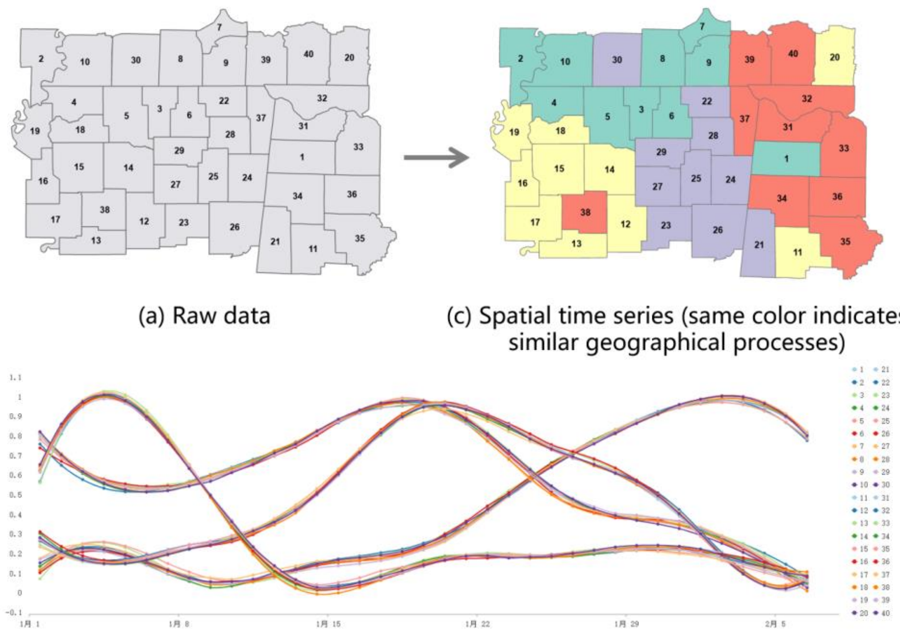
1.3 Operating Environment and Usage

The Geographical Process Regionalization Toolbox is implemented using Python 3.X, with the UI and tool framework based on ArcGIS Pro 3.3. These tools have been tested and confirmed to function correctly in ArcGIS Pro 3.X. It is recommended to use them within ArcGIS Pro 3.X. To use the toolbox, link to its location and find it in the "Catalog" panel. Double-click on any tool to run an analysis task.

Additionally, if you have programming experience and GIS knowledge, you can directly use the source code provided in the article. With appropriate configuration and restructuring, you can build your own tools without relying on ArcGIS or any third-party GIS platform. A Python interpreter is sufficient.

1.4 Experimental Data for Learning and Testing Tools

The Geographical Process Regionalization Toolset includes a dataset for learning to use the tools, testing them, and helping users understand geographical process regionalization. The data effects are shown in Figure 3. This dataset contains 40 spatial units, each corresponding to a time series, collectively forming a dataset of 40 spatial time series. By default, executing the regionalization tool with a configuration of 4 regions yields optimal results. However, this does not mean that 4 must be used as the threshold in specific operations. Understanding the mechanism of geographical regionalization, one realizes that theoretically, any threshold less than the total number of spatial time series can be used. Setting more than 4 regions will result in the model further dividing similar spatial time series into more regions based on spatial proximity constraints. A similar principle applies if the number of regions is set to less than 4.



(a) Raw data

(c) Spatial time series (same color indicates similar geographical processes)

(b) Spatial time-series of each basic areal unit

Figure 3. Test data for PGR tools

2 Tool Usage Instructions

2.1 Geographical Process Regionalization Tool

The Geographical Process Regionalization Tool performs Process-oriented Geographical Regionalization (PGR) tasks based on the Spatial Time Series (STS) formed by given regional units and their time series data. The tool interface is shown in Figure 4 and includes 8 input parameters. The meaning of each parameter is as follows:

① Area feature layer: A polygon feature layer used to determine the basic spatial analysis units and spatial adjacency relationships between units.

② Area ID field: The field in the Area feature layer that uniquely identifies each polygon.

③ Spatial time-series: Spatial time series data including polygon IDs, in csv file format, which can be generated using the "Spatial Time Series Generation Tool".

④ Spatial neighbor matrix data: Spatial adjacency matrix data in csv file format, which can be generated using the "Spatial Adjacency Matrix Generation Tool".

⑤ Similarity method: The method used to measure the similarity between two geographical processes (STS) during regionalization, with Euclidean distance as the default. The main methods include:

I. Value similarity based on Euclidean distance

II. Value similarity based on DTW

III. Shape similarity

IV. Periodicity similarity based on Fourier transform

⑥ Number of regions: Used to set the number of regions in the regionalization result, to be customized by the user.

⑦ Minimum number of regions: Used to set the minimum number of basic polygons within each region in the regionalization result.

⑧ Storage path for the regionalization result feature class.

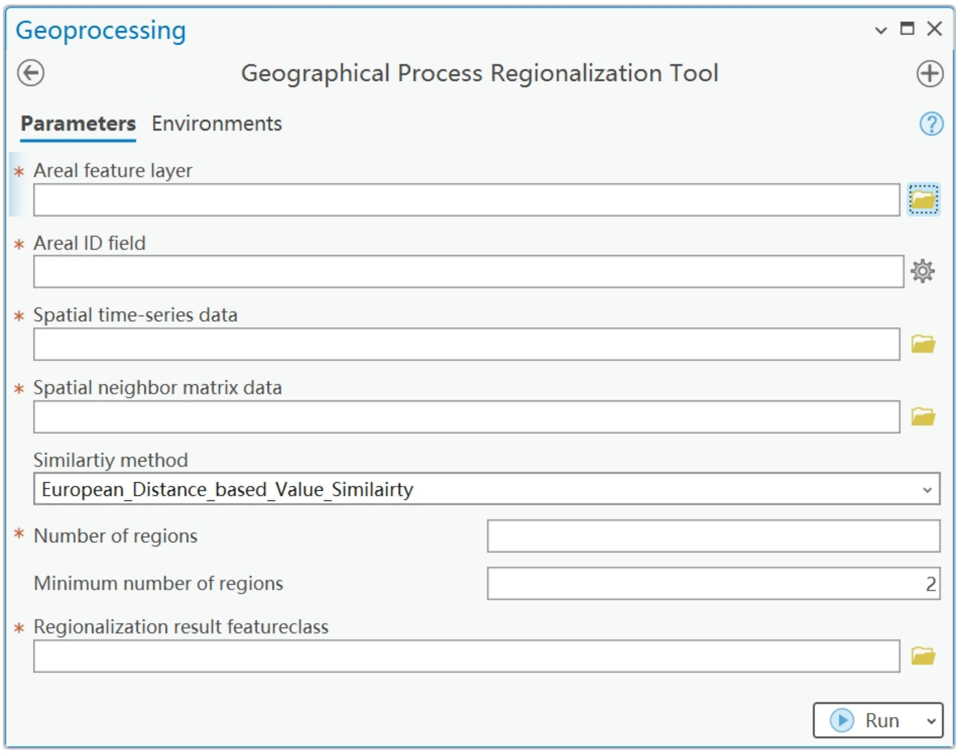


Figure 4. UI of Geographical Process Regionalization Tool

Based on value similarity of simulated data, geographical process regionalization is performed using the Euclidean distance method. The threshold for the number of regions is set to 4 and 5 respectively, with results shown in Figure 5. Additionally, regionalization tasks can be executed with different quantity thresholds using the provided datasets for shape similarity, periodicity similarity, and others.

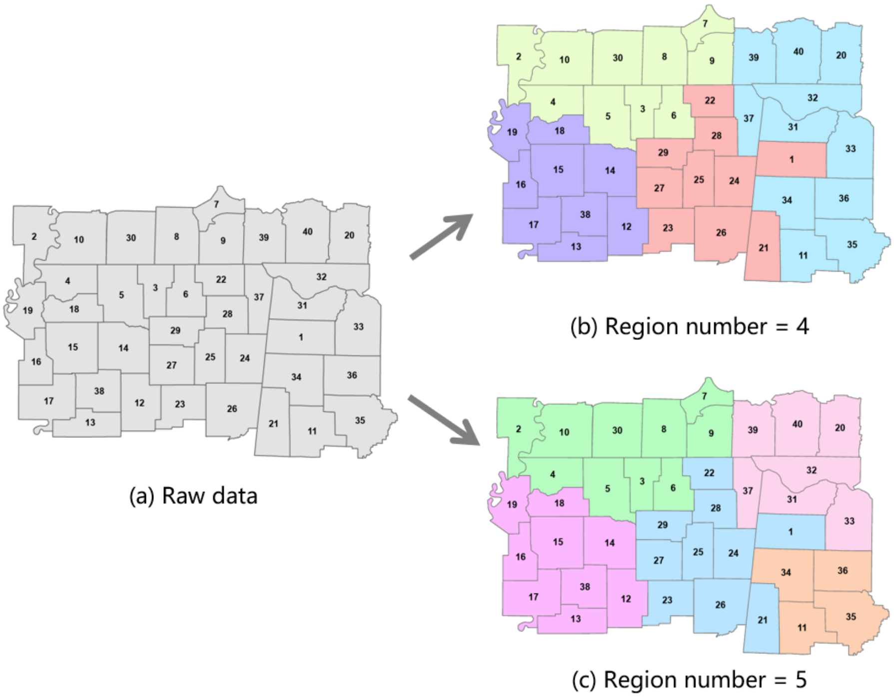


Figure 5. Results of Geographical Process Regionalization Based on Value Similarity

Comparing the data with predefined labels to the results after regionalization, it is evident that our model produces excellent regionalization results (Figure 6). Units with geographical processes that are significantly different from their surrounding areas are merged into their respective regions. Our quantity constraint also serves to avoid the problem of single or few units forming regions as much as possible.

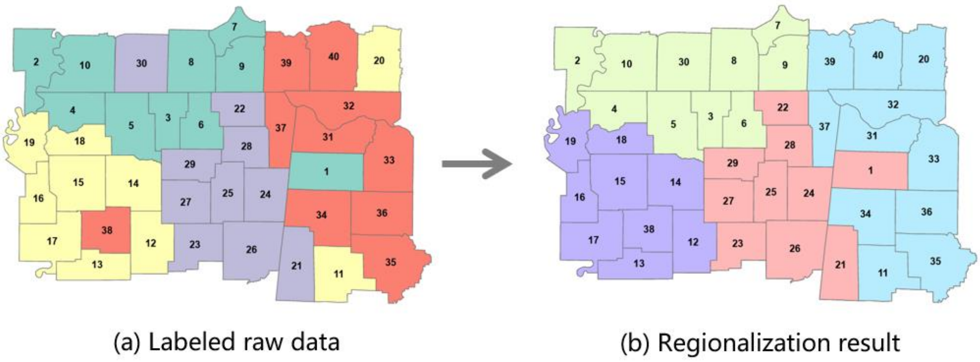


Figure 6. Predefined Label Data and Model Execution Results

2.2 Geographical Process DEA: One-to-Many Similarity Mapping Tool

The One-to-Many Similarity Mapping Tool is used for exploratory assessment of the degree of similarity between the geographical processes of one region and all other regions. As explained in our theoretical paper on geographical process regionalization, when geographical processes lack or have very weak proximity similarity, the regionalization results lose their significance. This tool explores the degree of proximity similarity in the simplest and most direct way. It's worth noting that for a single variable, direct visual mapping is possible, but geographical processes cannot be directly visualized.

The One-to-Many Similarity Mapping Tool task is executed based on given Spatial Time Series (STS). The tool interface is shown in Figure 7 and includes 6 input parameters. The meaning of each parameter is as follows:

① Areal feature layer: A polygon feature layer used to determine the basic spatial analysis units and spatial adjacency relationships between units.

② Area ID field: The field in the Area feature layer that uniquely identifies each polygon.

③ Focal area ID: Identifies the polygon from the Areal feature layer to be used as the reference target. This is achieved using the Area ID of the target polygon, which needs to be manually entered by the user.

④ Spatial time-series: Spatial time series data including polygon IDs, in csv file format, which can be generated using the "Spatial Time Series Generation Tool".

⑤ Similarity method: The method used to measure the similarity between two geographical processes (STS), with Euclidean distance as the default.

⑥ Output similarity mapping layer: Outputs the analysis results. The value for the target polygon (or focal area) will be the maximum, as it is completely identical when compared to itself.

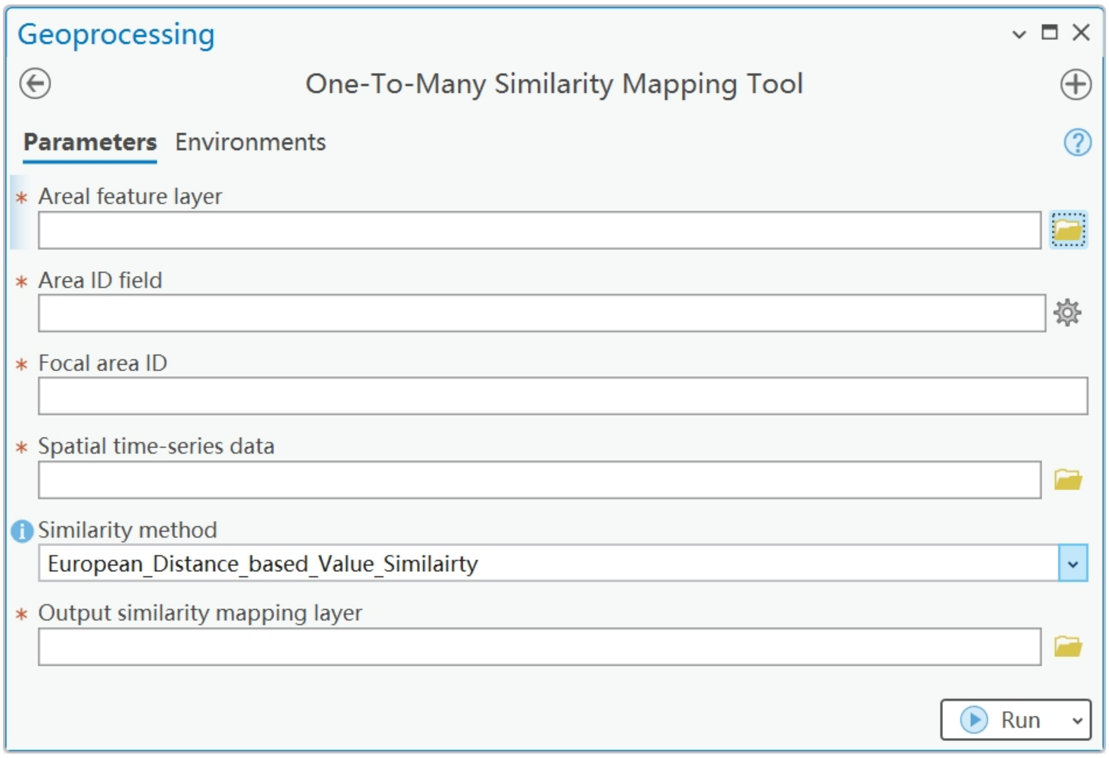
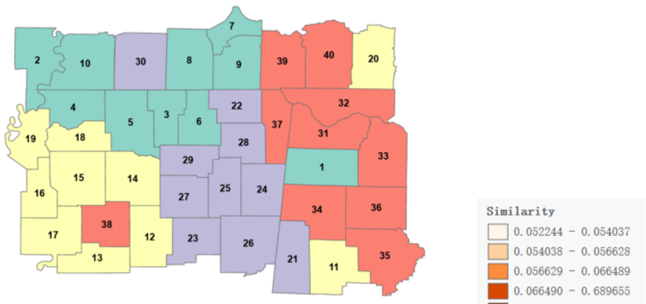
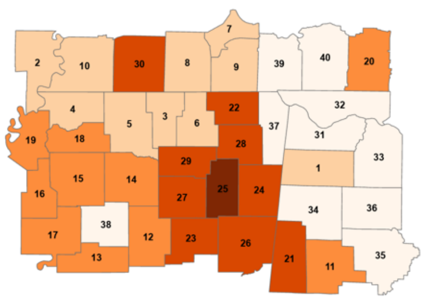


Figure 7. One-to-Many Similarity Mapping Tool

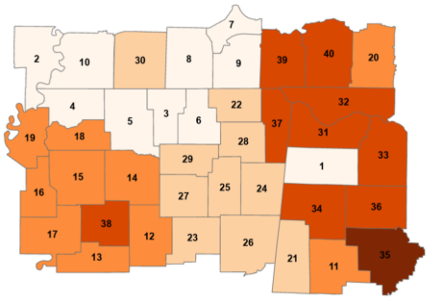
Here, we take area IDs 25 and 35 as examples to execute this tool based on value similarity. The results are shown in Figures 8(b) and 8(c) respectively. Comparing these with the original labeled data shown in Figure 8(a), we can see the rationality of the analysis results. At the same time, the results demonstrate that this method plays a simple yet highly effective role in exploratory data analysis of geographical processes.



(a) Labeled raw data



(b) Area 25 as the reference unit



(b) Area 35 as the reference unit

Figure 8. Demonstration of One-to-Many Similarity Mapping Tool Application Cases

2.3 Geographical Process DEA: Neighborhood of STS Statistical Mapping Tool

The Neighborhood Similarity Statistical Mapping Tool is used for exploratory regionalization to assess the degree of similarity or consistency between the geographical processes of each polygon and its neighboring polygons. The tool interface is shown in Figure 9. This tool provides two statistical methods: mean and variance. The statistical principle is to calculate the similarity of geographical processes between any given polygon and its neighboring polygons, then use either the mean or variance of these similarities as the statistical value for that polygon. The meaning of each parameter is as follows:

① Areal feature layer: A polygon feature layer used to determine the basic spatial analysis units and spatial adjacency relationships between units.

② Spatial time-series: Spatial time series data including polygon IDs, in csv file format, which can be generated using the "Spatial Time Series Generation Tool".

③ Spatial neighbor matrix data: Spatial adjacency matrix data in csv file format, which can be generated using the "Spatial Adjacency Matrix Generation Tool".

④ Similarity method: The method used to measure the similarity between two geographical processes

(STS), with Euclidean distance as the default.

⑤ Statistical type: Used to determine the type of statistics for geographical process similarity, including mean and variance.

⑥ Output feature class: Outputs the statistical analysis results, with a new field "Mean" for mean statistics and "Variance" for variance statistics.

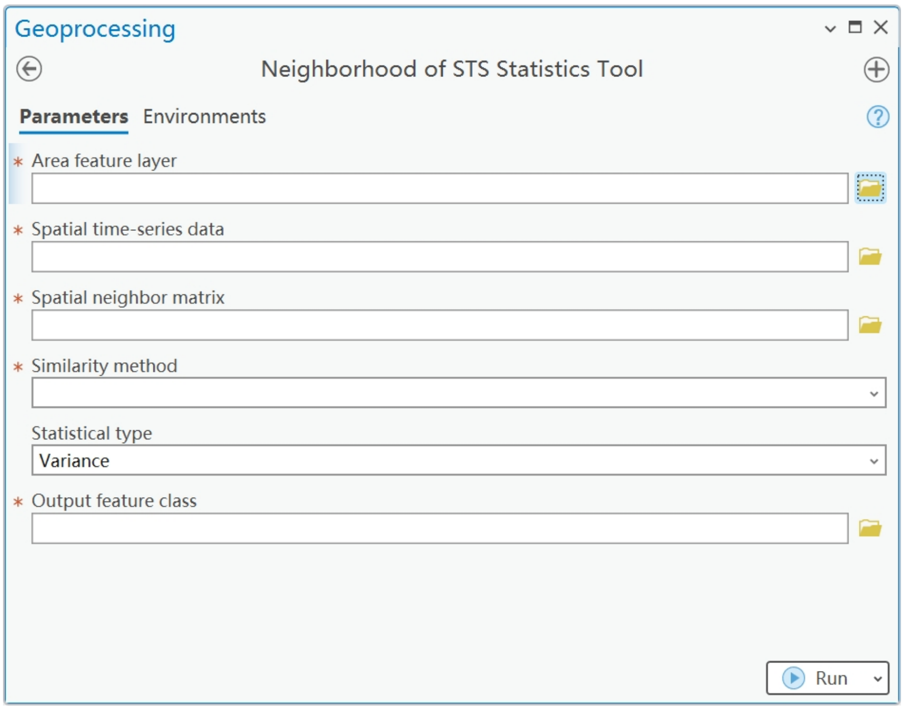
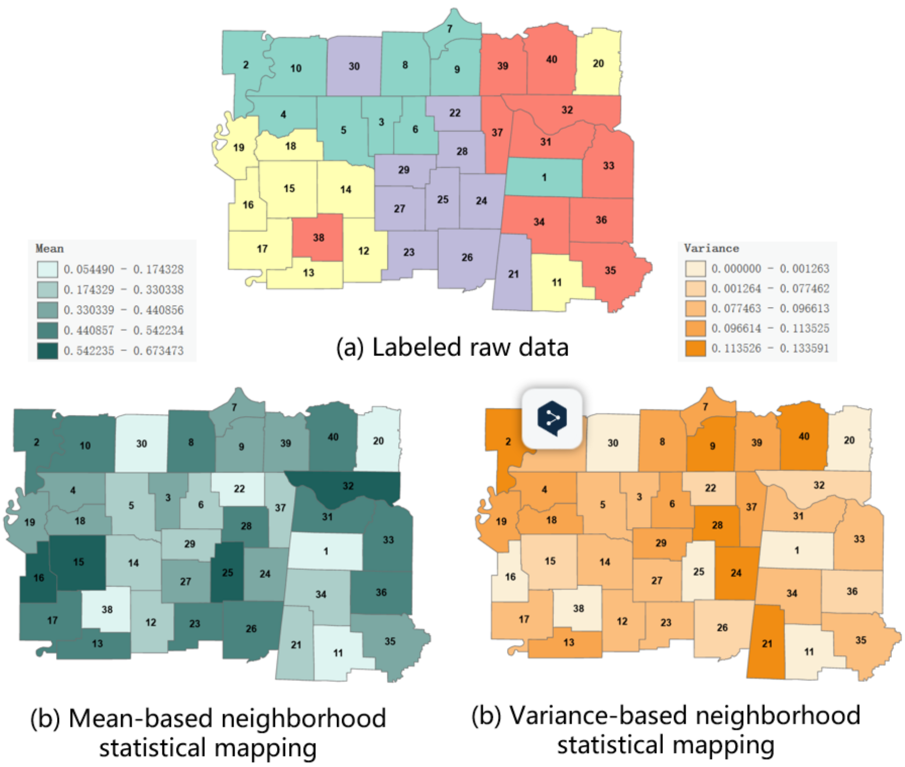


Figure 9. Neighborhood Similarity Statistical Mapping Tool

Based on value similarity, this tool is executed using both mean and variance statistics, with results shown in Figure 10. Essentially, both statistical methods reflect similar statistical characteristics. When the mean statistic for a polygon is large, it indicates that there is a significant difference between its geographical process and those of its surrounding neighbors. The variance can be understood similarly, though variance reflects the cumulative level of difference, while the mean reflects the average level of difference.



(a) Labeled raw data

(b) Mean-based neighborhood statistical mapping

(b) Variance-based neighborhood statistical mapping

Figure 10. Demonstration of Neighborhood Similarity Statistical Mapping Tool Application Cases

2.4 Spatial Time-Series Generation Tool

The Spatial Time Series Generation Tool includes two functions:

1. Generate standardized spatial time series data required for the geographical process regionalization tool
2. Scale the time resolution of time series

To enhance the ease of use of the geographical process regionalization tool, this tool uses CSV as both input and output file format. As long as the user's spatiotemporal data contains columns for location-associated fields, time series, and observation values, this tool can convert the spatiotemporal data into spatial time series data files suitable for geographical process regionalization.

Scaling temporal resolution is also a common application requirement. For example, if the original data is on a daily basis, but the final spatial time series needs to be on a monthly basis, this can be achieved through a scaling factor.

The tool interface is shown in Figure 11. The meaning of each parameter is as follows:

① Input spatiotemporal data: Original spatiotemporal data, which needs to include location ID, observation value, and time fields. Note: When using spatial time series data generated by this tool as input for scaling or other purposes, a row name field needs to be added to the input data.

② Region ID field: The column field in the Input spatiotemporal data that uniquely identifies each polygon (spatial time series).

③ Value field: The column field in the Input spatiotemporal data used to construct the spatial time series observation values.

④ Datetime field: The date field in the Input spatiotemporal data.

⑤ Scale rate: The scale rate field is used to determine the degree of temporal resolution scaling. If no scaling of dates is required, set this value to 1. Otherwise, set it to any value greater than 0 and less than 1. For example, if the original data is daily and you want a 10-day interval, set this value to 0.1.

⑥ Output spatial time-series: The standardized format or scaled spatial time series data generated.

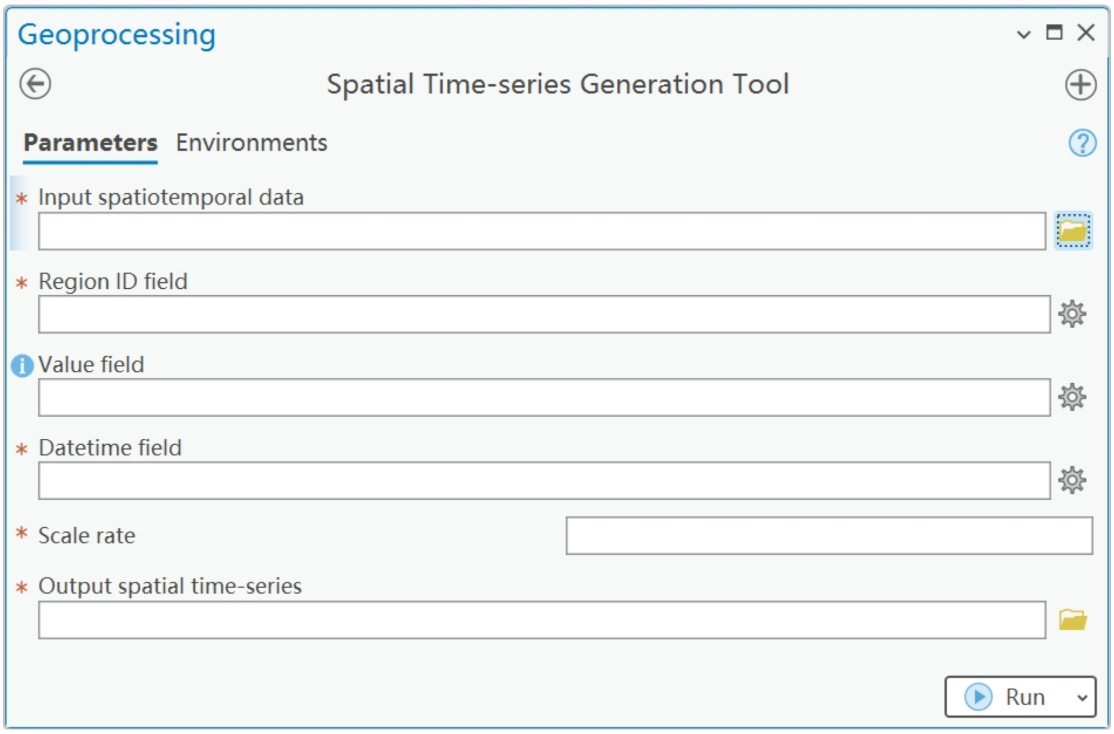


Figure 11. Spatial Time-series Generation Tool

2.5 Neighborhood Matrix Generation Tool

The Neighborhood Matrix Generation Tool is used to create the neighborhood matrix required for executing the geographical process regionalization tool. Since geographical regionalization must ensure spatial continuity within each region, "having common edges or common nodes" is considered the only adjacency rule and is set as the default for this tool. This version of the geographical process regionalization tool only supports polygon data and its time series, so no other rules are needed. If the original spatiotemporal data consists of point data and its time series, Thiessen polygons can first be constructed based on the points, and then these Thiessen polygons can be used as polygon data for geographical process regionalization.

The interface of this tool is shown in Figure 12, and the meanings of the input parameters are as follows:

① Input areal feature layer: A polygon feature layer used to determine the basic spatial analysis units and spatial adjacency relationships between units.

② Region ID field: The column field in the Input areal feature layer that uniquely identifies each polygon (spatial time series).

③ Output spatial neighbor matrix file: The output file containing the generated spatial adjacency matrix.

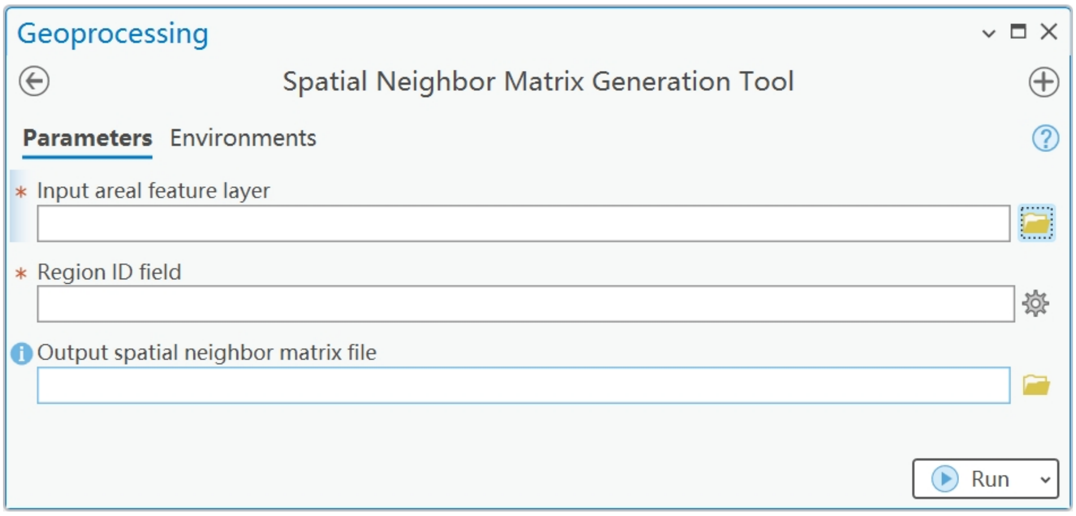


Figure 12. Spatial Neighborhood Matrix Generation Tool

3. Summary

Indeed, the importance of both core and auxiliary tools in geographical process regionalization. Let me summarize and expand on your statement:

When it comes to the primary function of geographical process regionalization, the "Geographical Process Regionalization Tool" alone is sufficient to execute this task. However, more often than not, exploratory data analysis and data processing to meet regionalization requirements are necessary steps.

Although the latter four tools are not part of the core toolset for geographical process regionalization, they play a crucial role in executing the regionalization process effectively. These auxiliary tools facilitate data preparation, exploration, and analysis, which are essential steps before applying the core regionalization algorithm.

To recap, the toolset currently consists of:

1. One core tool:
   - Geographical Process Regionalization Tool
2. Four auxiliary tools:
   - One-to-Many Similarity Mapping Tool
   - Neighborhood Similarity Statistical Mapping Tool

- Spatial Time Series Generation Tool
- Neighborhood Matrix Generation Tool

These auxiliary tools support tasks such as similarity analysis, statistical mapping, data formatting, and spatial relationship definition, which are all important precursors to the actual regionalization process.

The provision of both core and auxiliary tools creates a more comprehensive and user-friendly toolkit for researchers and practitioners working with geographical process regionalization. It acknowledges the complexity of the process and provides support at various stages of the analysis.

As you mentioned, both the core and auxiliary tools are expected to continue expanding in future versions. This ongoing development will likely enhance the capabilities and flexibility of the toolkit, allowing for more sophisticated analyses and accommodating a wider range of research needs.

The anticipation for these future developments is shared among the user community, as each new tool or improvement has the potential to unlock new insights in geographical process analysis and regionalization.

If you'd like to understand the detailed principles behind geographical process regionalization, please refer to the following literature:

Haiping Zhang, Xingxing Zhou, Haoran Wang, YuYang, Xinyue Ye, Guoan Tang. Advancing process-oriented geographical regionalization model. *Annals of the American Association of Geographers*, (2024). DOI Link: https://doi.org/10.1080/24694452.2024.2380893

Contact us: zhanghaiping@igsnrr.ac.cn