Manuals for GeoCA

Version 2.4

Authors: Ruifan Wang, Yao Yao

High-performance Spatial Computation Laboratory
School of Geography and Information Engineering
China University of Geosciences (Wuhan)

Nov. 9th, 2020

Catalogue

Introduction to functions 3
Introduction to function modules 4
Set the module according to the input Simulation data 4
Function description 4
Interface description 4
Output Settings 7
Function description 7
Interface description 7
Simulation Settings 10
Function description 10
Interface description 10
CA Iteration Parameters Setting 12
Rule-mining Settings 13
Function description 13
Interface description 13
Validation Settings 16
Function description 16
Interface description 16
Land Erosion Mapping 18
Function description 18
Interface description 18
Load and Export Params and Run 20
Function description 20
Interface description 20
Process Log 21
Function description 21
Interface description 22
Test data, parameters and output results 22
Description of files 24
Input file 24
Land use data file 24
Land use data code file 24
Land use reclassification data file 25
Spatial auxiliary data file 25
Additional probability data file 25
Output file 26
Urban conversion file 26
Reclassification of land use files 26
Development Suitability file 26
Random point file for regression analysis (data mining) 27
Urban land use conversion probability file(non urban to urban) 27
9-band conversion probability file 27

Simulation result file....28
Rule mining precision report file....28
Model validation file....28
Source / Target city reclassification file....28
Simulation result file....29
Output precision file....29
Land erosion mapping file....30
Land use LULC file....30
Simulation result file....30
Output mapping file....30
Log file....31
Parameter file....33
Contact....33
References....34

Introduction to GeoCA

The geographical process, especially the process of urban development, is extremely complex. In order to accurately simulate the urban development process and predict the future urban form, we provide a set of geographic simulation app via cellular automata (GeoCA).

GeoCA is a set of software used to simulate urban development process from large-scale pixel level (grid level). GeoCA has a complete set of high-performance massive spatial data reading and writing infrastructure, which can support multi-threaded reading and writing, automatic memory allocation and spatial data coordinate automatic alignment.

GeoCA encapsulates the urban simulation model supported by coupling cellular automata and multi machine learning model, which has been widely used in the analysis of urban development process, land erosion, ecological environment and urban planning.

GeoCA includes the following functional modules: input simulation data, output simulation results, simulation parameter setting, rule mining parameter setting, simulation result check, land erosion mapping, model execution, and processing log output. The design pattern of data loose coupling is adopted between modules.

GeoCA software is developed by the high performance space computing intelligent laboratory of China University of Geosciences (Wuhan) (HPSCIL@CUG). The copyright belongs to China University of Geosciences (Wuhan) high performance space computing intelligent laboratory (HPSCIL@CUG).

The theory of urban simulation model based on cellular automata are attached with references at the end of the instruction, and relevant materials can also be downloaded here:http://www.urbancomp.net/2020/10/01/traditional-ca/

The function, algorithm and operation steps of each module are described in below.

Introduction to function modules

Set the module according to the input Simulation data

Function description

The function of the module is to set the input data of the model.
The input data includes three parts

1. Source filepath
2. Landcover files
3. Auxiliary files
4. Additionally probability file

Interface description

The interface is shown in Figure 1.
The input is divided into four parts: source filepath (required), landcover files (required), auxiliary files (required) and additional probability file (optional).

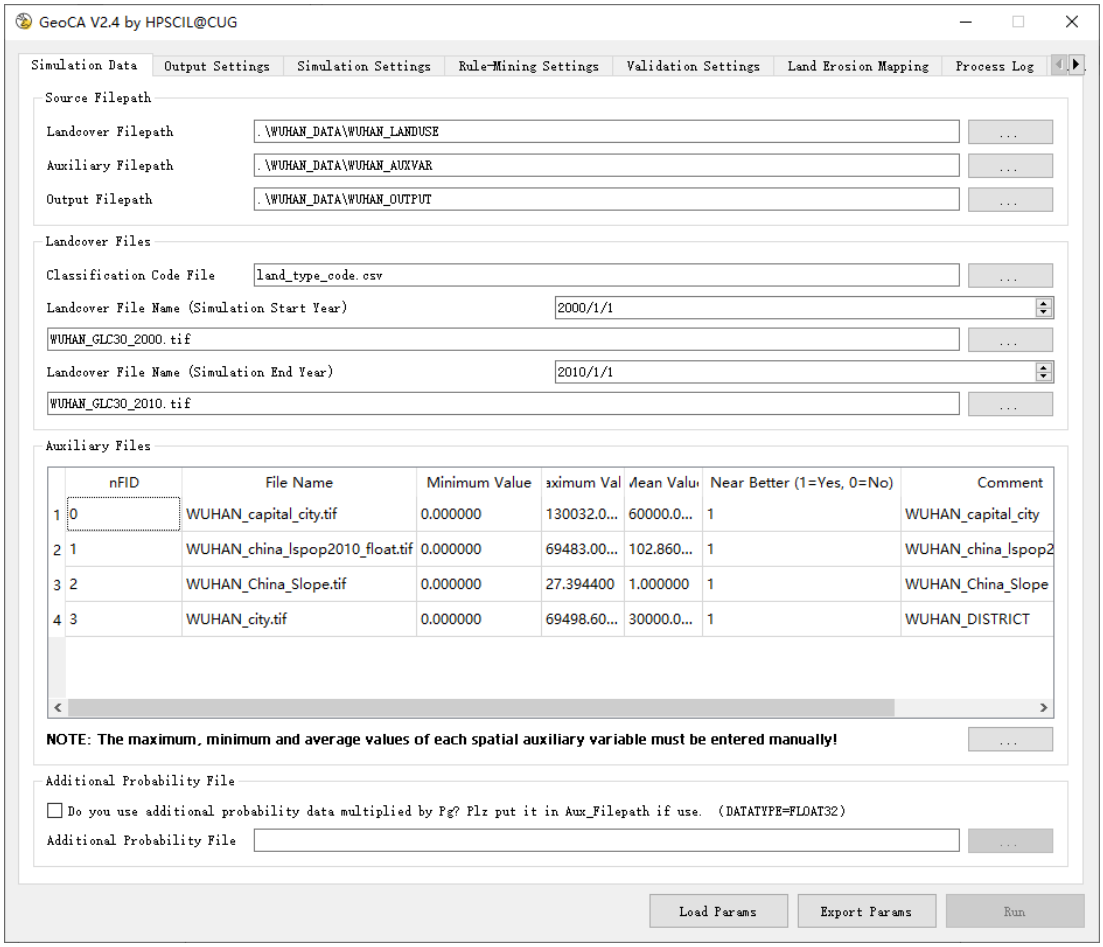

Figure 1. Input simulation data interface.

Source Filepath

Landcover Filepath：the folder path of landcover documents. It is recommended that all landcover documents should be placed in this folder.
Auxiliary Filepath：the folder path of auxiliary documents. It is recommended that all auxiliary documents should be placed in this folder.
Output Filepath：the folder path of output files. The program will place all output files in this folder.

Landcover Files

Classification Code File: The land use reclassification file is placed in the landcover filepath. The file type is TXT or CSV text file. Each class stores DN, NAME, TYPE_ ID, RE_ TYPE, RE_ TYPE_ ID, COMMENTS attributes for each land use and land cover type。
Landcover File Name (Simulation Start/End Year): landcover files for simulation start / end year, GeoTIFF format. Year and file name are required.

Auxiliary Files

The button can open auxiliary files in batches.
Each data is in Geotiff format, and the recommended data type is FLOAT32.
Each auxiliary data is displayed in the form of rows, and nFID is the unique identification code of auxiliary data.
In order to normalize the data, you need to manually modify the minimum value, maximum value and mean value of each auxiliary data. The default minimum value is 0 and the maximum value is 1. The mean value is 0.5, which can be modified as needed.
Whether the closer is better: 1 for "Yes" and 0 for "No". If you fill in 0, the normalization will be in reverse order.
Comments: optional.
Theoretically, the number of auxiliary data does not exceed the number of sampling random points (the default is 5000).

Additionally Probability File

This file is used to mine the probability of urban growth rules and multiply another probability value.
The file can be used or not. The data type is FLOAT32 and the data range is not limited. If this probability is used, Simulation Settings > > CA Iteration Settings > > Minimum Transition Probability Threshold is also recommended to be modified according to the value range of the file.
The additional probability data storage location should be located in Simulation Data > > Auxiliary Filepath.
Tip: this file can be used to control whether growth is allowed (data is 0 or 1), or to control (increase or decrease) the probability of a certain feature being eroded.

Output Settings

Function description

The function of the module is to set the output files.

Interface description

The interface is shown in Figure 2.
The output results of GeoCA mainly include:

1. Urban Change File Name (Geotiff format, BYTE type)
2. Reclassification LULC (Start/End Year) File Name (Geotiff format, BYTE type)
3. Development Suitability File Name (Geotiff format, BYTE type)
4. Random Points for Regression File Name (Text data)
5. Land-use Change Probability (N-U) File Name (Geotiff format, FLOAT32 type)
6. Iteration Results (No File Extension required) (Geotiff format, BYTE type)
7. Nine-band Probability File (optional, Geotiff format, FLOAT32 type)

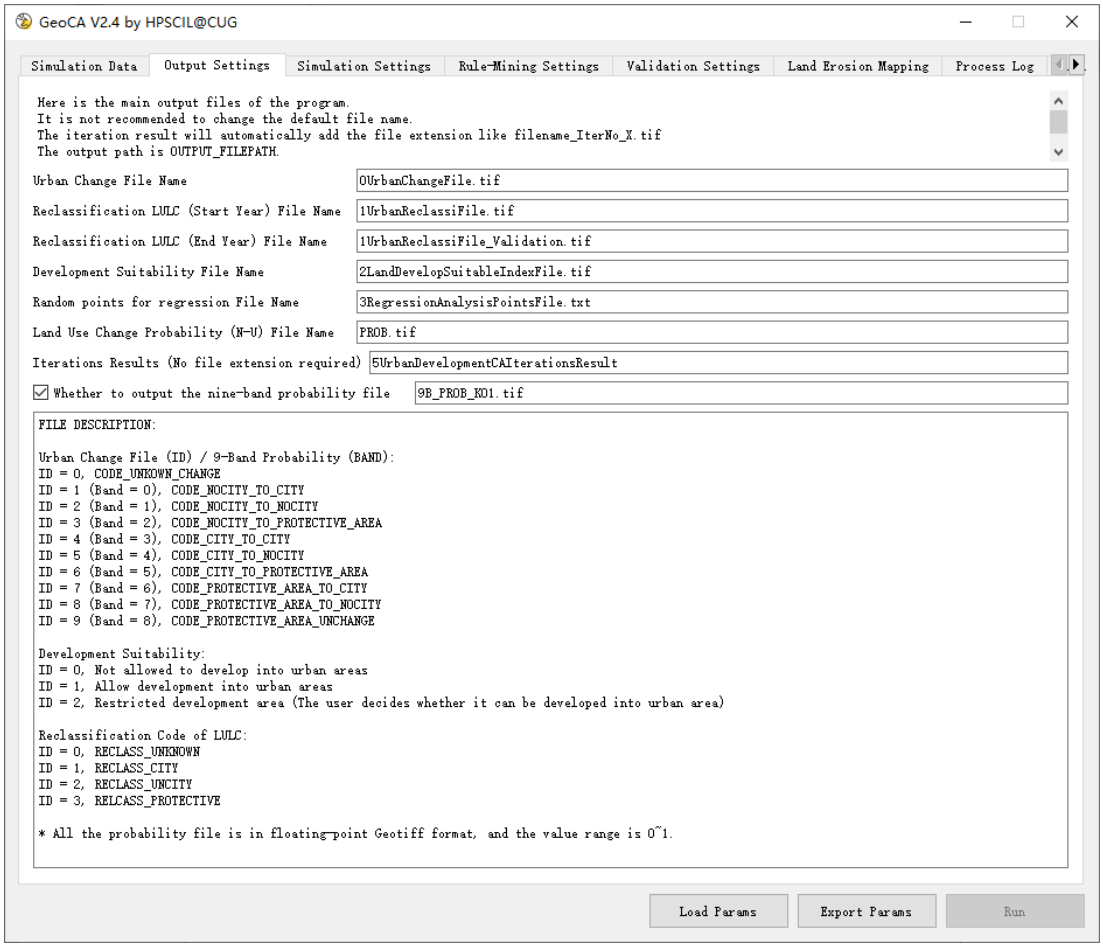

Figure 2. Output settings interface.

Output filepath

All data will be output in the Source Filepath >> Output Filepath folder.

Urban Change File Name

The difference in land use type conversion between start year and end year of statistics. The extension type is tif, the file type is Geotiff, and the data type is BYTE.

Reclassification LULC (Start/End Year) File Name

The results of reclassification of land use data from the start and end years according to the land conversion code settings in "Simulation Data >> Classification Code File".
The extension type is tif, the file type is Geotiff, and the data type is BYTE.

Development Suitability File Name

According to the land conversion code set in the "Simulation Data >> Classification Code File", each grid in the study area is set to the suitability of the development for the city (permitted development, non-permitted development, restricted development areas).
The extension type is tif, the file type is Geotiff, and the data type is BYTE.

Random Points for Regression File Name

With 9 different conversion types, random sampling is performed on land use conversion and auxiliary spatial data, and stored as random point-conversion type data. It can be used to check the data later and test the accuracy of the data mining model.
This file is a text file that stores the longitude, latitude, auxiliary data set and conversion type of each random point.

Land-use Change Probability (N-U) File Name

According to the obtained random points and the rule mining model selected by the user, the Pg probability map file of non-urban to urban conversion is obtained.
The extension type is tif, the file type is Geotiff, the data type is BYTE, the numerical range is 0~1.

Iteration Results (No File Extension required)

According to the CA model and simulation settings, the urban development simulation iteration result file is obtained.
Note that there is no need to set the file extension name, the default file name is 5UrbanDevelopmentCAIterationsResult.

The iteration result will be automatically named 5UrbanDevelopmentCAIterationsResult_IterNo_*.tif according to the number of iterations, where * is the number of iterations.

Nine-band Probability File

The file is large, and whether to output is optional. It is not output by default.
According to the obtained random points and the rule mining model selected by the user, 9 types of converted Pg probability map files are obtained.
The extension type is tif, the file type is Geotiff, the data type is BYTE, the numerical range is 0~1, and it has 9 bands.

Simulation Settings

Function description

The function of the module is to set simulation parameters and CA iteration parameters.

Interface description

The interface is shown in Figure 3.
It mainly includes two settings: simulation parameters setting and CA iteration parameters setting.

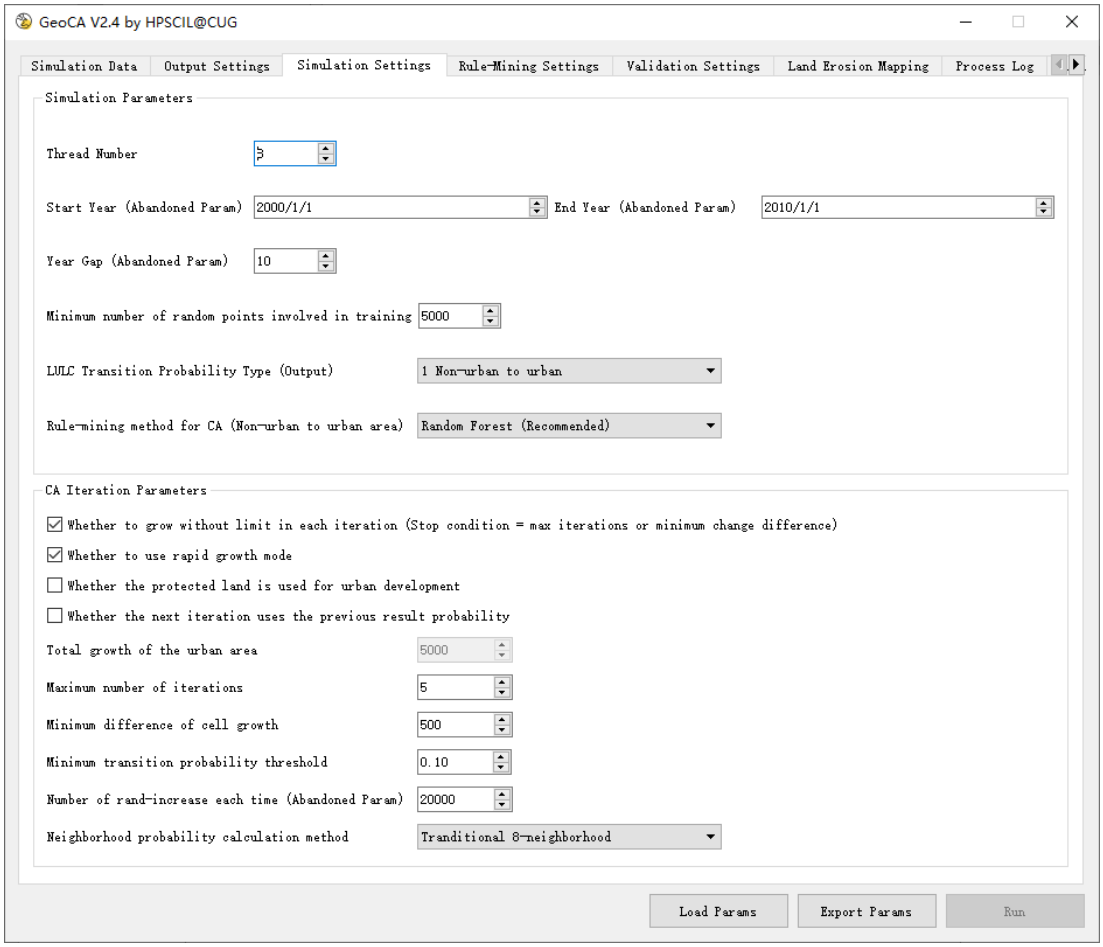

Figure 3. Simulation settings interface.

Simulation Parameters Setting

Thread Number: Set the number of multithreading. The program will automatically get the number of multithreading that can be used. Whether to use multithreading is up to the user. If multithreading is not used, the entry is set to 1.

Start Year、End Year、Year Gap: These three parameters are obsolete and do not need to be set.

Minimum number of random points involved in training: the minimum number of random points collected by the program. In general, the program will collect 2 times random points to enter the rule mining process.

LULC Transition Probability Type: The Pg probability type is 1 by default (non city to city). Do not modify to other types if not necessary.

Rule-mining method for CA: The rule mining methods are used to calculate the Pg probability. At present, GeoCA provides five models: random forest model (recommended), neural network model (recommended), logic model (recommended), logic regression (not recommended) and immune algorithm (not recommended).

CA Iteration Parameters Setting

Total growth of the urban area: Set the number of cells for urban growth.
Maximum number of iterations: If the number of iterations exceeds the value, stop iteration.
Minimum difference of cell growth: If the cell growth number between this and last iteration is less than the value, the iteration will be stopped.
Minimum Transition Probability Threshold: Non urban development is the minimum development probability of a city. The probability formula is P = Pg * Ω * RA, where Ω is the neighborhood probability and RA is the normal random number.
Number of rand-increase each time: Obsolete parameter, no need to set.

Neighborhood Probability Calculation Method:
GeoCA provides three kinds of neighborhood probtability calculation models:
Tranditional 8-neighborhood：traditional 8-neighborhood probability (City is 1, non city is 0)
8-neighborhood with Probability：improved 8-neighborhood model considering non urban development probability (city is 1, non city is Pg)
Tan-curve model：Neighborhood probability based on tangent model (Tan (0.25 * number of neighboring cities - 1.2) + 3.0) / 4.0-0.09

Whether to grow without limit in each iterations:
√ If this box is checked:
Total urban growth will be unlimited (0 in parameter file).
GeoCA will transform every cell that can grow into a city.
The iteration stop condition is the maximum number of iterations or the minimum growth interval difference .

○ If this box is unchecked:
The total urban growth is set for users.
The number of each growth = Total urban growth / Maximum number of iterations.
The minimum growth interval difference is not the iteration stop condition.

Whether to use rapid growth mode:
√ If this box is checked:
In the case of a small amount of data, directly count the increasing city data.
Use with unlimited development.

○ If this box is unchecked:
According to the total urban growth and the total number of urban expansion to simulate.
Use with unlimited development.

Whether the protected land is used for urban development:
√ If this box is checked: Protected land will develop into urban land in the iterative process.
○ If this box is unchecked: Protected land remains unchanged.

Whether the next iterations uses the previous result probability:
Not recommended.
√ If this box is checked: After each iteration, the overall development probability Pg will be replaced by P, and the calculation convergence speed of P will be accelerated.
○ If this box is unchecked: The overall development probability Pg remains unchanged.

Rule-mining Settings

Function description

The function of the module is to set the machine learning model parameters of the overall development probability Pg.

Interface description

The interface is shown in Figure 4.
According to the selected machine learning model, GeoCA will automatically open the corresponding parameter setting area.

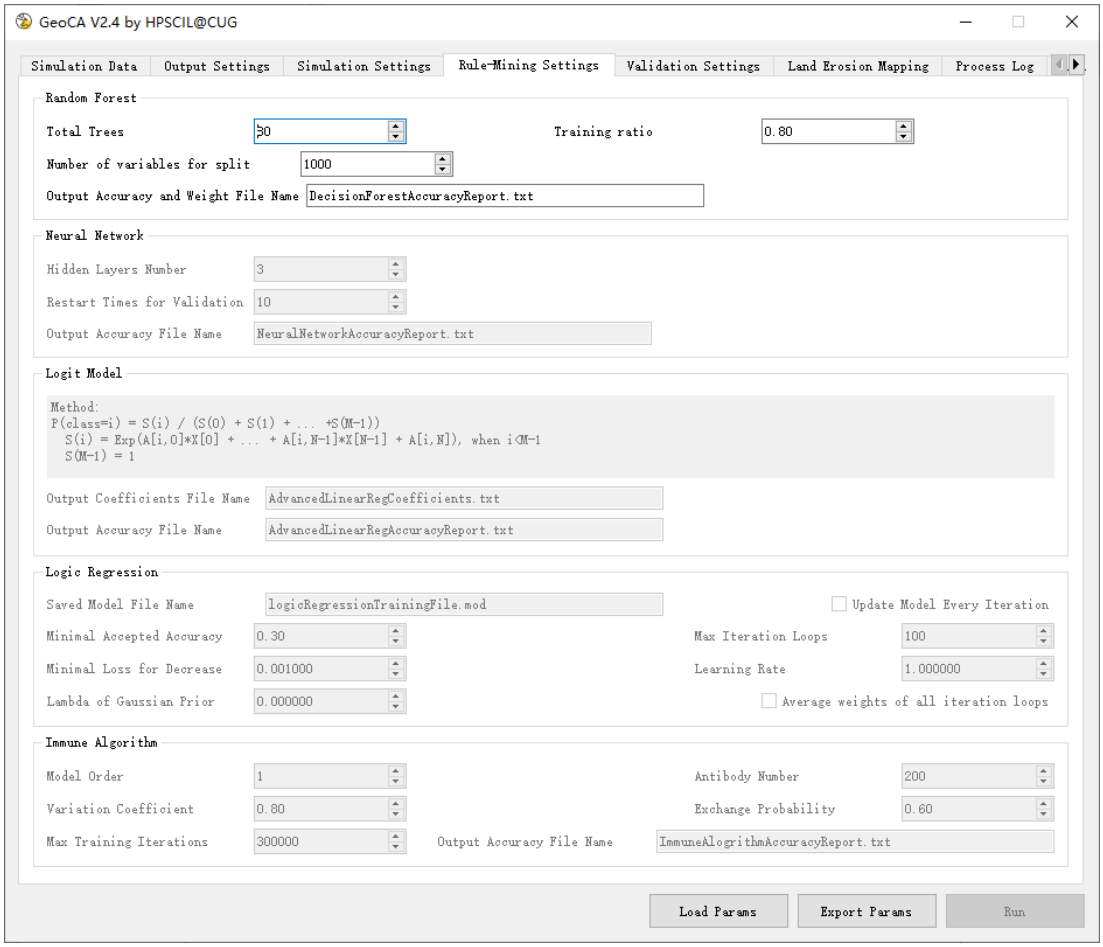

Figure 4. Rule-mining settings interface.

Random Forest

Toal trees: total number of trees used to construct a random forest.
Training ratio: training data set ratio.
Number of variables for split: the number of variables used for tree splitting.
Output Accuracy and Weight File Name: The output random forest precision and variable weight file is stored in Simulation Data > > Output Filepath..

Neural Network

Hidden Layers Num: hidden layers num in neural network.
Restart times for validation: When it is set to 10, it is 10-fold cross validation.

Output Accuracy File Name: The output neural network cross validation precision file is stored in Simulation Data > > Output Filepath.

Logit Model

Output Cofficients File Name: Output the variable weight of each category.
Output Accuracy File Name: The output logic model training precision file is stored in Simulation Data > > Output Filepath.

Logic Regression

Saved model file name: model file saved by logic regression. It is stored in Simulation Data > > Output Filepath;
Minimal accepted accuracy: This is the iteration stop condition. When the model accuracy is greater than the value, the training model is stopped.
Max iteration loops: This is the iteration stop condition. When the number of iterations reaches the value, the training model is stopped;
Minimal loss for decrease: loss function parameters in model iteration.
Lambda of Gaussian Prior: Gaussian kernel parameters of the model.

Learning rate: Setting of learning rate in gradient descent.

Update Model Every Iteration:
√ If this box is checked: Load the existing model to update at startup.
○ If this box is unchecked: The model will be retrained.

Average weights of all iteration loops:
√ If this box is checked: Variable weights will be updated with the mean of all iterations.
○ If this box is unchecked: Variable weights will be the result of the last iteration.

Immune Algorithm

Model order: Algorithm order of immune model.

Antibody number: Number of antibodies.
Variation Coefficient: Difference level.
Exchange Probability: Probability of antibody exchange gene.
Max Training Iterations: Maximum iterations of model training.
Output Accuracy File Name: The output immune algorithm model training precision file is stored in Simulation Data > > Output Filepath.

Validation Settings

Function description

The module is used to verify the CA simulation results, and the accuracy of each iteration result can be calculated.
This function is optional.

Interface description

The interface is shown in Figure 5.
According to the input validation data, the accuracy evaluation of CA simulation iteration results is supported.

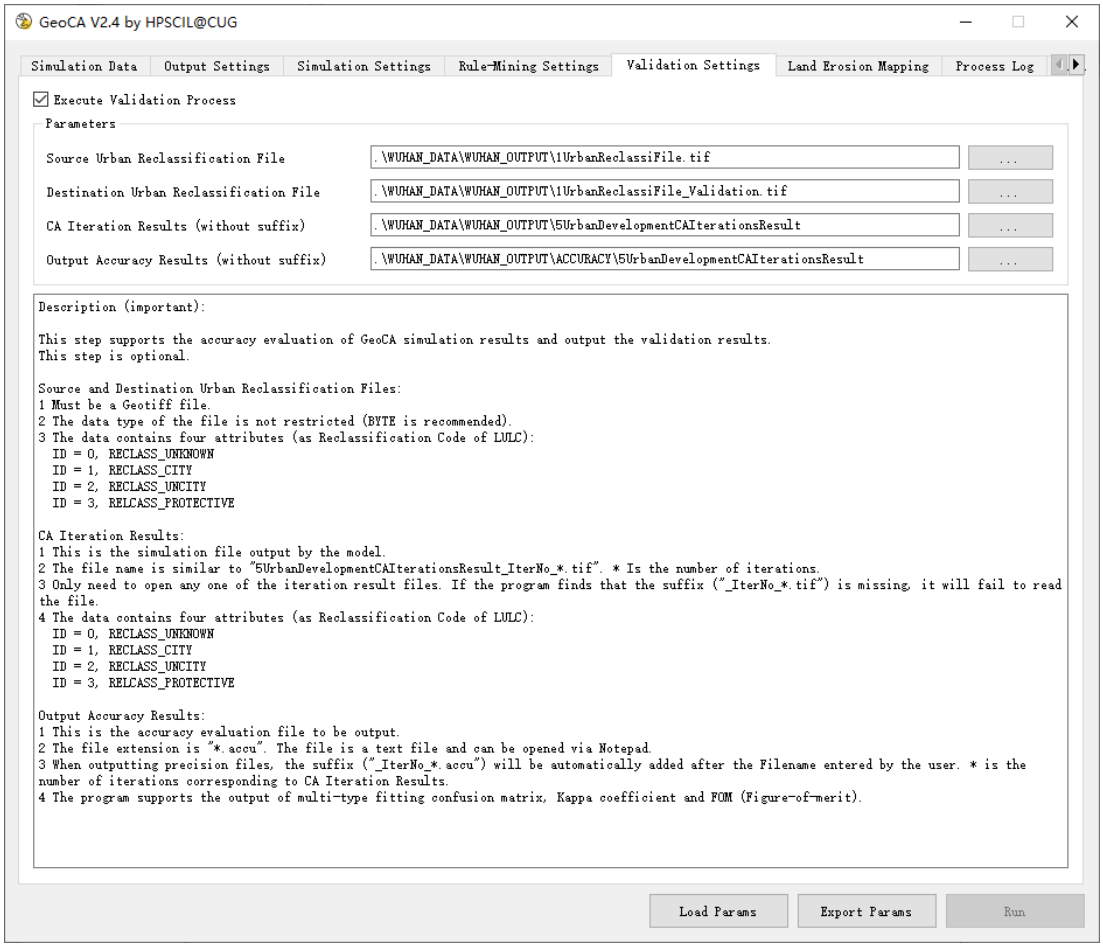

Figure 5. Validation settings interface.

Source Urban Reclassification File

Simulate the land use reclassification file for the start year.
The extension type is tif, the file type is Geotiff, and the data type is BYTE.
Absolute path.

Destination Urban Reclassification File

Land use reclassification file for the validation year (simulate target year).
The extension type is tif, the file type is Geotiff, and the data type is BYTE.
Absolute path.

CA Iteration Results without Suffix

Absolute path.
File extension name is not required. The default setting is Simulation Data > > Output Filepath, and the file name is Output Settings > > Iteration Results
CA iteration results should be named as this type: filename_ IterNo_ *. TIF, where * is the number of iterations
If the manually selected file is missing "_ IterNo_ ", GeoCA will refuse to load.

Output Accuracy Results without Suffix

Absolute path.
File extension name is not required. The default setting is Simulation Data > > Output Filepath, and the file name is Output Settings > > Iteration Results
The iteration results are automatically named according to the number of iterations: 5UrbanDevelopmentCAIterationsResult_ IterNo_ *. accu, where * is the number of iterations. *.accu is a text file that can be opened in Notepad. GeoCA automatically outputs confusion matrix, kappa coefficient, various misclassification and misclassification errors, various user and producer accuracy, Figure-of-Merit (FOM), user precision and producer accuracy of FOM.

Land Erosion Mapping

Function description

This module is a typical CA application, which is used to analyze the proportion of land eroded in the process of urban development, and the distribution map of erosion.
This function is optional.

Interface description

The interface is shown in Figure 6.
According to the input CA simulation iteration results and the third party land use data, the erosion analysis and mapping of the user selected land are carried out.

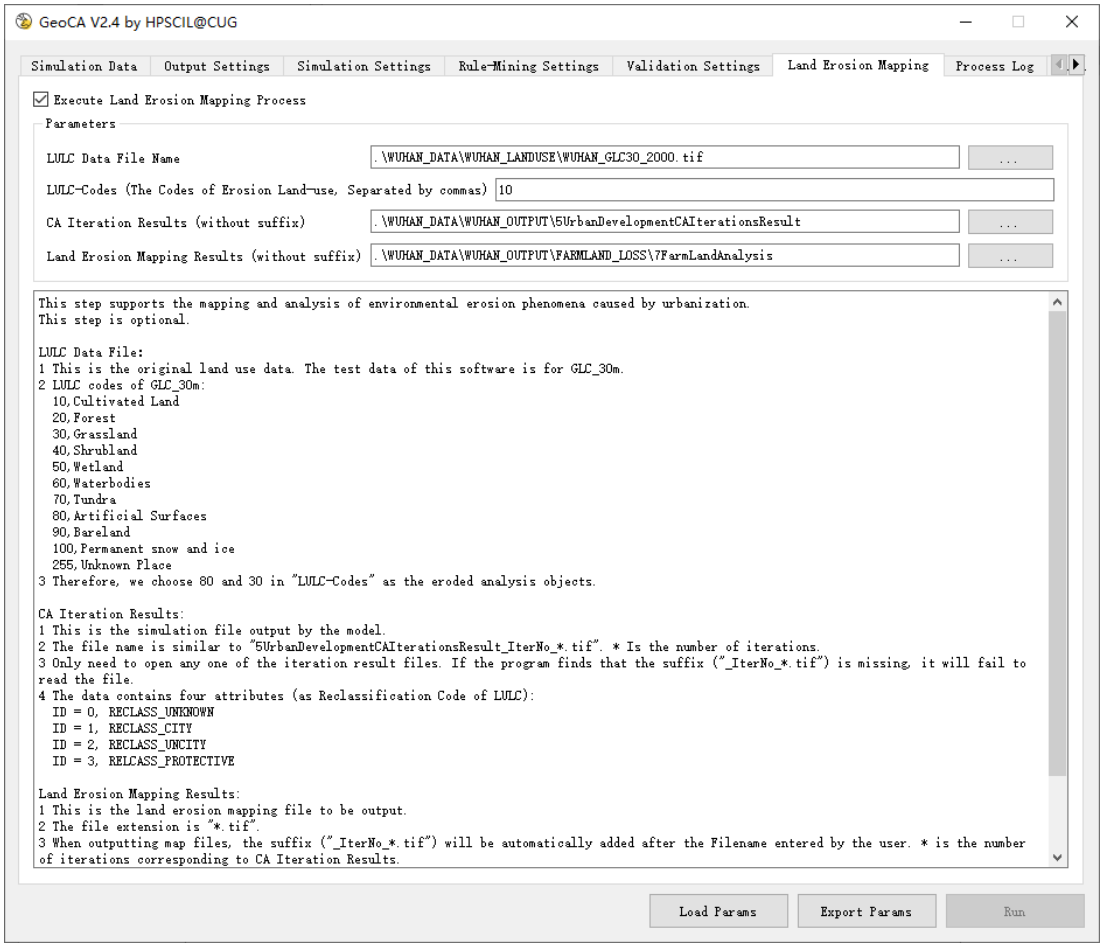

Figure 6. Land erosion mapping interface.

LULC Data File Name

Absolute path.
The file type is Geotiff, and the data type is BYTE.

LULC-Code

In the third-party land use file, the land erosion situation that needs to be analyzed.
Multiple land use codes can be entered for analysis, and each code is separated by ",".

CA Iteration Results without Suffix

Absolute path.
File extension name is not required. The default setting is Simulation Data > > Output Filepath, and the file name is Output Settings > > Iteration Results.
CA iteration result naming should be of this type: filename_ IterNo_ *. TIF, where * is the number of iterations.
If the manually selected file is missing "_ IterNo_ ", GeoCA will refuse to load.

Land Erosion Mapping Results without Suffix

Absolute path.
File extension name is not required. The default setting is Simulation Data > > Output Filepath, and the file name is "7FarmLandAnalysis"
The iteration results are automatically named: 7FarmLandAnalysis_IterNo_*.tif, where * is the number of iterations.
The erosion result file type is Geotiff and the data type is BYTE.
The statistical land erosion proportion can be queried in the log file.

Load and Export Params and Run

Function description

The module is used to load, output and execute parameters.

Interface description

The interface is shown in Figure 7.
Support loading and saving of GeoCA parameter file and running simulation process.

Figure 7. Load and export params and run.

Load Params

Support loading existing parameter files.
The parameter file is in XML format.

Export Params

Support to save the parameters set by the user.
The parameter file is in XML format.

Run

After the user saves the setting parameters, clicking the button will execute the whole simulation process.
When the model is executed, it will automatically jump to the log output module.
Note: this button will not be valid until the parameters are saved.

Process Log

Function description

The module outputs and displays the log when the model is executed.

Interface description

The interface is shown in Figure 8.
The log files are automatically saved in the logs folder (Simulation Parameters > > Output Filepath).
*.log file is a text file, which can be opened in Notepad.

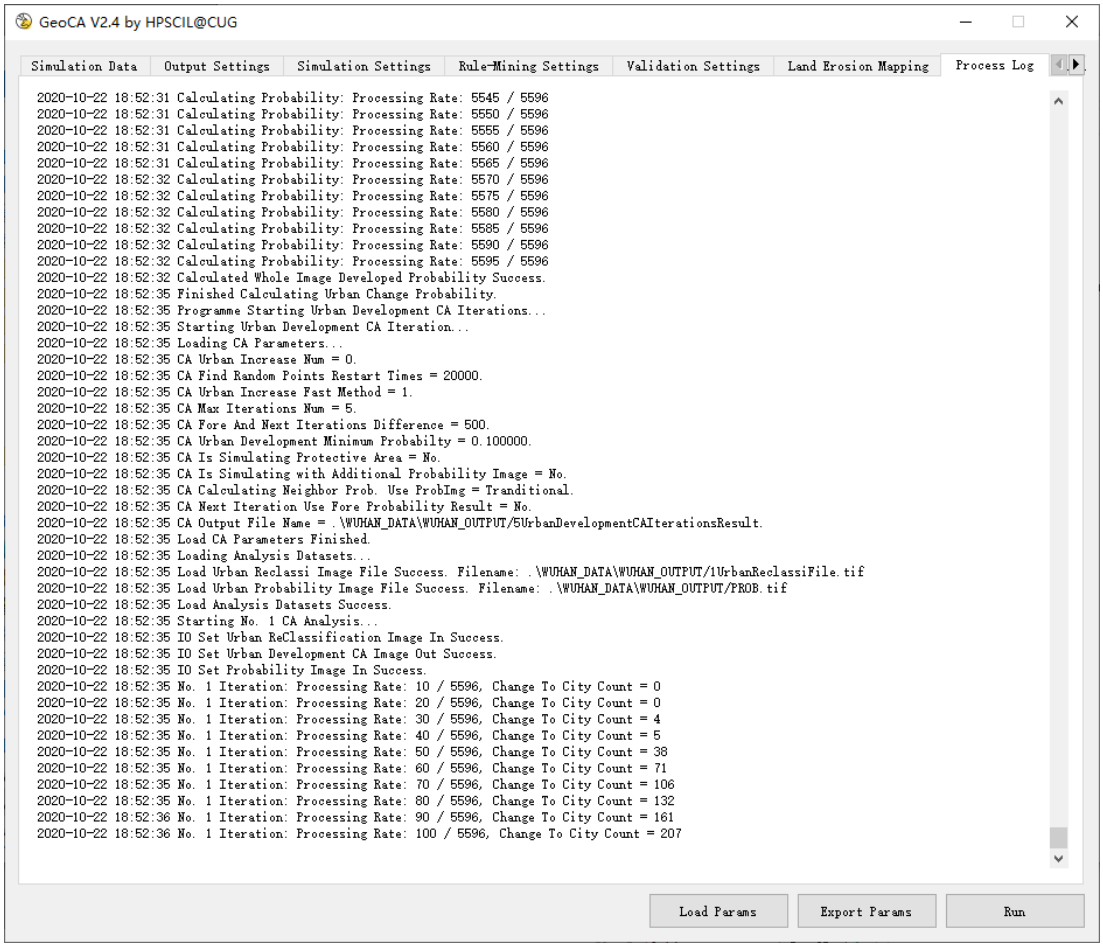

Figure 8. Process log.

Test data, parameters and output results

In order to let users understand the function and file requirements of GeoCA more quickly, GeoCA provides a set of test data and test parameter files.
The test data set is located in the ./WUHAN_DATA folder under the program directory (Figure 9).

The corresponding test parameter file is in the program directory ./sim_params.xml (Figure 10).

You can use the module provided by GeoCA to load parameters and save them (to a third-party file) for direct execution.

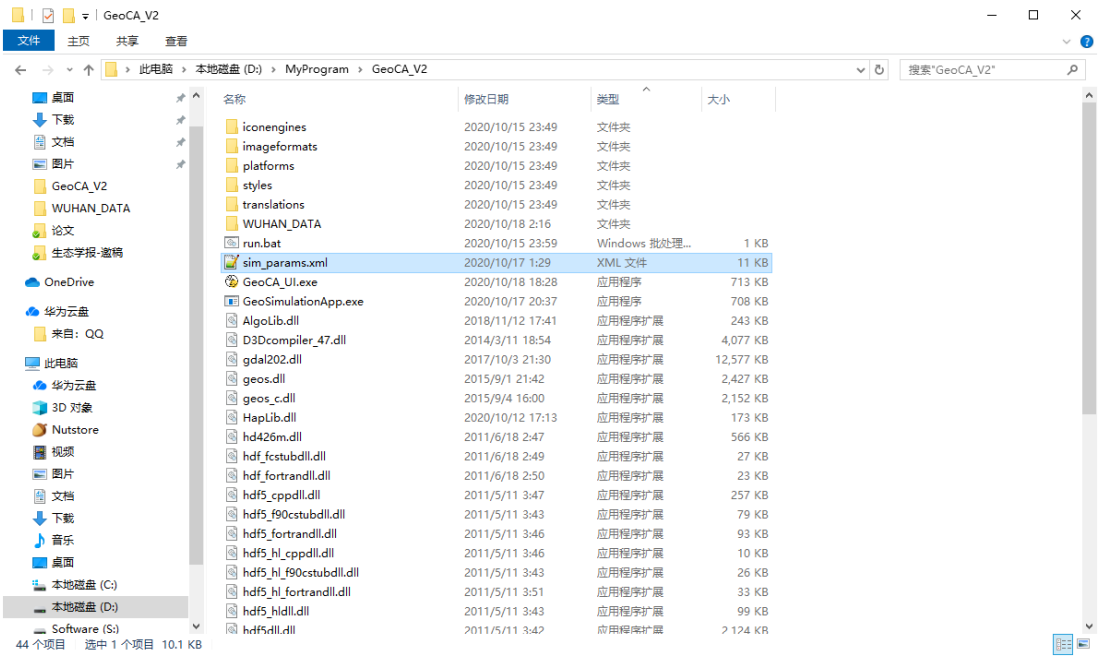

Figure 9. Location of test data and parameters.

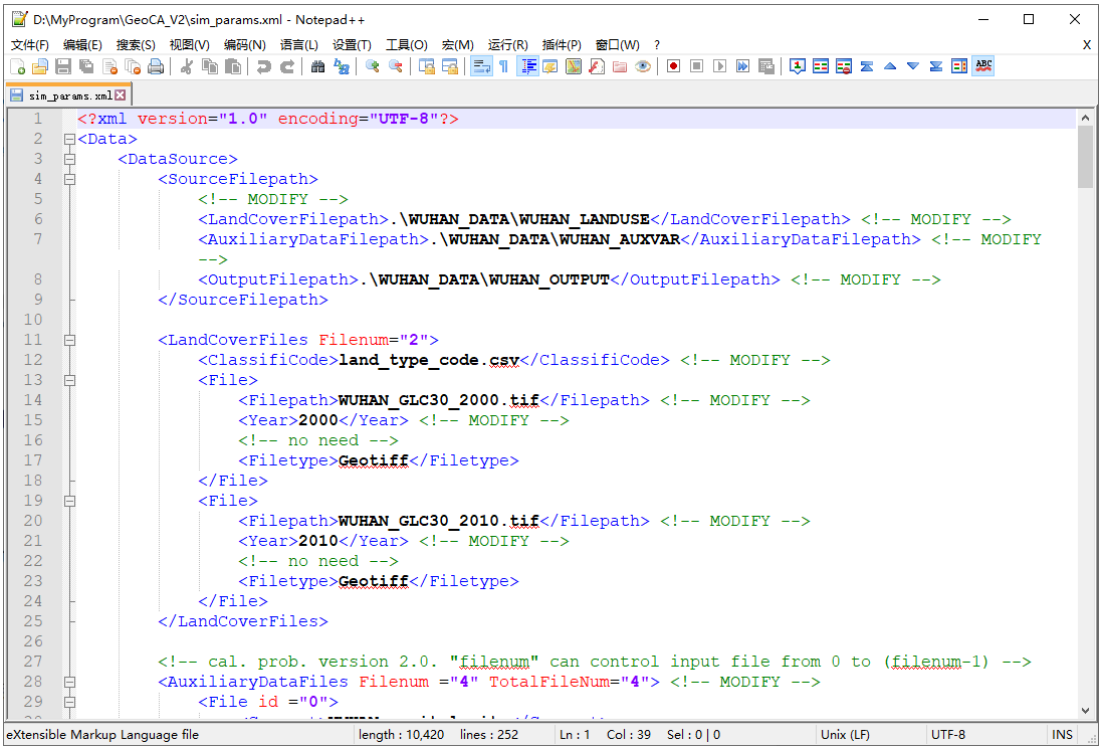

Figure 10. Using Notepad++ to open test parameter file.

Description of files

Input file

Land use data file

File format: Geotiff with projection coordinate information
Data type: BYTE, UINT16, FLOAT32, single band
Description: None

Land use data code file

File format: text file
Data type: txt or csv
Description: Text files separated by "," are recommended to be in English or numbers instead of Chinese.

First line: DN, NAME, TYPE_ID, RE_TYPE, RE_TYPE_ID, COMMENTS
DN: the corresponding value in land use data (DN value)
Name: Land Use / land cover name
TYPE_ ID: 1 ~ n, numbered according to the number of land use types
RE_ TYPE：NON_ URBAN_ AREA, URBAN_AREA, PROTECTED_ AREA
RE_TYPE_ID：NON_URBAN_AREA(0)、URBAN_AREA(1)、PROTECTED_AREA(-1)
Comments: optional

Land use reclassification data file

File format: Geotiff with projection coordinate information
Data type: BYTE, numerical range(0-3), single band
Description: Using land use data code file to reclassify land use data file
Data code:
0 Unknown area
1 Urban area
2 Non urban area
3 Protected area

Spatial auxiliary data file

File format: Geotiff with projection coordinate information
Data type: FLOAT32, single band
Description: Spatial auxiliary data files, such as distance from city center, distance from main roads, road network density, etc.
Quantity limit: no more than twice the number of random points set by the user (the default value is 5000, so there will be a large error in the spatial auxiliary data model when the input is more than 10000).

Additional probability data file

File format: Geotiff with projection coordinate information
Data type: FLOAT32, single band
Description: the additional probability file multiplied by the total probability file PG

Output file

Urban conversion file

File format: Geotiff with projection coordinate information
Data type: BYTE, numerical range(0-9), single band
Description: urban conversion files based on two land use reclassification files
Data code: Urban conversion code
0 unknown conversion
1 Conversion of non urban area to urban area
2 Non urban area has not changed
3 Conversion of non urban area to protected area
4 Urban area has not changed
5 Conversion of urban area to non urban area
6 Conversion of urban area to protected area
7 Conversion of protected area to urban area
8 Conversion of protected area to non urban area
9 Protected area has not changed

Reclassification of land use files

Refer to "input data > > land use reclassification data file".

Development Suitability file

File format: Geotiff with projection coordinate information
Data type: BYTE, numerical range(0-2), single band
Description: urban suitability conversion file based on land use data code file
Data code:
0 Conversion to urban area is not allowed
1 Conversion to urban area is allowed
2 Protected area (user decides whether it can be changed)

Random point file for regression analysis (data mining)

File format: text file
Data type: txt
Description: The first line is the number of random points, variables and conversion types
Longitude
Latitude
Feature-(nFID+1): nFID-th feature value
ClassiID: City conversion code (see "city conversion file" for more details)

Urban land use conversion probability file(non urban to urban)

File format: Geotiff with projection coordinate information
Data type: FLOAT32, numerical range(0-1), single band
Description: The probability that non-urban area and protected area converts into urban area.

9-band conversion probability file

File format: Geotiff with projection coordinate information
Data type: FLOAT32, numerical range(0-1), 9 bands
Description: The probability that non-urban area, protected area and urban area converts to each other.
Band description:
1 Conversion of non urban area to urban area
2 Non urban area has not changed
3 Conversion of non urban area to protected area
4 Urban area has not changed
5 Conversion of urban area to non urban area
6 Conversion of urban area to protected area
7 Conversion of protected area to urban area
8 Conversion of protected area to non urban area
9 Protected area has not changed

Simulation result file

File format: Geotiff with projection coordinate information
Data type: BYTE, numerical range(0-3), single band
Description: "_IterNo_*.tif" is in the file name, where * is the number of iterations.
Data code:
0 Unknown area
1 Urban area
2 Non urban area
3 Protected area

Rule mining precision report file

File format: text file
Data type: txt
Description: The first line is the report generation time.
Index description:
rel cls error: relative classification accuracy
avgce: average cross entropy
rms error: root mean square error(RMSE)
avg error: average error
avg rel error: average relative error
oob-*: out of bag error (Random forest)
Top vars: variable importance ranking (Random forest)
Var importances: variable importance value (Random forest)
Coefficients: variable weight parameter (logistic regression)
Number of independent variables: (logistic regression)
Number of classes: (logistic regression)

Model validation file

Source / Target city reclassification file

Refer to "Input data > > Land use reclassification data file".

Simulation result file

Refer to "Output data > > Simulation result file"

Output precision file

File format: text file (The extension type is *.accu)
Data type: accu
Description: There is "_IterNo_*.accu" in the file name, where * is the number of iterations (Figure 11). The first line is the file name of the simulation results of the evaluation, and the second line is the report generation time.
Index description:
Confusion Matrix
Overall Accuracy
Kappa Coefficient
Commission Error
Omission Error
Mapping Accuracy
User Accuracy
FOM: Figure-of-Merit
FOM_PA: Producer precision based on FOM
FOM_UA: User precision based on FOM

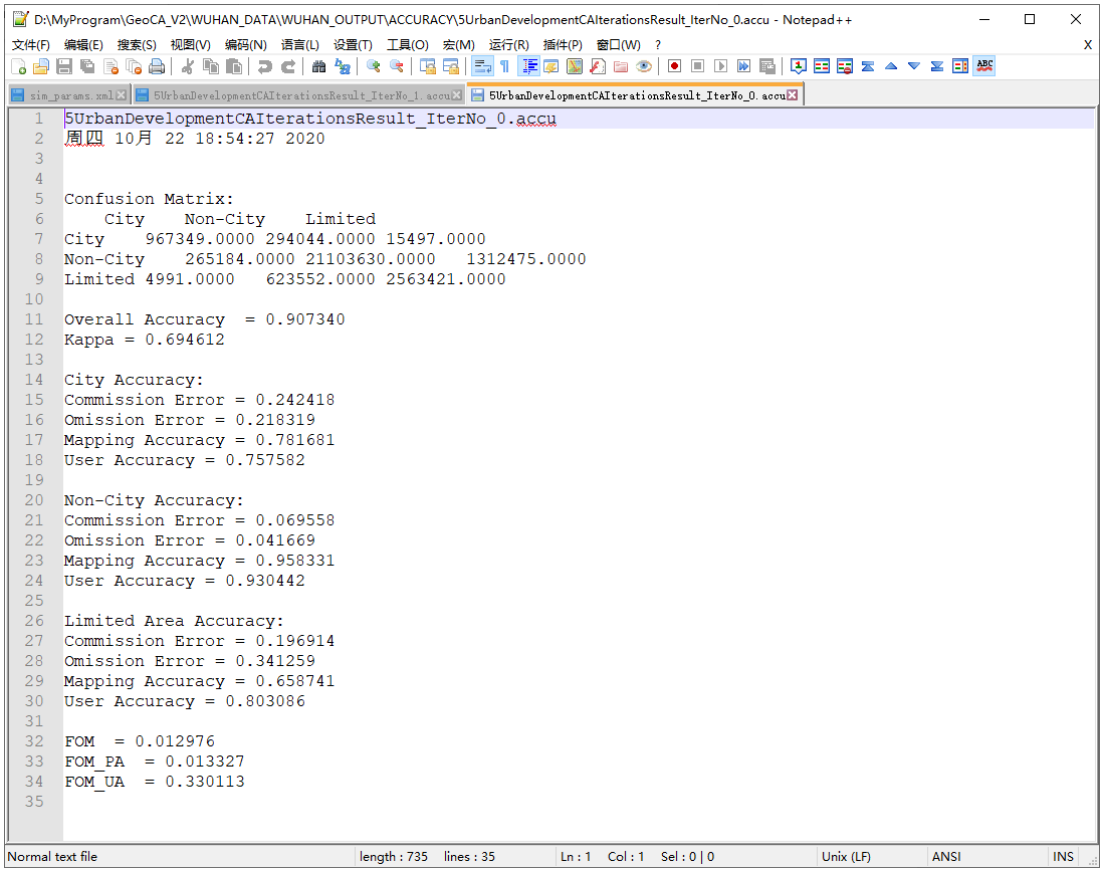

Figure 11. Output precision file.

Land erosion mapping file

Land use LULC file

Refer to "Input file > > Land use data file".

Simulation result file

Refer to "Output file > > Simulation result file".

Output mapping file

File format: Geotiff with projection coordinate information
Data type: BYTE, numerical range(0-2), single band
Description: Spatial mapping of land erosion in the future. There is "_IterNo_*.tif" in the file name, where * is the number of iterations (Figure 11). The erosion ratio can be viewed in the log file (Figure 12).
Data code:
0 unknown
1 eroded by urban area
2 not eroded by urban area

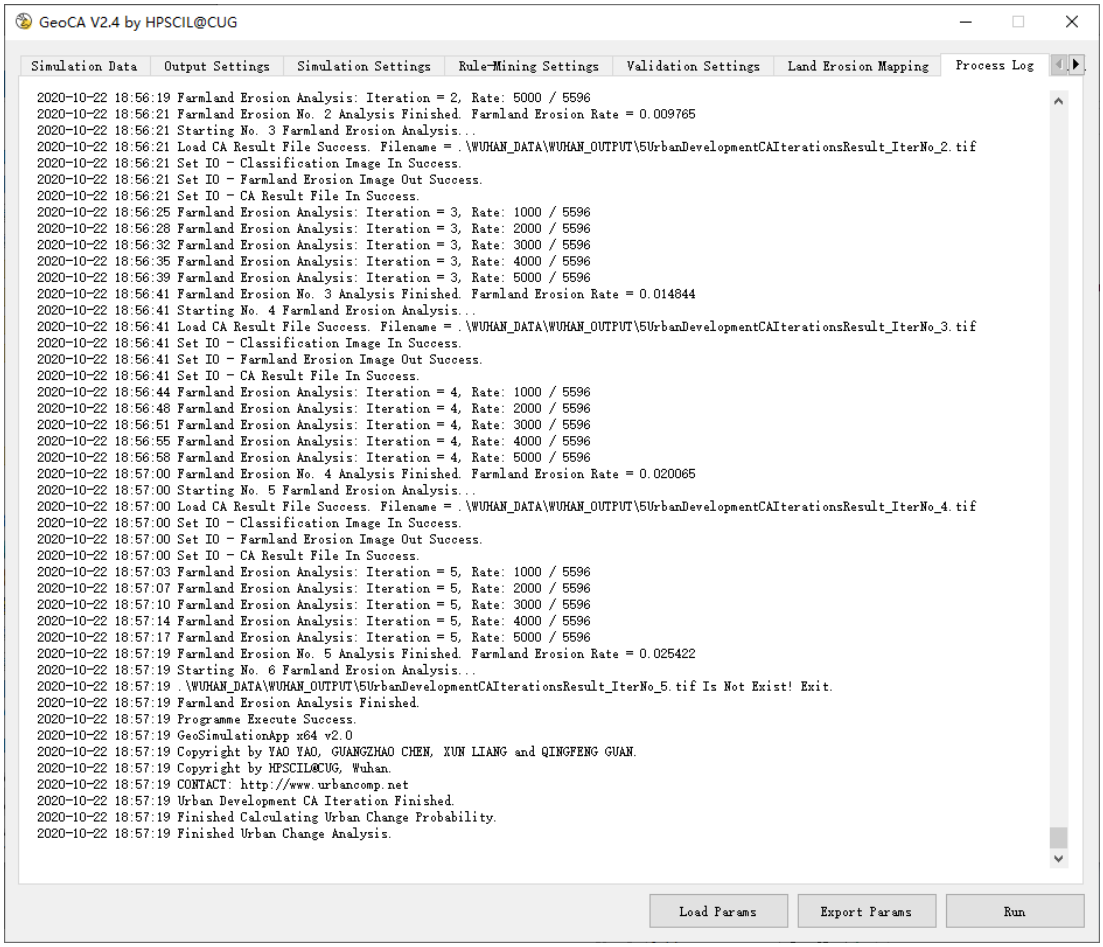

Figure 12. Output of land erosion analysis in log file.

Log file

The log file is the output file when the program is running and can be viewed in the processing log output module of GeoCA.

The log file is saved in the logs folder of Output Filepath by default (Figure 13).
The naming format of the log file is "HPSCIL_SIM_GEOCA_V2_HOUR_MINUTE_SECOND_DATE.log", and the time in the file name is the time of program execution.
The log file can be opened via Notepad (Figure 14).

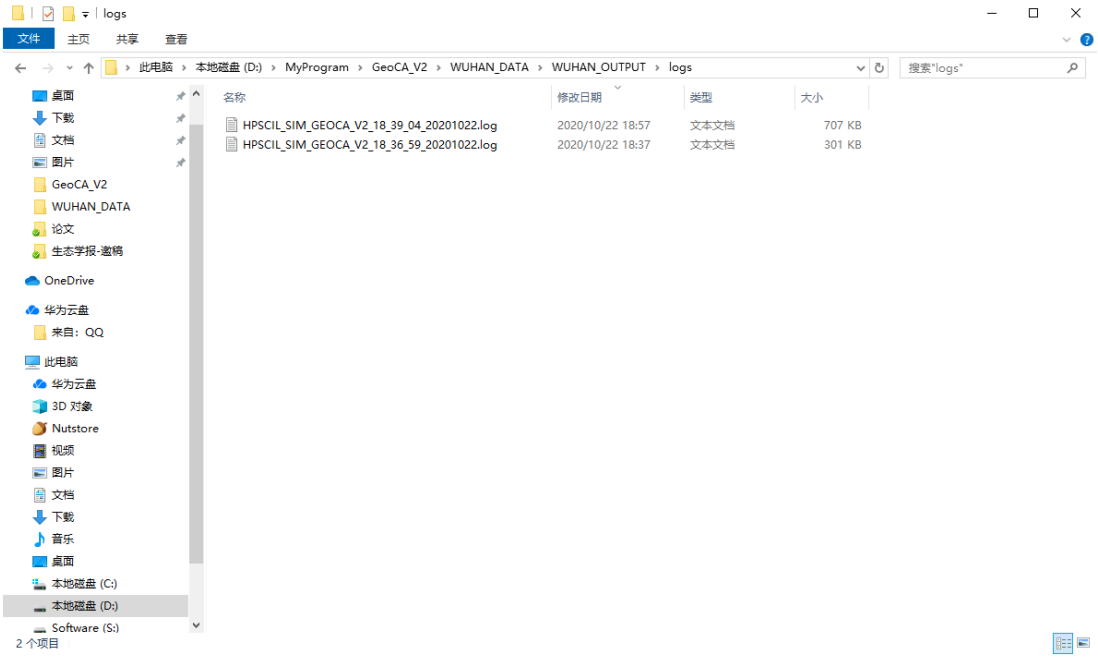

Figure 13. Log folder.

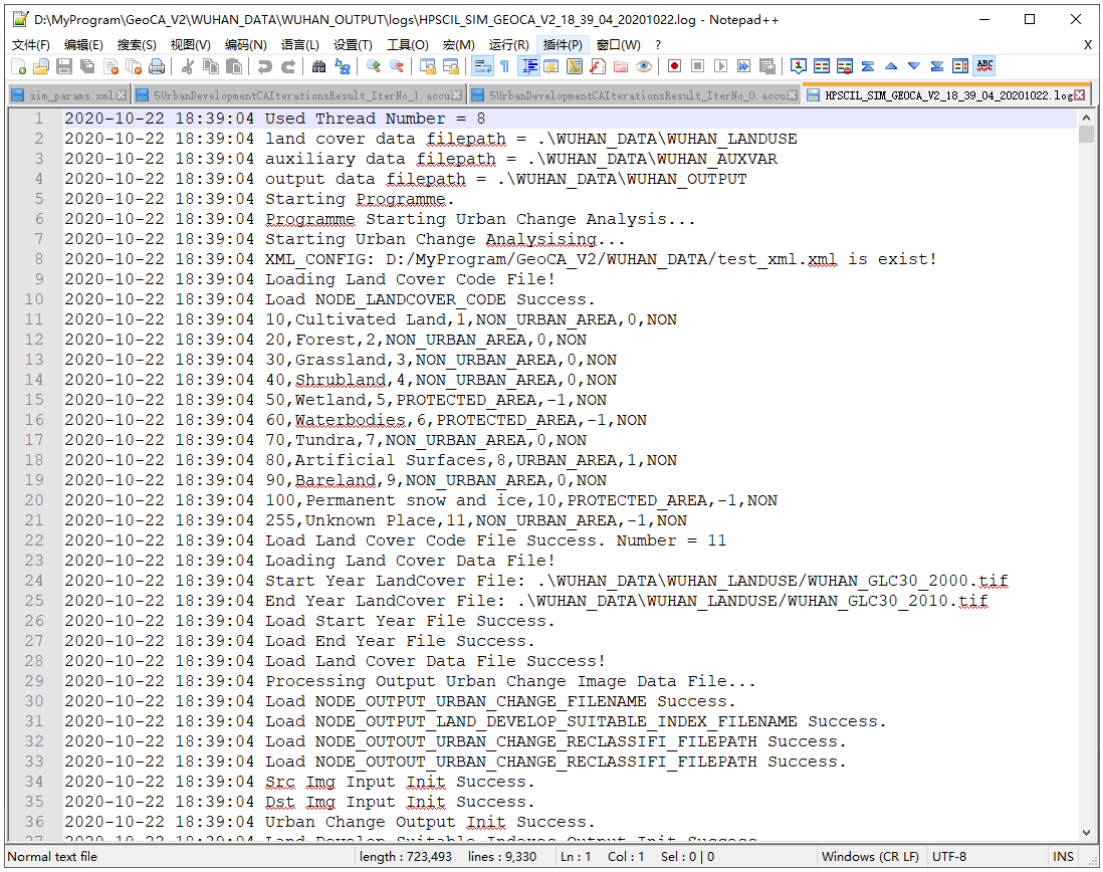

Figure 14. Log file content.

Parameter file

The parameter file is in XML format.
For XML file comments, you can check the comments in ".\sim_params.xml" in the program directory.
The user can modify the XML parameter file directly and modify the program directory ".\run.bat" in the program directory to execute in batch mode.

Contact

If you encounter problems in use, please contact us in time.
Please attach the parameter file (*.xml) exported by the model and the log file (*.log) of the model execution error in the email.

References

[1] Chen, D., Zhang, Y., Yao, Y., Hong, Y., Guan, Q., & Tu, W. (2019). Exploring the spatial differentiation of urbanization on two sides of the Hu Huanyong Line--based on nighttime light data and cellular automata. Applied Geography, 112, 102081.

[2] Zhang, D., Liu, X., Wu, X., Yao, Y., Wu, X., & Chen, Y. (2019). Multiple intra-urban land use simulations and driving factors analysis: a case study in Huicheng, China. GIScience & Remote Sensing, 56(2), 282-308.

[3] Yao, Y., Liu, X., Zhang, D., Liang, Z., & Zhang, Y. (2017). Simulation of Urban Expansion and Farmland Loss in China by Integrating Cellular Automata and Random Forest. arXiv preprint arXiv:1705.05651.

[4] He, Y., Ai, B., Yao, Y., & Zhong, F. (2015). Deriving urban dynamic evolution rules from self-adaptive cellular automata with multi-temporal remote sensing images. International Journal of Applied Earth Observation and Geoinformation, 38, 164-174.

[5] Li, X., & Yeh, A. G. O. (2002). Neural-network-based cellular automata for simulating multiple land use changes using GIS. International Journal of Geographical Information Science, 16(4), 323-343.