Patch-level Land Use Simulation model V1.0 (PLUS) User's Manual

Authors:

Xun Liang, Qingfeng Guan*, Keith C. Clarke, Shishi Liu, Bingyu Wang, Yao Yao

* Corresponding author. E-mail address: guanqf@cug.edu.cn (Qingfeng Guan).

Consultation

If you have technical questions regarding PLUS software, please contact Dr. Xun Liang (liangxun@cug.edu.cn)

Contact info:

High-performance Spatial Computational Intelligence Lab (HPSCIL) (https://github.com/HPSCIL)

School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei 430078, China

For any possible research collaboration, please contact Prof. Qingfeng Guan (guanqf@cug.edu.cn)

Abstract

The cellular automata (CA) model is developed to improve the representation of complex land-use and land-cover (LULC) systems. Previous studies focus on the improvement of technical modeling procedures, and little researches has pay attention to promote the understanding of the nonlinear relationships underlying LULC. Lacking of model ability to reflect patch-level landscape evolution also limits the applicablepotential of CAs for policy development. This study presents a patch-level land use simulation (PLUS) model that integrates a rule -mining framework based on Land Expansion Analysis Strategy (LEAS), and a CA model based on multi-type Random Seeds (CARS), which can illustratethe drivers of land expansion and project landscape dynamics. The PLUS model can obtain higher simulation accuracy and a more similar landscape pattern than other models. The LEAS can help researchers investigatethe underlying transition rules. The proposed method combines simulation, knowledge discovery and policy-making processes, which can provide vital information formultiply users (e.g., researchers, planners and policy makers).

PLUS was developed purely in the C++ language. The parallel technology of PLUS software is from High-performance Spatial Computational Intelligence Lab @ China University of Geosciences (Wuhan) (https://github.com/HPSCIL). The Random forest technique in our model is from a powerful open -source library called Alglib 3.9.2 (http://www.alglib.net/). The UI of the software is built using a famous open -source library Qt 5.13 (https://www.qt.io/download/). This UI provides a real-time display of dynamic changes of land use in the simulation process. Moreover, the using of open source library GDAL 2.0.2 (http://www.gdal.org/) allows our model to directly read and write raster data (.tif, .img, .txt files) that includes geographical coordinate information.

This user's manual intends to provide users a "quick start" on how to use the PLUS software. All of the necessary data and files for the tutorial have been provided and can be used as templates forpractices.

1. Download and start PLUS

The latest version of PLUS software and the user's manual are provided at a download link (https://github.com/HPSCIL/Patch-level_Land_Use_Simulation_Model). Click the executable files "**PLUS VX.0.exe**" to start the software.

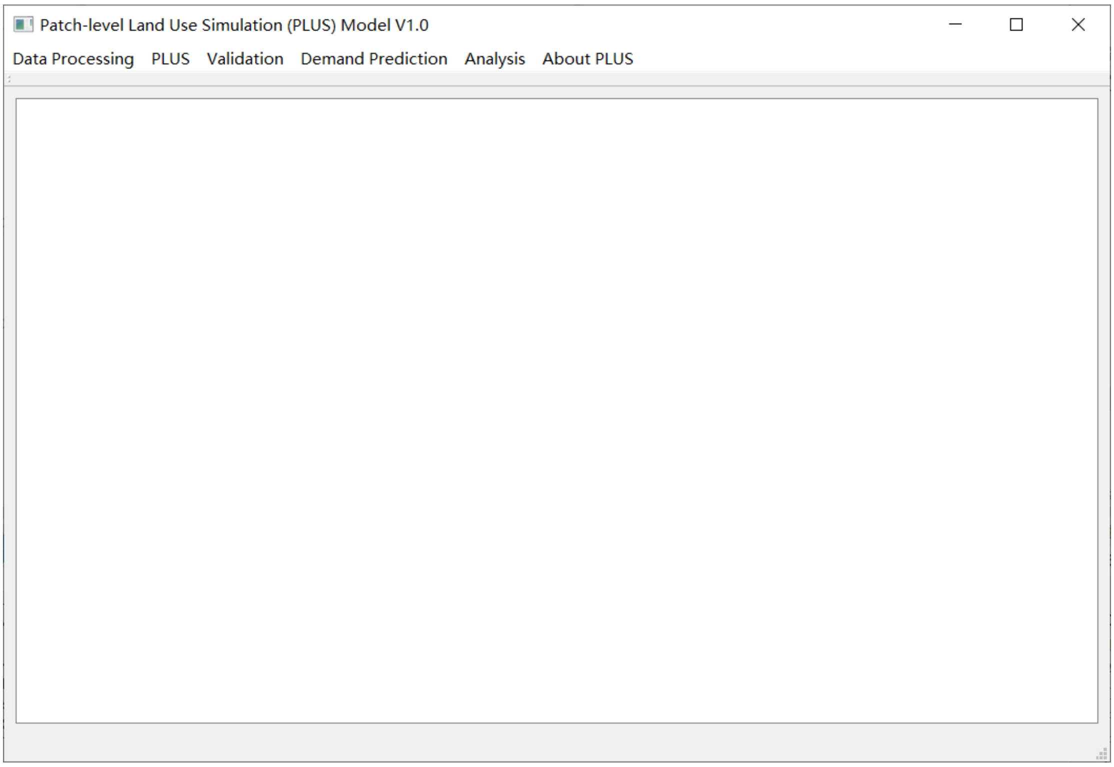


Fig. 1. The main interface of the PLUS model.

The main interface is a brief image browser that provides dynamic display function for the simulation process.

2. Some critical notes before applying the PLUS model

1) The input land use data must be in an ‘unsigned char’ format.

2) The LEAS do not require the input image data have the same number of rows and columns. An integral alignment mechanism is available. However, the coordinate or projection system of these images must be supported by GDAL. We recommend the World Geodetic System (WGS 84).

3) If a module of the PLUS software needs multiple land use maps as its inputs, these land use maps should have the same number of rows and columns.

4) In the step of simulating dynamics of land use change (CARS), the input image data need to have the same number of rows and columns, including the land use pattern, restricted area, and change probability data.

6) Cancel the dynamic display by unchecking the ‘Dynamic Display’ checkbox can make the model faster and more stable.

5) We strongly recommend the users do not include a non-English character or sign in the file paths of the PLUS model, **including the spaces**. The file names and folder names must begin with a letter, not a number.

3. Example data description

Tab. 1 List of test data

Category	Data	Filename	Description
Land use data	Land use data	wh2003_refy.tif wh2013_refy.tif	1: Grass, 2: Deciduous forest, 3: Cropland, 4: Urban land, 5: Bare land, 6: Water area, 7: Evergreen forest
Constraint data	Land use constraint	wh_open_water13.tif	
Socioeconomic data	Population	wh_Pop.tif	http://www.geodoi.ac.cn/WebCn/Default.aspx
	GDP	gdp2010.tif	
	Proximity to the highway	wh_dist_highway.tif	OpenStreetMap (https://www.openstreetmap.org/)
	Proximity to the arterial road	wh_dist_trunk.tif	
	Proximity to the primary road	wh_dist_primary.tif	
	Proximity to the secondary road	wh_dist_secondary.tif	
	Proximity to the tertiary road	wh_dist_teriary.tif	
	Proximity to railway	wh_dist_railway	
	Proximity to high-speed railway stations	wh_dist_highspdstation.tif	http://lbsyun.baidu.com/
	Proximity to governments	wh_dist_gov.tif	
Climatic and environmental data	Soil type	wh_soiltype.tif	HWSD v 1.2 (http://westdc.westgis.ac.cn/data/844010ba-d359-4020-bf76-2b58806f9205)
	Annual Mean Temperature	wh_df_tem.tif	WorldClim v2.0 (http://www.worldclim.org/)
	Annual Precipitation	wh_df_pre.tif	
	DEM	wh _df_dem.tif	NASA SRTM1 v3.0

Slope	wh_df_slope.tif
Proximity to water	wh_dist_openwater

The test dataset can be found in the ***TestData*** folder. Use this dataset to run an example model or compared it with your simulation results. In practical, more human and natural environment effects can be taken into consideration. The study region in this example is Wuhan city, located in Hubei province, central China.

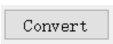
4. User interfaces of PLUS model

4.1. Data processing

First of all, data processing is critical! The PLUS model only supports the ‘unsigned char’ format land use/land cover (LULC) data, with encoded land use category as consecutive integers start from 1 (e,g., 1, 2, 3, 4, 5, 6,……). Also, the PLUS model only supports the ‘unsigned char’ land use constraint data, which is a binary value image that ranges from 0 to 1.

If the users use ArcGIS software to reclassify the LULC data, the ArcGIS usually convert the original ‘unsigned char’ LULC data to ‘int’ or ‘unsigned int’ format, which is not allowed in the PLUS model. Thus we provide a conversion tool. It can convert LULC data and constraint data (in all formats) to ‘unsigned char’ format. If you are not sure theformat in your LULC or constraint data, just apply the conversion tool to generate the ‘unsigned char’ data that are suitable for the PLUS model. This tool is very fast and convenient to use.

Unfold the “**Data Processing**” menu and click the “**Convert LULCs to Unsigned Char Format**” to start the module that has the same name as the item. Click the 

button to input theland use or constraint data: ‘wh2003_refy.tif’ and ‘wh2013_refy.tif’ in the pop-up dialog. Then the users can specify the output folder and click the 

button to convert the data format. The new LULC files ‘wh2003_refy_uc.tif’ and ‘wh2013_refy_uc.tif’ will be exported by the conversion tools. Note that the format of all the test data is suitablefor the PLUS software, this tutorial just provide a guide for how to use the PLUS software. What is more, all the LULC images need to have the same number of rows and columns.

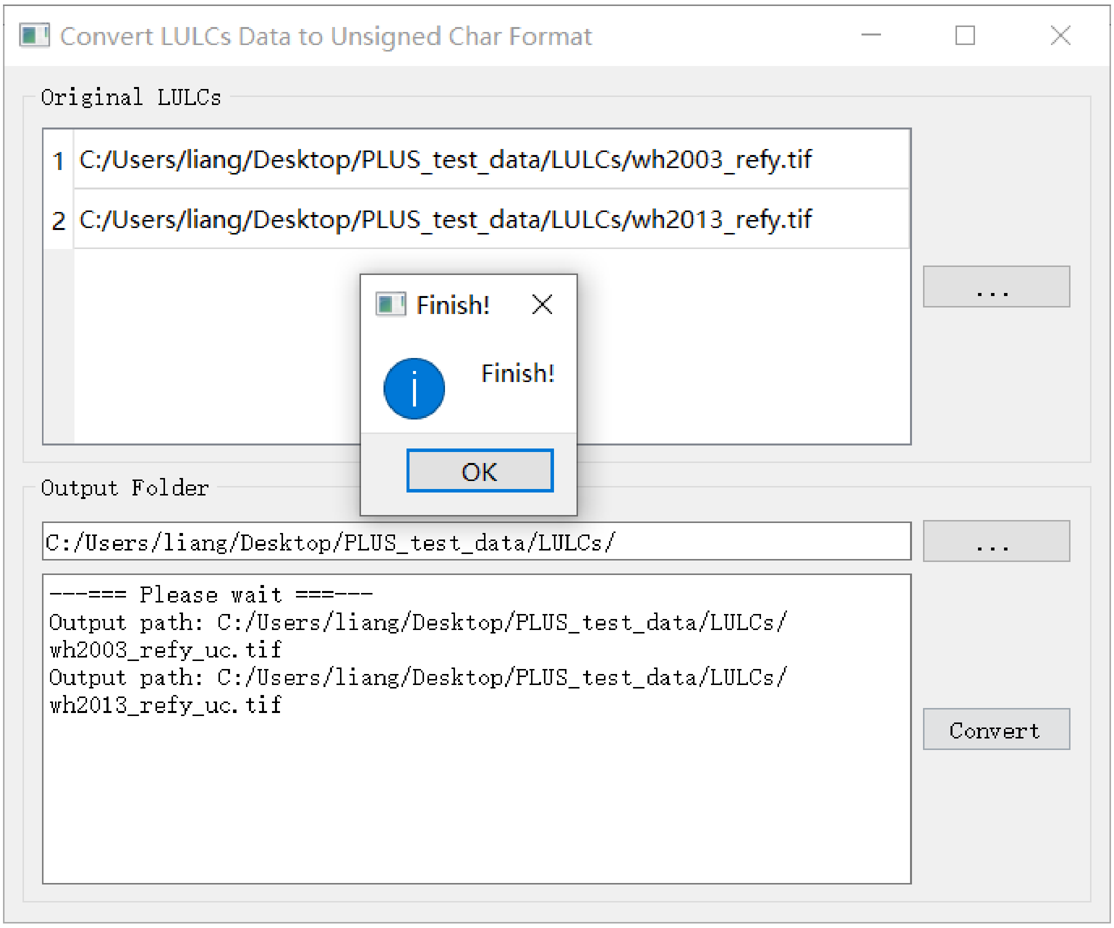


Fig. 2. The UI of the data conversion tool.

4.2. Extract Land expansion

The PLUS consists of three main parts, 1) Extract Land expansion, 2) Land expansion analysis strategy (LEAS), and 3) CA based on multiple random seeds (CARS). Users are allowed to start the corresponding module by selecting the option from the "**PLUS**" button's pull-down menu on the main menu.

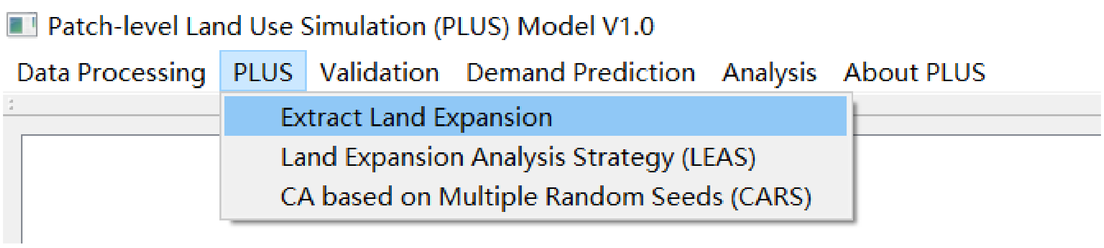
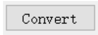


Fig. 3. Starting the 'Extract Land Expansion' module.

Click the  button and select theland use or constraint data: ‘wh2003_refy.tif’ and ‘wh2013_refy.tif’ in the pop-up dialog. Then the users can specify the output file path and click the  button to extract the land expansion map from 2013 to 2013.

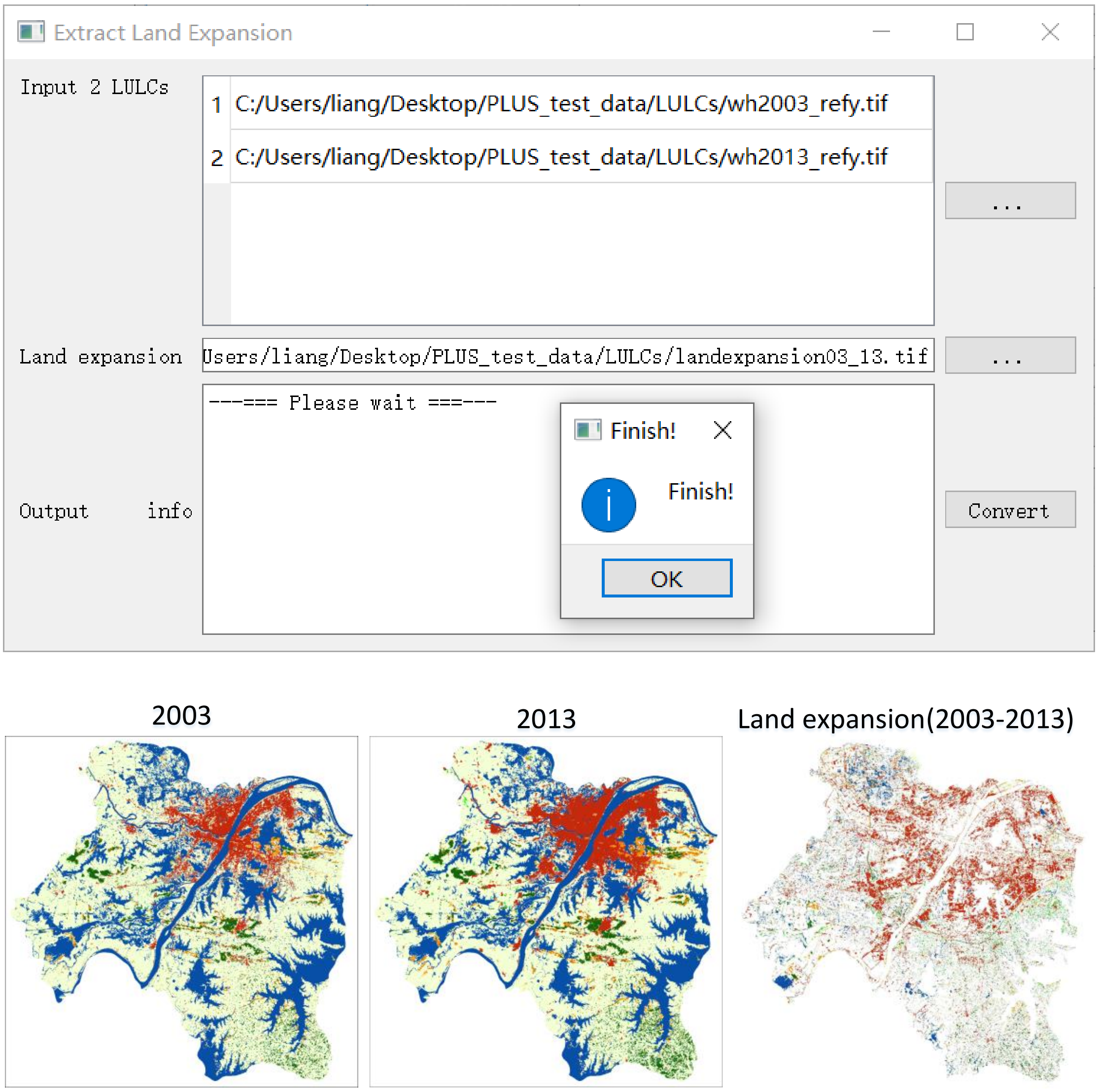


Fig. 4. Extracting land expansion from two periods of land use data.

4.3. Land expansion analysis strategy (LEAS)

4.3.1. Start the module

Click the “**Land Expansion Analysis Strategy (LEAS)**” item to start the module.

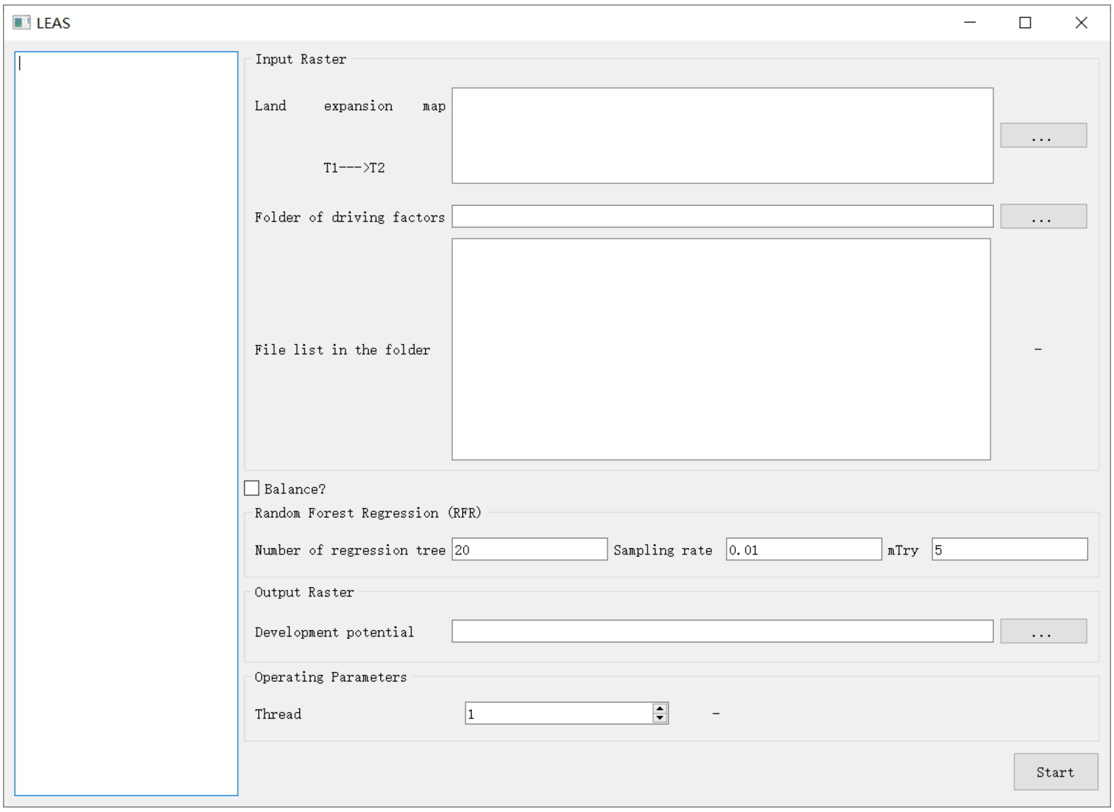


Fig. 5. UI of the 'Land expansion analysis strategy' module.

4.3.2. Input land expansion data

Click the  button of the "Land expansion map" text box and select the land expansion map "landexpansion03_13_landuse_1to2.tif" outputted in the previous step in the pop-up dialog.

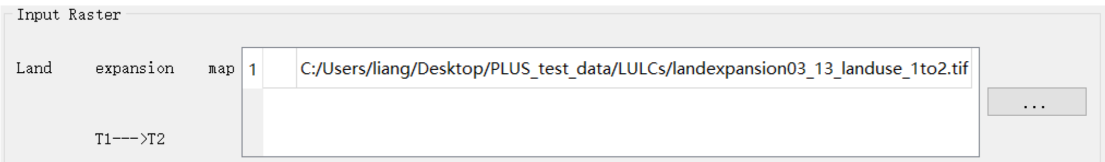


Fig. 6. Select the land expansion map outputted in the last step.

4.3.3. Input driving factors

Click the  button in the "**Folder of driving factors**" group box and select the folder of driving factors in the pop-up dialog. The PLUS software would automatically load the “tiff” File in the folder. We provide 16 driving factorsimplementing the simulation.

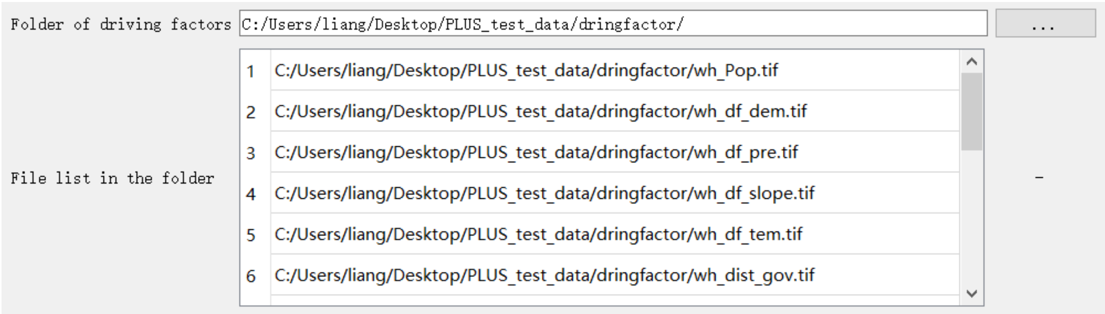
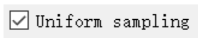


Fig. 7. Chose a folder and the software will automatically input all the driving factors within the folder..

4.3.4. Setting the training parameters and output path

Users can set the parameters of **Random Forest Regression (RFR)**. The sampling rate is set to 0.01 by defaultwhich means about one percent of pixels is selected for training. The number of regression trees is set to 20 in this experiment. mTry means the feature number used to train the RFR model. The max feature number is 16 in this experiment, equal to the number of driving factors. If the checkbox 

is checked, the sampling points for each land use type will be the same. If not, the sampling points will randomly scatter across the study area. Then, click the 

button in the “**Output Raster**” group box and select the change probability in the pop-up dialog. In the test data, the generated development potential of each land use type is named “devPotential.tif”. Users can increase the number of parallel internal threads to accelerate the running speed.

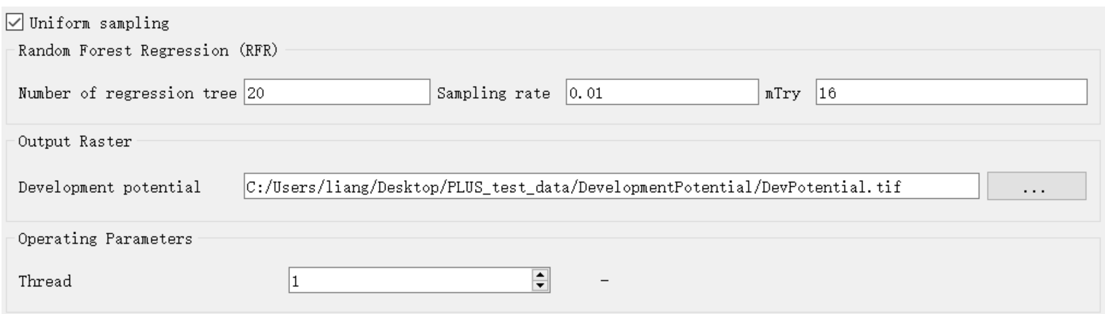


Fig. 8. Determining the training parameters and output path of the RFC.

4.3.5. Running program and examine the result

Once the setting-ups complet, click the button  for running the module. A message box will come out as below when the model process run isfinished.

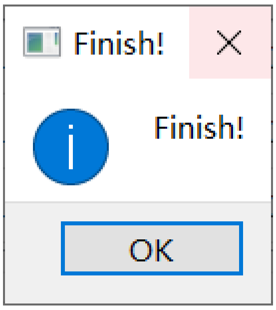


Fig. 9. A message that shows the program run successfully.

The text panel at the left of the interface will output the training accuracies of the RFCs for each land use component. The RFC outputs two accuracy indicators: RMSE and OOB-RMSE for each land use type.

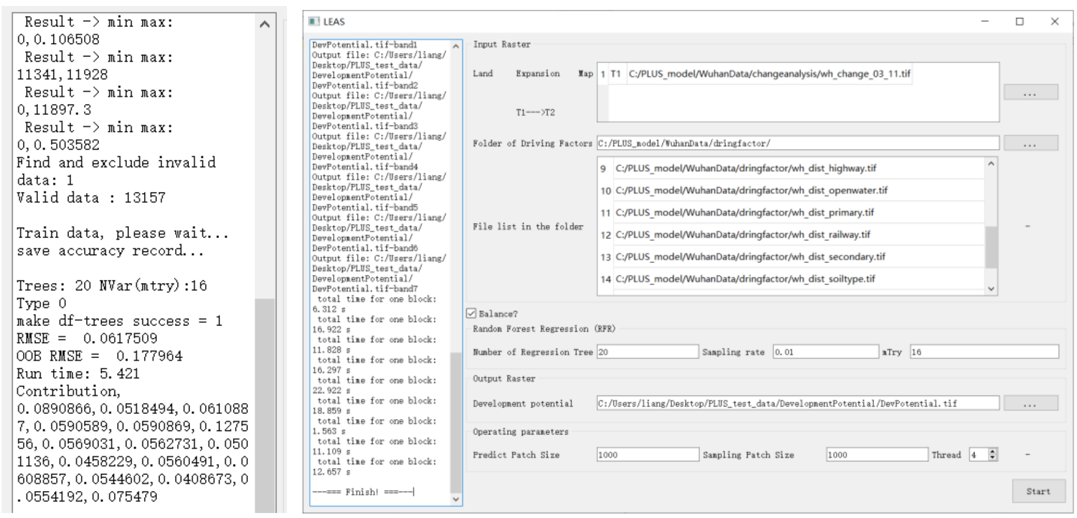


Fig. 10. Accuracy for each land use type.

What is more, the contribution of all the driving factors for the expansion of each land use type can be exported by the RFC, which is very important for researchers to understand the underlying causes of land use change. The contribution of the driving factors for all land use types are saved in CSV files in the ‘Parameterfile’ folder .

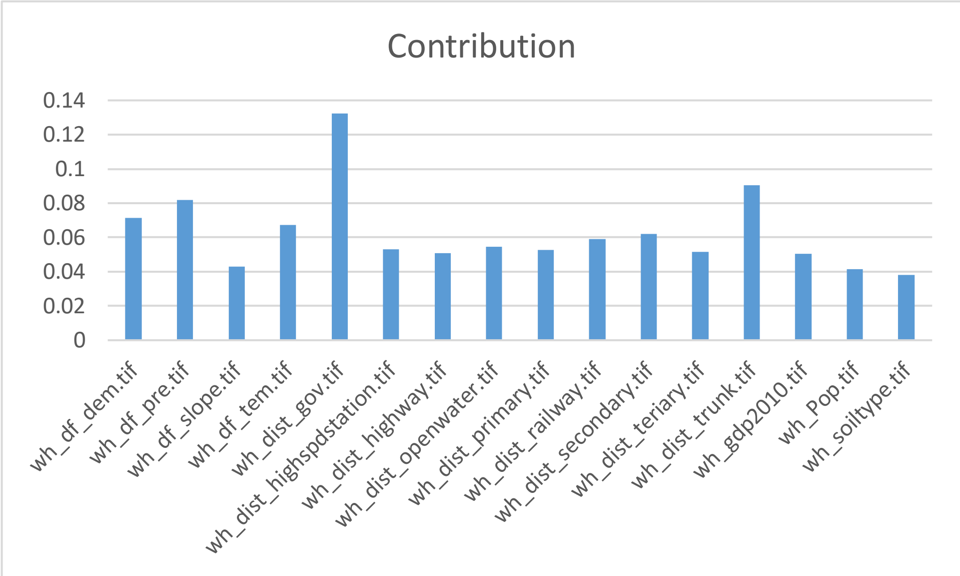


Fig. 11. Contribution of each driving factor tothe expansion of grassland.

For example, the file 'contribution0.csv' recorded the contribution of various driving factors on the expansion of grassland. We find that the proximity to governments has the most influence on the growth of grass. This result indicates that the grass is most likely to grow in the place with less human activities because the surrounding areas ofgovernments are generally the most active places of the human being.

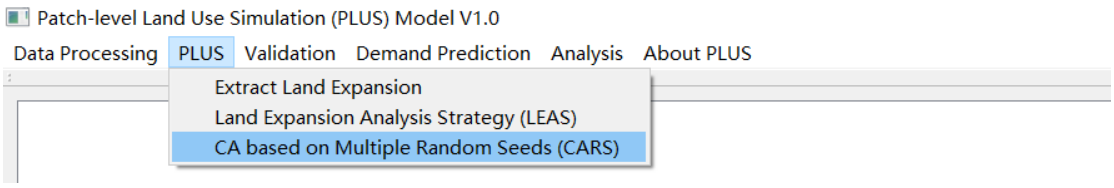
After the training process, we obtained 7 RF classifiers corresponding to 7 land use types. These classifiers are applied to predict the development potential of each land use type on each cell. Finally, the PLUS model output 7 development potential maps.

Fig. 12. Development potential map for each land use type.

4.4. Simulating patch-level land use change

4.4.1. Start the module

Click the “**CA -based on multiple random seeds (CARS)**” item to start the simulation module.



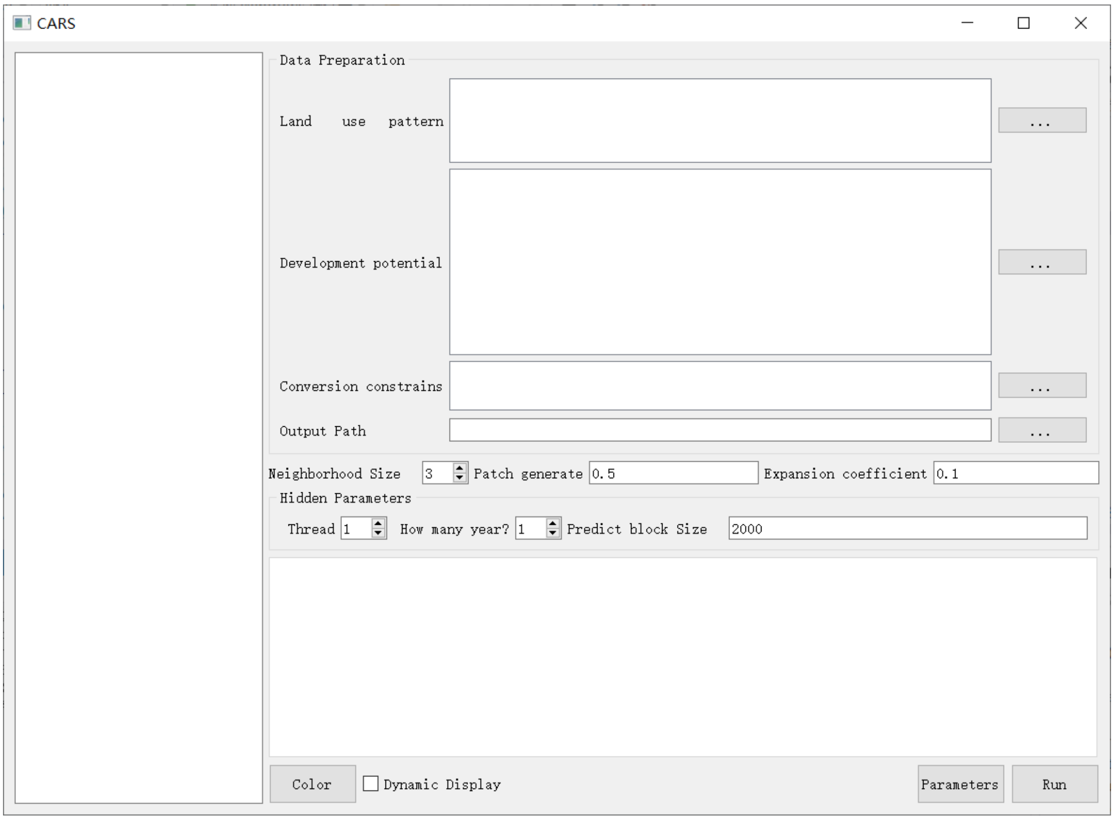


Fig. 13. The UI of the ‘CARS’ module.

4.4.2. Input initial land use data

Click the  button in the “**Land use pattern**” text line and select the land use data of the start year “wh2003_refy.tif” in the pop-up dialog.

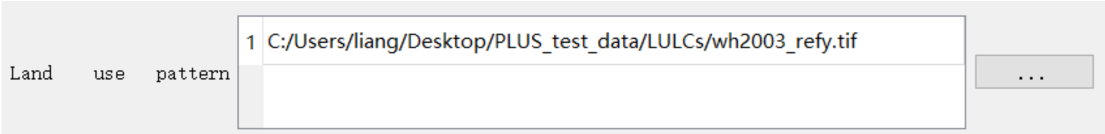


Fig. 14. Select the land use structure data at the start year.

4.4.3. Input development potential maps

Input the development potential file generatedfrom the ‘LEAS’ module in the “**Development Potential**” group box.

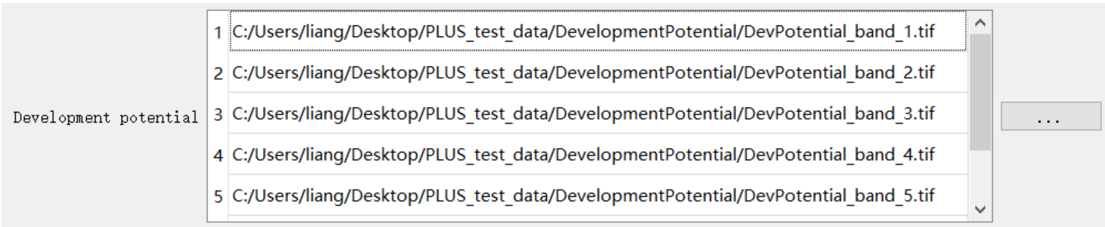


Fig. 15. Select the potential development maps generated by the LEAS module.

4.4.4. The constraint of open water (Spatial policies)

Some policies restrict all types of land change in designated areas, such as open water and some nature conserved areas. Users should prepare a binary image of restricted area that only contains value of 0 or 1. The value 0 means no conversion while the value 1 meansconvertible. In this case, we assume that open water (a sub-category of waterbody) is not allowed to be converted to other land use components.

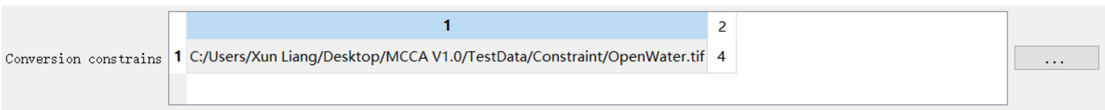


Fig. 16. Select the change probability data.

4.4.5. Set the path of saving simulation result

Select the save path of the simulation result in the “**Output path**” group box.

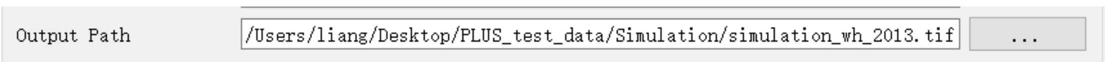


Fig. 17. Select the output path.

4.4.6. Simulation parameters

Simulation parameters include the land use demand, transition matrix, and neighborhood weights. The land use demand is an vital parameter of the PLUS modelthat needs to be firstly set according to the actual situation of the study region. We suggest users to apply external models to determine the land use demand for future scenarios.**In the testing, the time span for the land use change simulation is from 2003 to 2013, so we set the land use demand equal to the land use map of 2013.** <u>We also provide TWO simple tools for predicting future land demand (section 3.4)</u>. We will introduce the two modules in the following section.

The PLUS model will automatically stop when the allocated area equals to the demanded area for all land use types. In the tab widget, users can input the future land use demand of each land use type in the second row “Future Amounts” of the “Land Demand” page is. The “Start Amounts” will automatically load. Users can also input the future cell amount of each land use type in predicted year as figure 18 shows.

Land Demands | Transtion Matrix | Neighborhood Weights

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
Start Amounts	0	0	0	0	0	0	0
Future Amounts 1	43394	231113	2477678	929417	39413	1453586	150790

Fig. 18. Input the future land use demand.

In addition, the neighborhood effect of PLUS is similar to the traditional CA model. Here the default value of the **neighborhood effect** is 3. The **patch generation** is a decay threshold for generating new patches, ranging from 0 and 1. A higher decay threshold means a more conservative transition strategy, which makesthe cells with lower overall probability are less likely to to change. The **expansion coefficient** is a parameter adjust the ability of model to generate new land use patches, ranging from 0 to 1. A higher expansion coefficient means a higher ability to generate new patches.

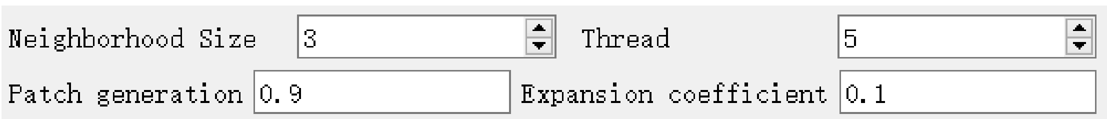


Fig. 19. The simulation parameters of the PLUS model.

Then, switch the tab widget to the “**Transition Matrix**” page. The columns of this matrix indicate the current land use types, and the rows indicate the future land use types. A value of 1 means the conversion is allowed, while a value of 0 indicates that the conversion is not possible. A **Transition Matrix** is a collection of a series of prior knowledge of the expertise about the study area. An example of the transition matrix is shown below:

Land Demands | Transtion Matrix | Neighborhood Weights

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
Type 1	1	1	1	1	1	1	1
Type 2	0	1	0	0	0	0	0
Type 3	1	1	1	1	1	1	1
Type 4	0	0	0	1	0	0	0
Type 5	1	1	1	1	1	1	1
Type 6	1	1	1	1	1	1	1
Type 7	0	0	1	0	0	0	1

Fig. 20. Transition Matrix.

In the "**Neighborhood weights**" page, users need to set the neighborhood weights of different land use types. The neighborhood effects may be different for different land use types in a unique study region. The value of the neighborhood weight for each land use type can be determined according to expert knowledge and a series of model tests, ranging from 0-1. Or the users can also determine the neighborhood weight of each land-use type by calculating the ratio of the expansion areas of a land-use type accounting for the total land expansion based on the land expansion map (the map outputted in section 4.2). The users can still modify them according to the expert knowledge. According to the land expansion map, we set the parameter as below:

Weights | Transtion Matrix | Land Demands

	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6	Type 7
Weights	0.033354448	0.072010689	0.214762891	0.449313208	0.027947715	0.130501503	0.072109546

Fig. 21. Neighborhood weight for each land use type.

4.4.7. Setting the colors of all land use types for dynamic display

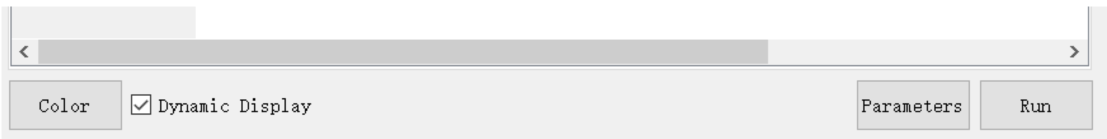


Fig. 22. The position of the ‘Color’ button on the UI.

Click the “**Color**”button to activate the ‘Set Color Display’ module.

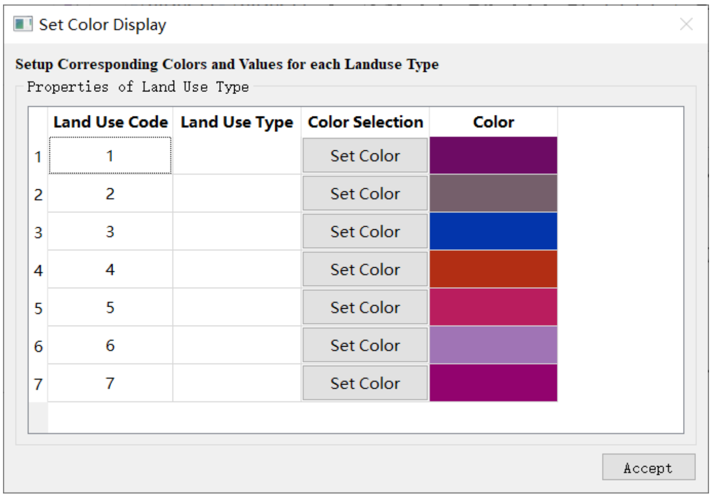


Fig. 23. The ‘Set Color Display’ interface.

Click the “Set Color” button of each row in the list box to set an RGB value (such as rgb(150,200,50)) for all land use types. Then, input the name of each land use type in the ‘Land Use Type’ column. Finally, click the “**Accept**” button, confirm the above settings.

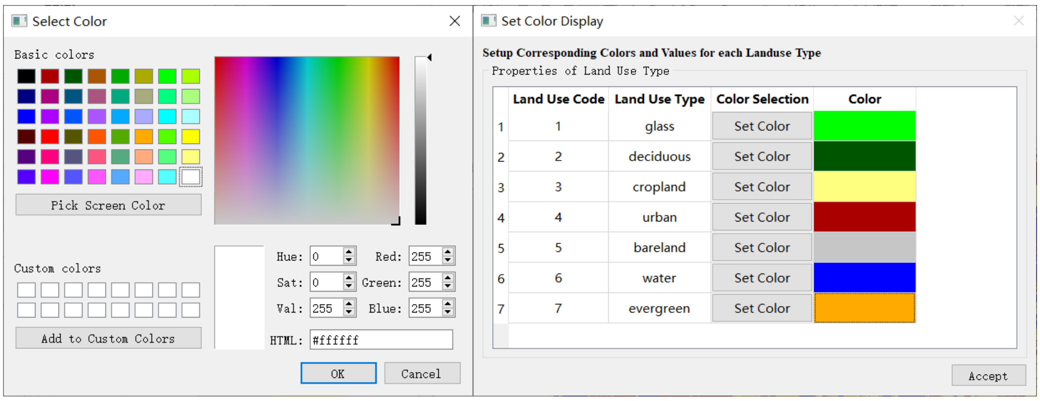


Fig. 24. Setting color for all land use types.

Note that the PLUS model allowed the users to not activated the dynamic display function by unchecking the ‘Dynamic Display’ checkbox. Because the display function can cause software crashes in some situations. Running the PLUS model without a dynamic display can also make the model run faster.

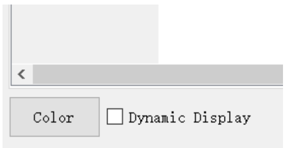


Fig. 25. Cancel the dynamic display by unchecking the ‘Dynamic Display’ checkbox.

4.4.8. Running the CARS

Finally, click the button Run to start the simulation. The information of each iteration is shown on the left of the interface of ‘Simulating dynamics of land use structures’ module. Also, the PLUS model can presentthe process of land use change dynamically, including the land cover spatial change, value change and change curve of each land use type on the main interface.

The visualization panel allows the users to zoom in and observe the land use dynamics processes of study regions. The horizontal axis of the dynamic chart represents the number of iteration, and the vertical axis represents the amount of cell.

The model will stop and save the simulation result when the allocated area equals to the demanded area for all land use types. The PLUS software will save all the parameters to two configuration files: “CARSparameters.tmp” in the ‘Parameterfile’ folder. The module will load the parameters that are recorded in the configuration files when clicking the Parameters button, which is convenient for the user to use to repeat their experiments.

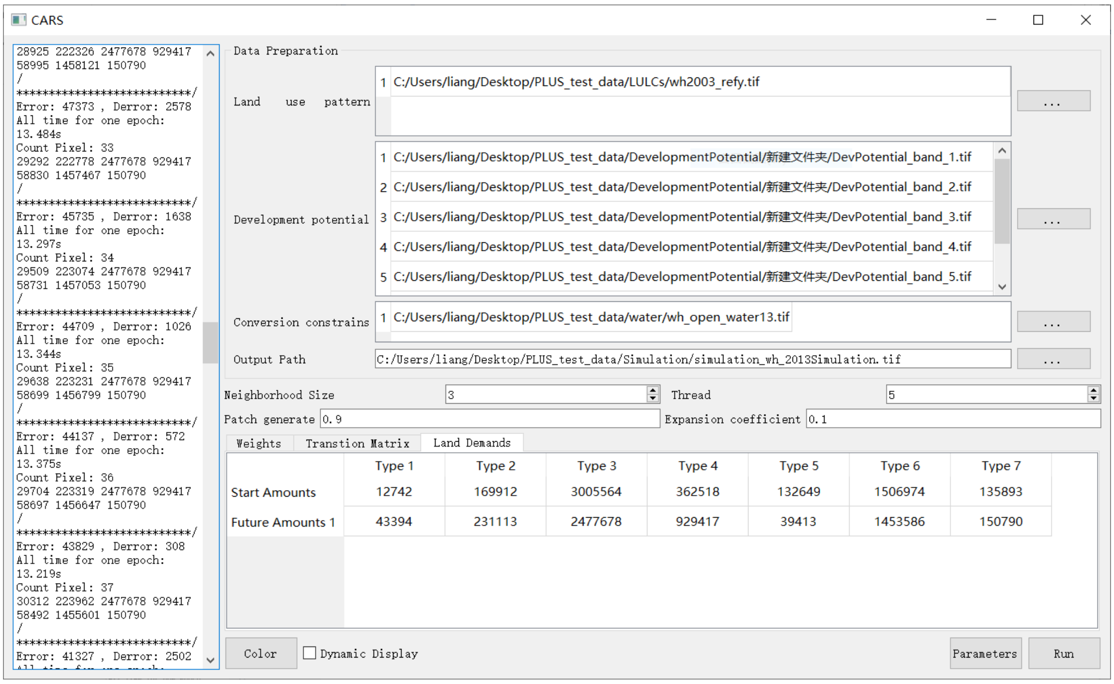
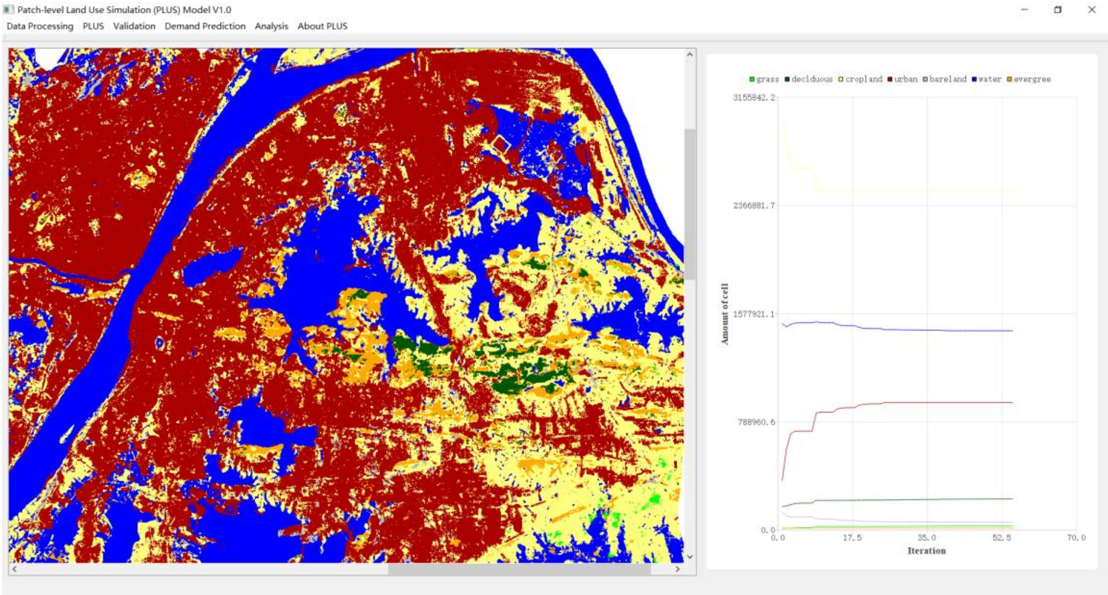


Fig. 26. The dynamic information of each iteration is shown on the left.



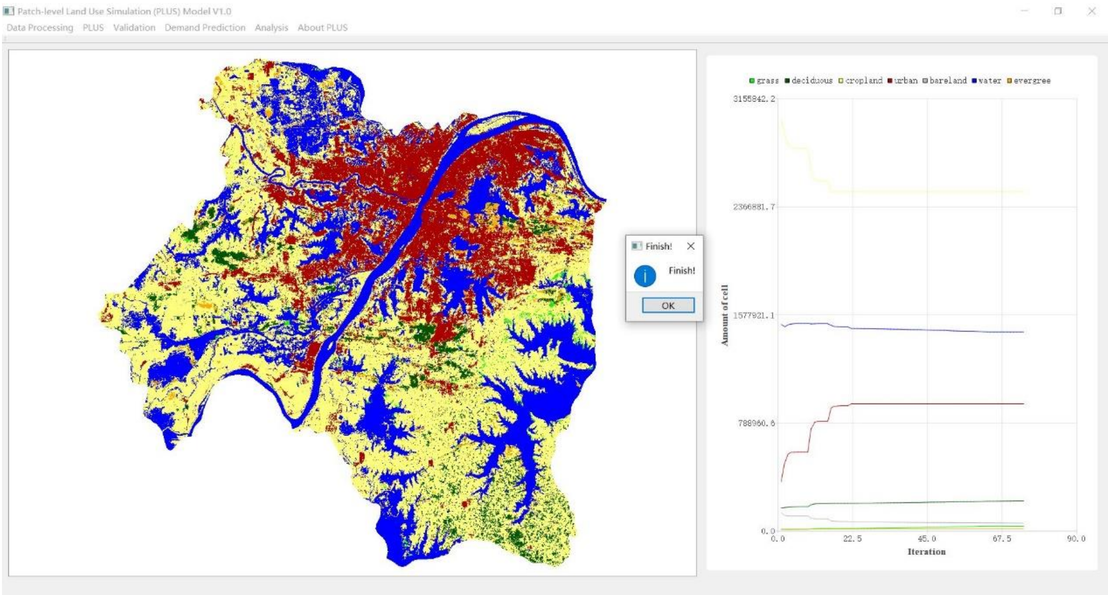


Fig. 27. The dynamic display of the simulation process on the main interface.

4.5. Accuracy validation of simulation

The PLUS software provides a kappa statistic tool and an FoM statistic tool for evaluating the agreement between simulation results and actual land use patterns.

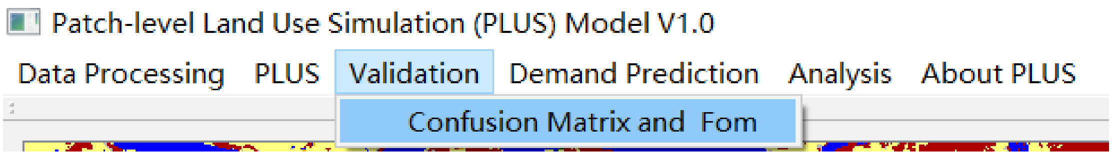


Fig. 28. Start the validation module.

4.5.1. Kappa statistic tool

Click the **Validation → Confusion Matrix and Fom** item on the main menu to start the kappa statistic tool.

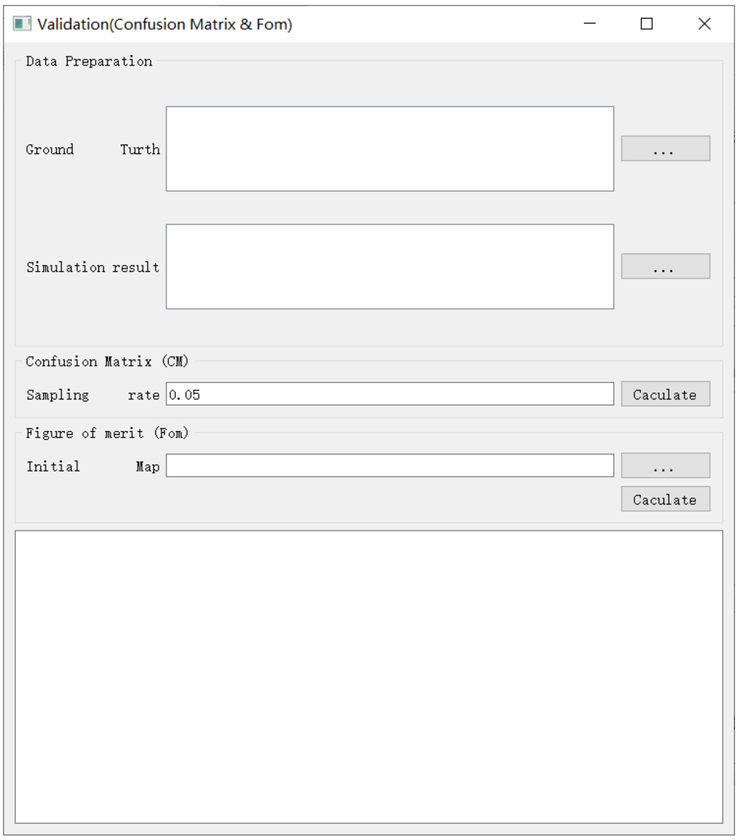


Fig. 29. UI of the validation module.

4.5.1.1. Load data and set the sampling rate

Click the  button in the "Ground Truth" row，to input the actual land use data"wh2013_refy.tif". Then click the  button in the "Simulation Result" row to input the simulated land useresults. In this case, we set the sampling rate to 10%.

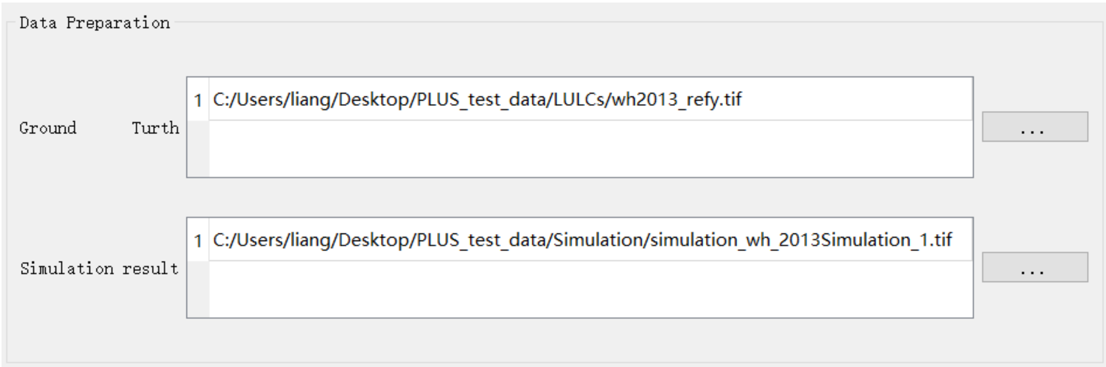
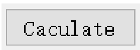


Fig. 30. Input the ground truth and simulation result.

4.5.1.2. Start calculating and save the result

Click the 

button to start calculating the Kappa coefficient. A message box will pop up when the program completes the calculation.

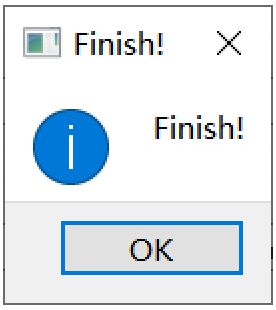


Fig. 31. The message box indicates that the calculation has finished.

The results include the Kappa coefficient, overall accuracy and confusion matrix, which will be saved in a file named "Kappa.csv" in the 'Parameterfile' folder.

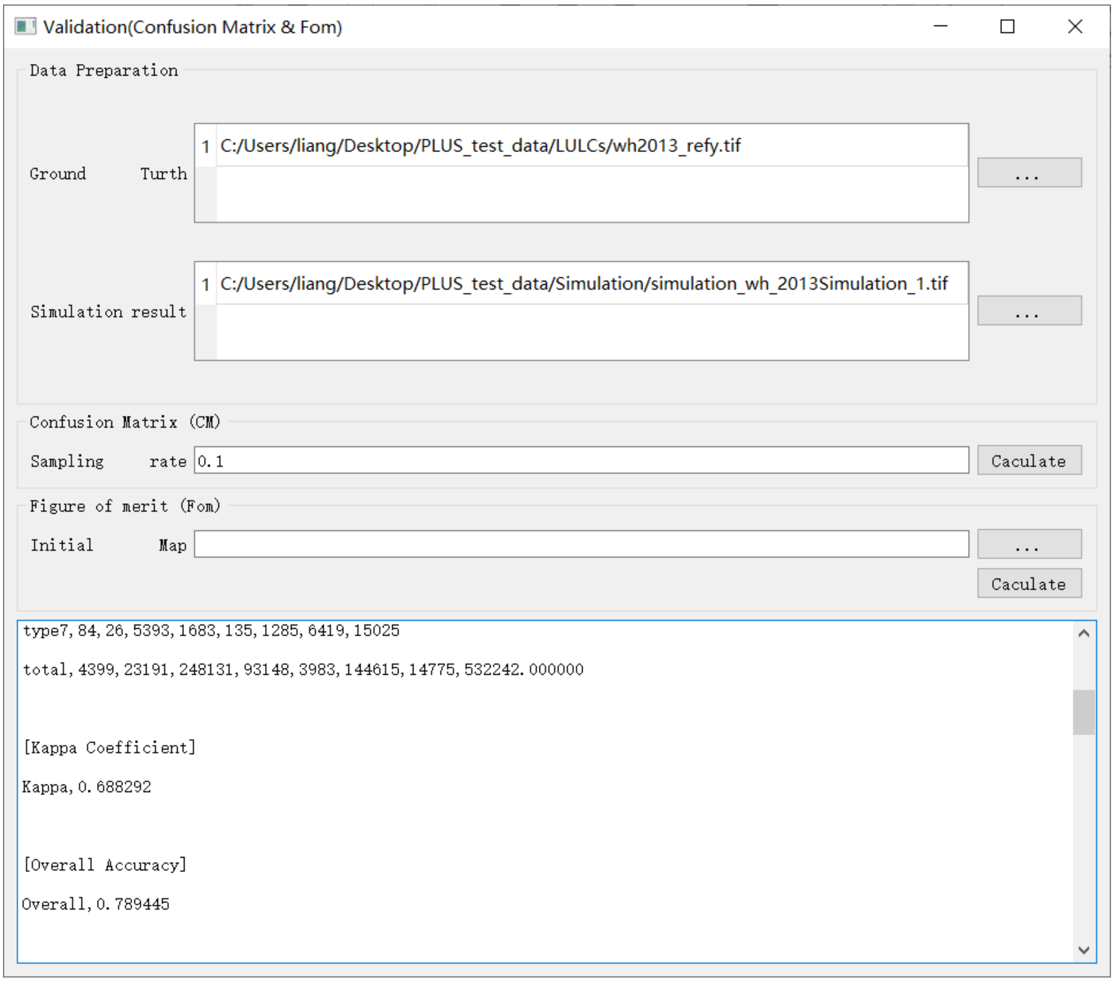
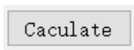


Fig. 32. Kappa coefficient and overall accuracy of the simulation result.

4.5.2. FoM statistic tool

Then click the  button in the "**Initial Map**" row, to input the start land use data "wh2003_refy.tif" in 2003. Click the 

button to start calculating the FoM coefficient. A message box will pop up when the program completes the calculation.

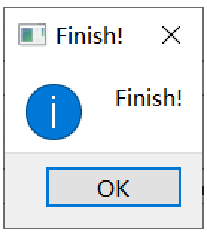


Fig. 29. The message box indicates that the calculation has finished.

The results include FoM statistics and producer's accuracy, which will be saved in a file named "FoM.csv" in the 'Parameterfile' folder.

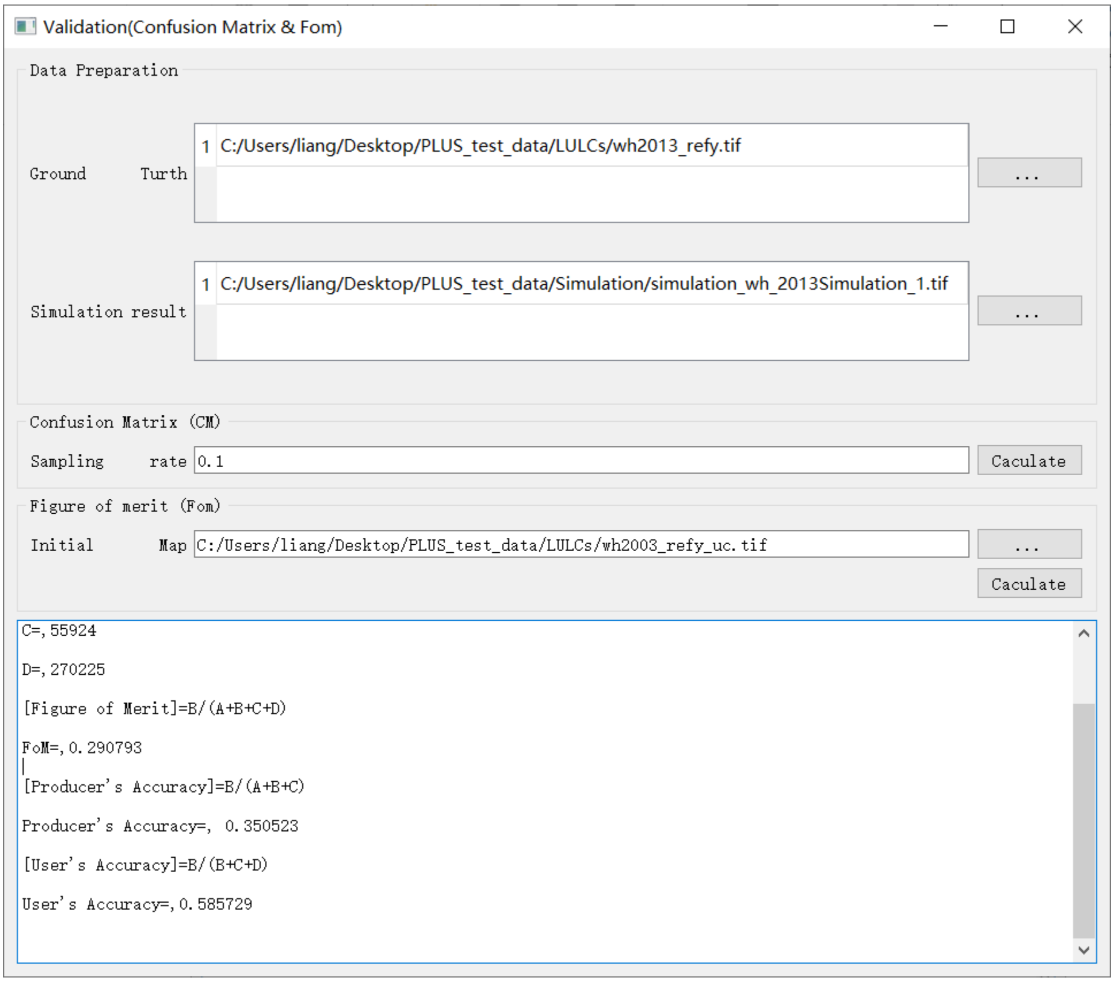


Fig. 33. The figure of merit of the simulation result.

4.6. Projecting future land use demand

Future land use demand can be determined by many methods, such as expert experience, linear regression, Markov chains, the system dynamics model, or an integrated assessment model. The PLUS provides a linear regression method and Markov chain method. This experiment projects future land use demand based on historical data.

4.6.1. Linear regression method

Click the "Linear regression" item to start the prediction module.

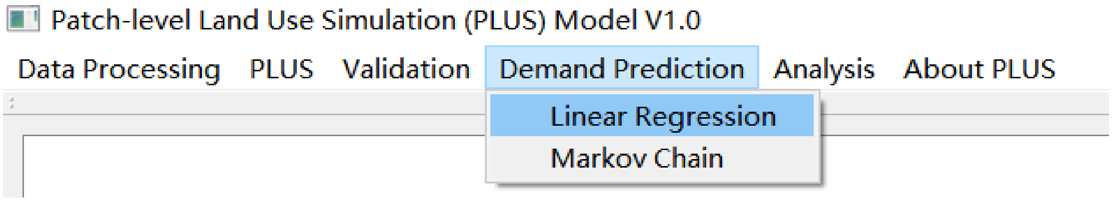


Fig. 34. Activating the land use demand projection module.

4.6.1.1. <u>Input historical land use data</u>

Click the  button, select the historical land use data. In this case we picked the data in the year of 2003, 2005, 2010, and 2013.

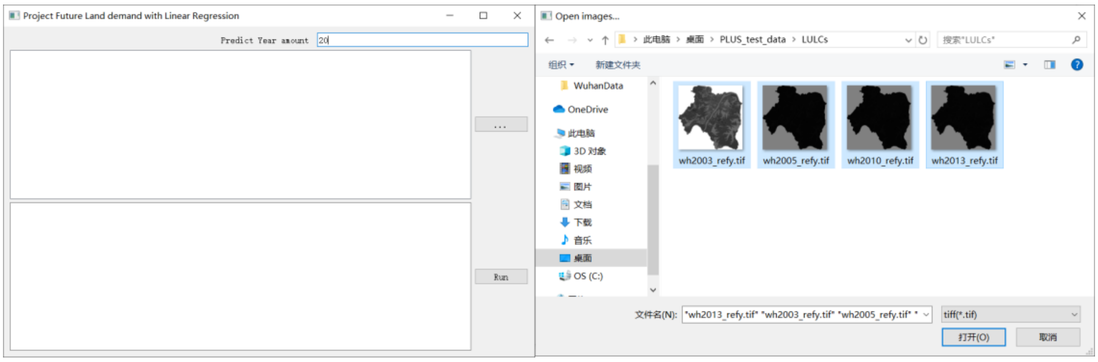


Fig. 35. Select the historical land use data.

The program can automatically read the year on the filename. The uses can also manually input the year to the ‘Year’ column on the chart of data. We set the predict year amount to 20 and the module will predict the future land use demand from 2013 to 2033 (2013+20).

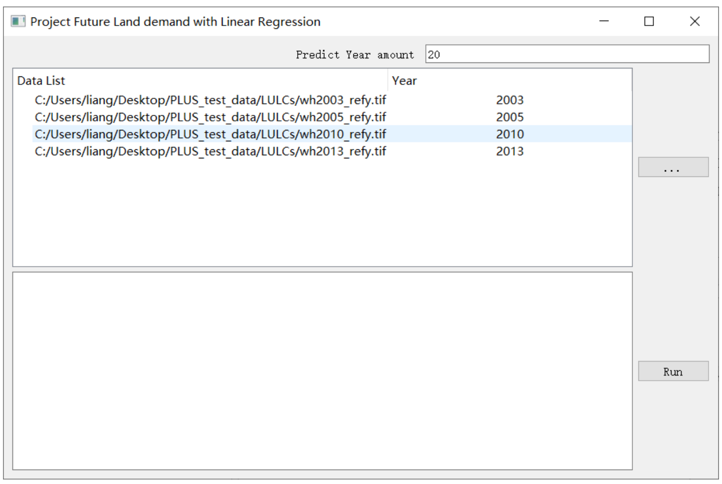
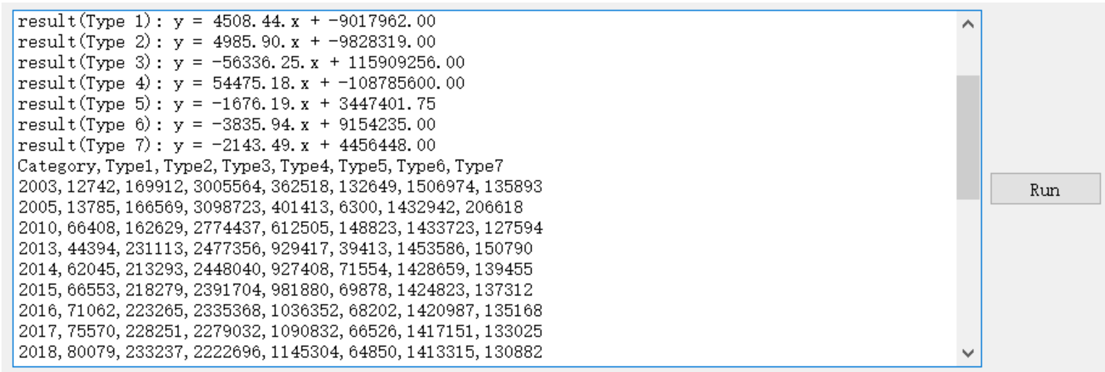


Fig. 36. Inputting the historical land use data to the Demand projection module.

4.6.1.2. Running linear regression and predicting future demands

Click the 

button to start calculating future land use demand. A message box will pop up when the program completes the calculation. The results will be showed on the interface and saved in a file named "PredictDemand.csv" in the folder 'Parameterfile'.The formulas of linear regression can also be shown.



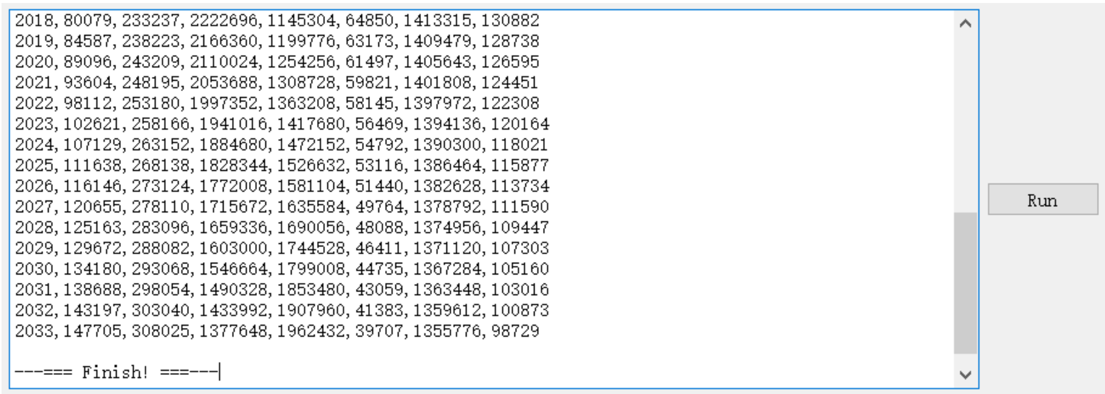


Fig. 37. Running the linear regression and predicting future land use demands.

4.6.2. Markov chain method

4.6.2.1. Input historical land use data

Click the “Markov chain” item to start the prediction module.

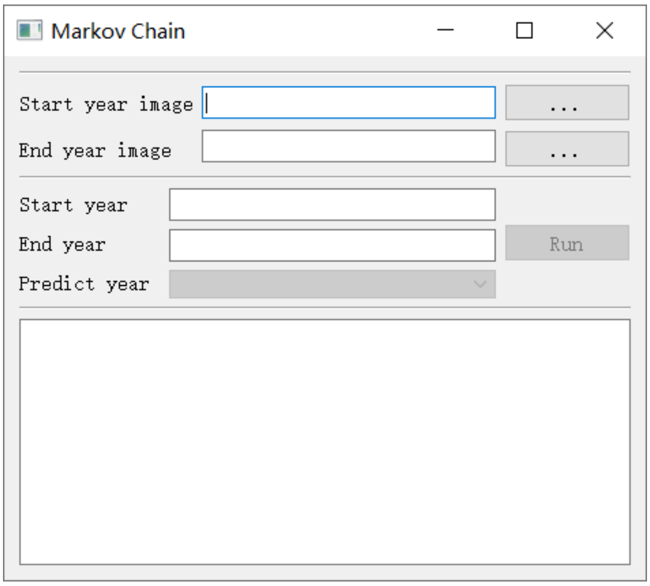


Fig. 38. The interface for the Markov chain.

Click the  button in the “**Start year image**” row to input the land use data of start year “wh2003_refy.tif”. Then click the  button in the “**End year image**” row to input the land use data of end year “wh2013_refy.tif”. Then click the text plain below and activated the “**Predict year**” pull-down menus. Users can choose the future year they want to predictsuch as 2033 year.

4.6.2.2. Predicting future land demands with Markov chain

Click the 

button to start calculating future land use demand. A message box will pop up when the program completes the calculation. The results will be showed on the interface and saved in a file named “MakovChain.csv” in the folder ‘Parameterfile’.

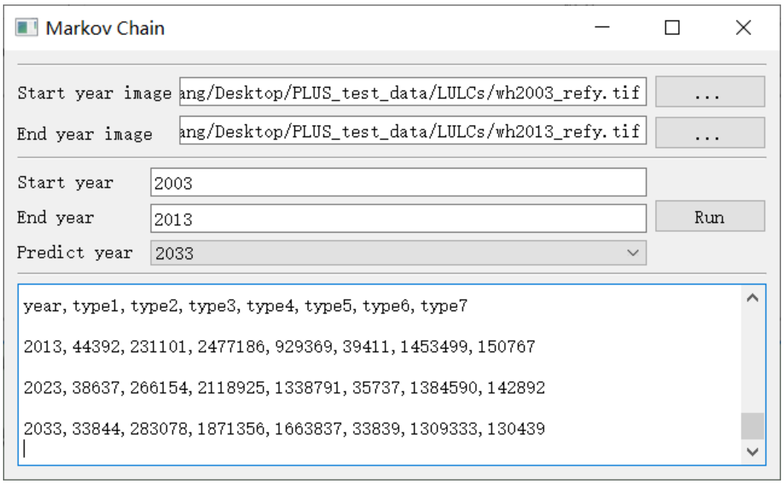


Fig. 39. Predict future land use amount with Markov chain.

4.6.3. Simulating future land use from 2013 to 2033

Click the  button in the “Land use pattern” text line and select the land use data of start year “wh2013_refy.tif” in the pop-up dialog. Then input the development potential file from 2003 to 2013 generated by the ‘LEAS’ module in the “**Development Potential**” group box. Input the open water map in 2013 “wh_open_water13.tif”as land use constraints, and define an output path for the simulation result.

Other simulation parameters (Neighborhood size, Patch generating factor, Expansion coefficient, Neighborhood weights, and Transition matrix) remain the same as the validation process (section 3.4.6). Then users need to import the predicted land use demand in the year of 2033. We analyzed two future scenarios in this tutorial: 1) Linear regression scenario and 2) the Markov chain scenario with separately running the PLUS model. Note that the final land use amount will approximate but may not exactly equal to the land use demands. Because the final land use amounts not only determined by the land use demands from “top-down” scale but also determined by the local geographical conditions from the “bottom-up” scale. The final land use amounts are final results of the interactions between “top-down” and “bottom-up” effects.

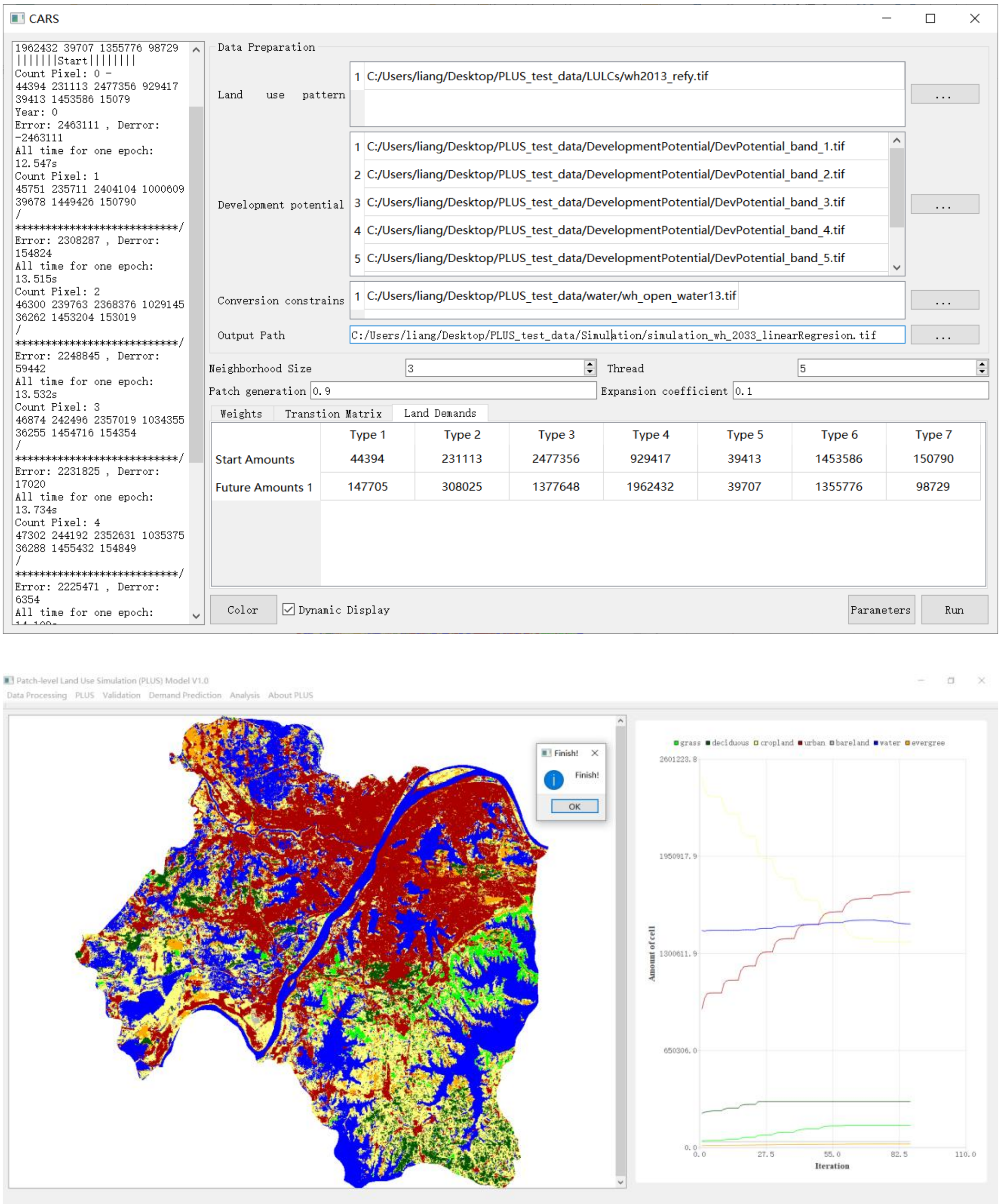


Fig. 40. Parameter setting and simulation result from for the year 2033 under the Linear regression scenario.

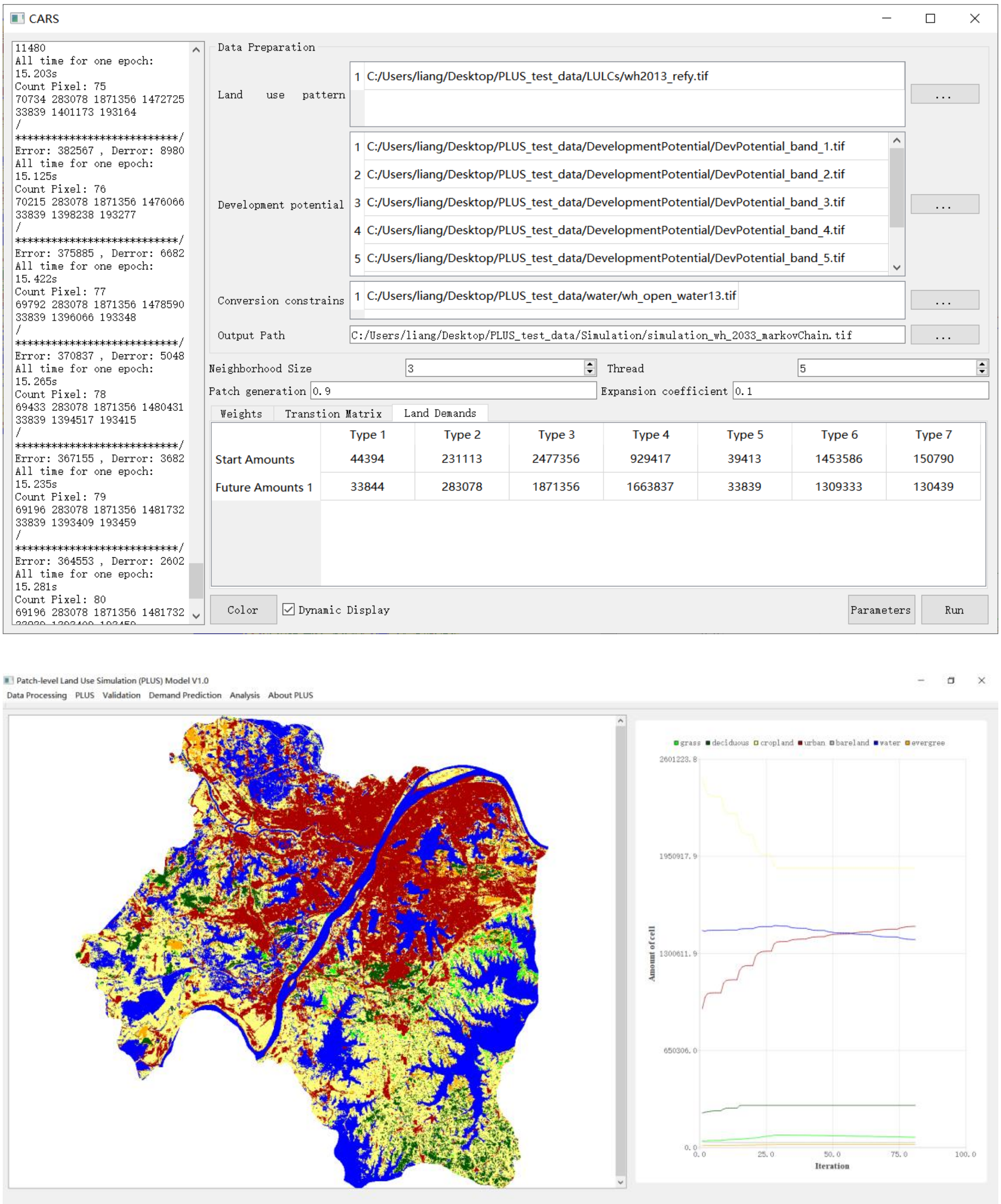


Fig. 41. Parameter setting and simulation result for the year 2033 under the Markov chain scenario.

The PLUS software will save all parameters to two configuration files: "CARSparameters.tmp" in the 'Parameterfile' folder. The module will load the parameters that are recorded in the configuration files when clicking the Parameters buttom, which is convenient for the user to use to repeat their experiments.

4.7. Calculation of scenario diversity

Scenario diversity describes the different of pixels between various scenarios. Areas show more variations between scenarios are more sensitive to the scenario with unique regional environment, while the area likely to remain unchanged regardless of scenario is less sensitive. Areas that change between scenarios (3 scenarios or more) also represent the hotspots of variability in the study regions. This, the PLUS software provides a tool for calculating the scenario diversity.

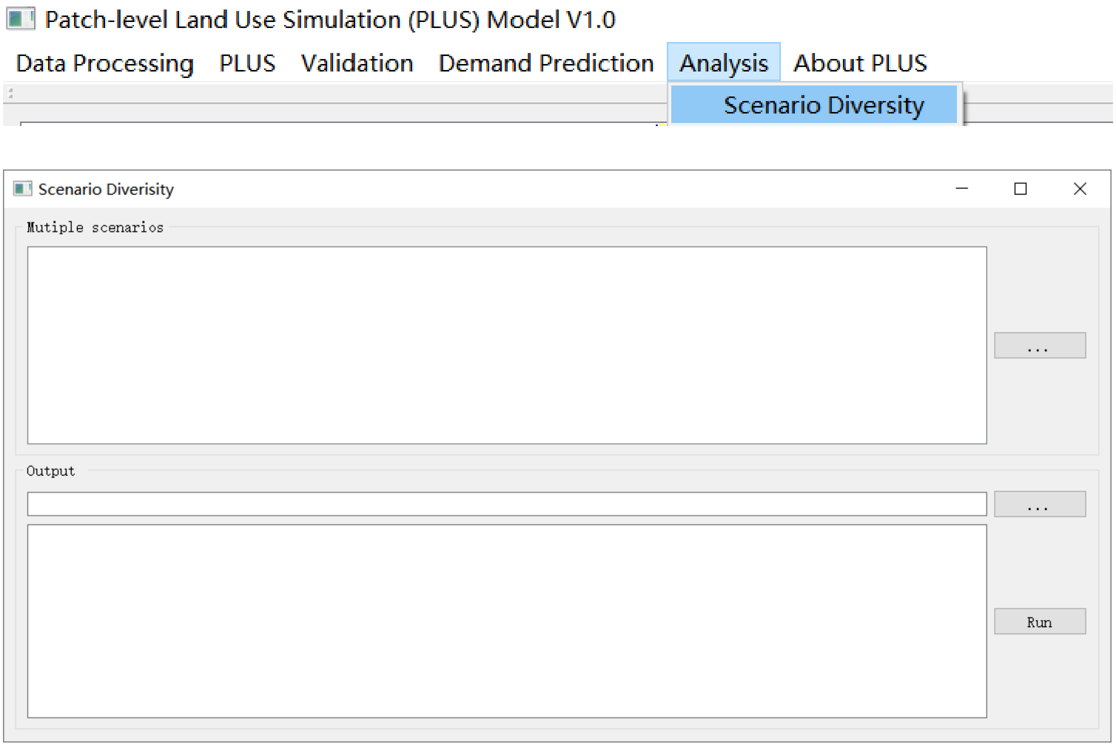


Fig. 42. Activating the ‘Scenario Diversity’ module.

4.7.1. Input multiple simulation results under different scenarios

Click the  button in the “**Multiple scenarios**” row, select the simulation result in 2033 under the two mentioned above scenario in the pop-up dialog. Then click the  button in the“**Output**” row to set the output path of the diversity image.

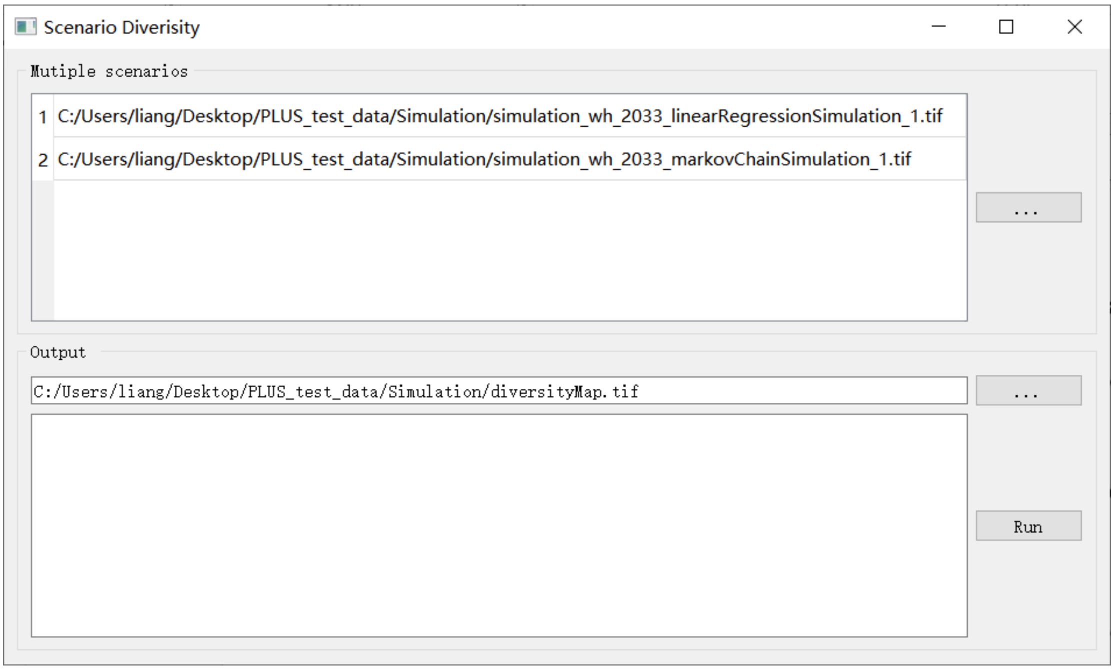


Fig. 43. Set the input and output paths for the ‘Scenario Diversity’ module.

4.7.2. Calculating the land use mixture of each cell

Click the 

button to generate the scenario diversity map in 2033. Note that the ‘Scenario Diversity’ module supports multiple scenarios. This tutorial only provides a simple case study with 2 scenarios.

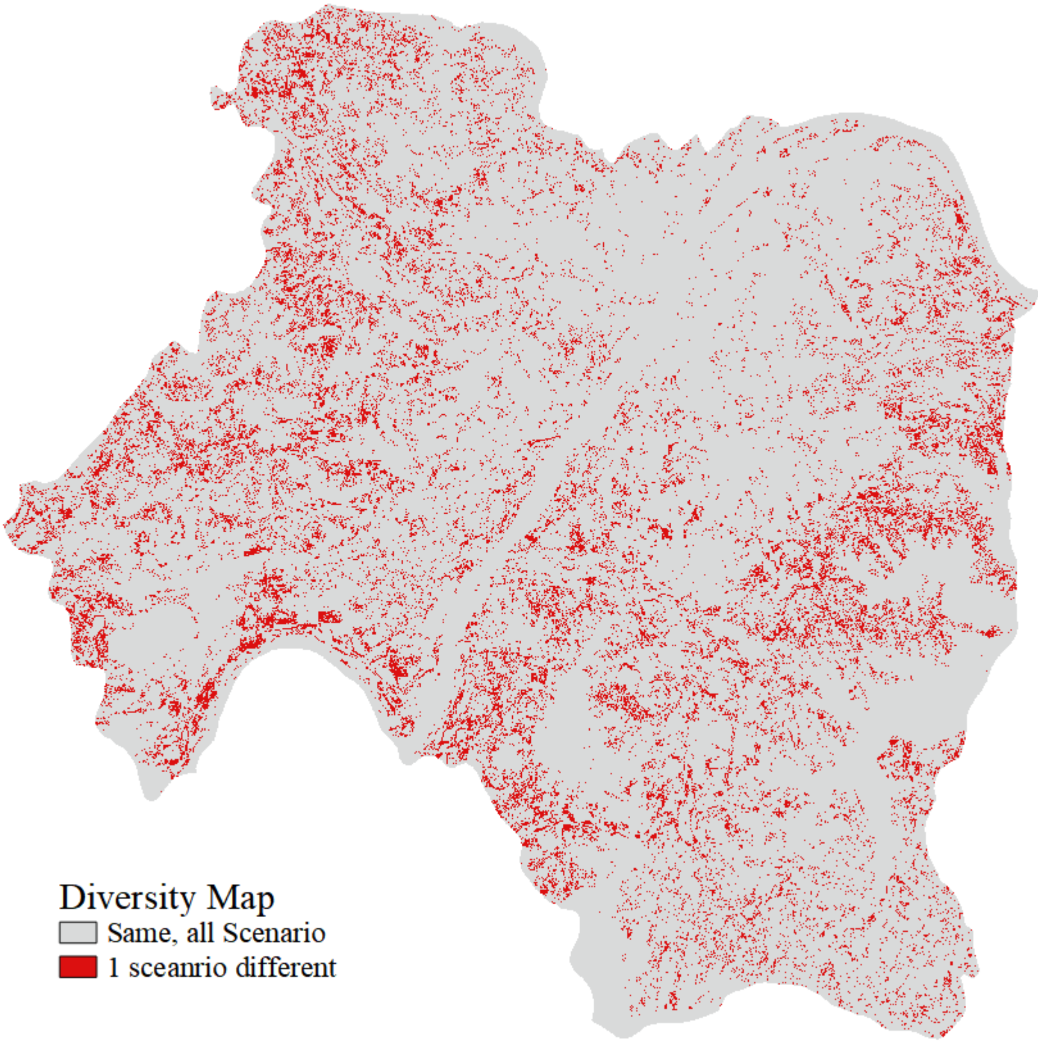


Fig. 44. Generating the scenario diversity map in 2033.