Patch-level Land Use Simulation model V1.0 (PLUS)

User's Manual

Authors:

Xun Liang, Qingfeng Guan*, Shishi Liu, 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

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, little researches have pay attention to promote understanding of the nonlinear relationships underlying LULC. A lack of model ability to reflect patch-level landscape evolution also limit the application potential 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 was used to understand the drivers of land expansion and project landscape dynamics. The PLUS model can obtain higher simulation accuracy and more similar landscape pattern than other models. The LEAS can help researchers find some underlying transition rules. The proposed method combines simulation, knowledge discovery and policy-making, which can provide vital information for both researchers 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 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 is intended to provide users of PLUS a “quick start” on how to use the software. All of the necessary data and files for the tutorial have been provided and these can be used as templates for how to format your own files latter on.

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/). Click the executable files "**PLUS VX.0.exe**" to start the software.

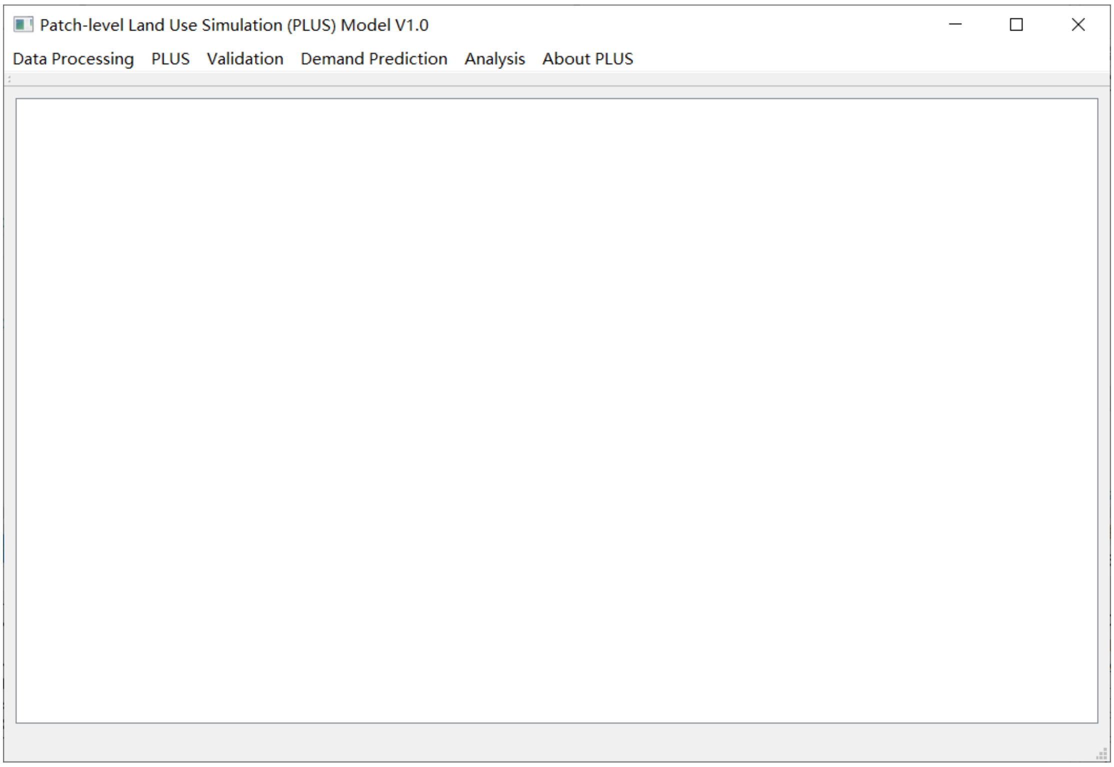


Fig. 1. Main interface of PLUS model.

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

2. 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 highway	wh_dist_highway.tif	OpenStreetMap (https://www.openstreetmap.org/)
	Proximity to arterial road	wh_dist_trunk.tif	
	Proximity to primary road	wh_dist_primary.tif	
	Proximity to secondary road	wh_dist_secondary.tif	
	Proximity to 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	

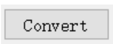
The test dataset are found in the ***TestData*** folder. Use this dataset to run example model or compared with your simulation results. In practical applications, more human-related and natural environment effects can be taken into consideration. The study region in this example is Wuhan, located in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

3. User interfaces of PLUS model

3.1. Data processing

First of all, data processing is extremely important! The PLUS model only support the 'unsigned char' format land use/land cover (LULC) data, whose land use types are encoded as consecutive integers that start from 1 (e,g., 1, 2, 3, 4, 5, 6,……). In addition, 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 you use ArcGIS software to reclassify the LULC data, the ArcGIS usually convert the original 'unsinged 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 be used to convert your LULC data and constraint data (in all format) to 'unsigned char' format. If you are not sure which format your LULC or constraint data are, 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 your land 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 correct for the PLUS software, this tutorial just provide a guidance for how to use the PLUS software. What's more, all the LULC images need to have the same number of rows and columns.

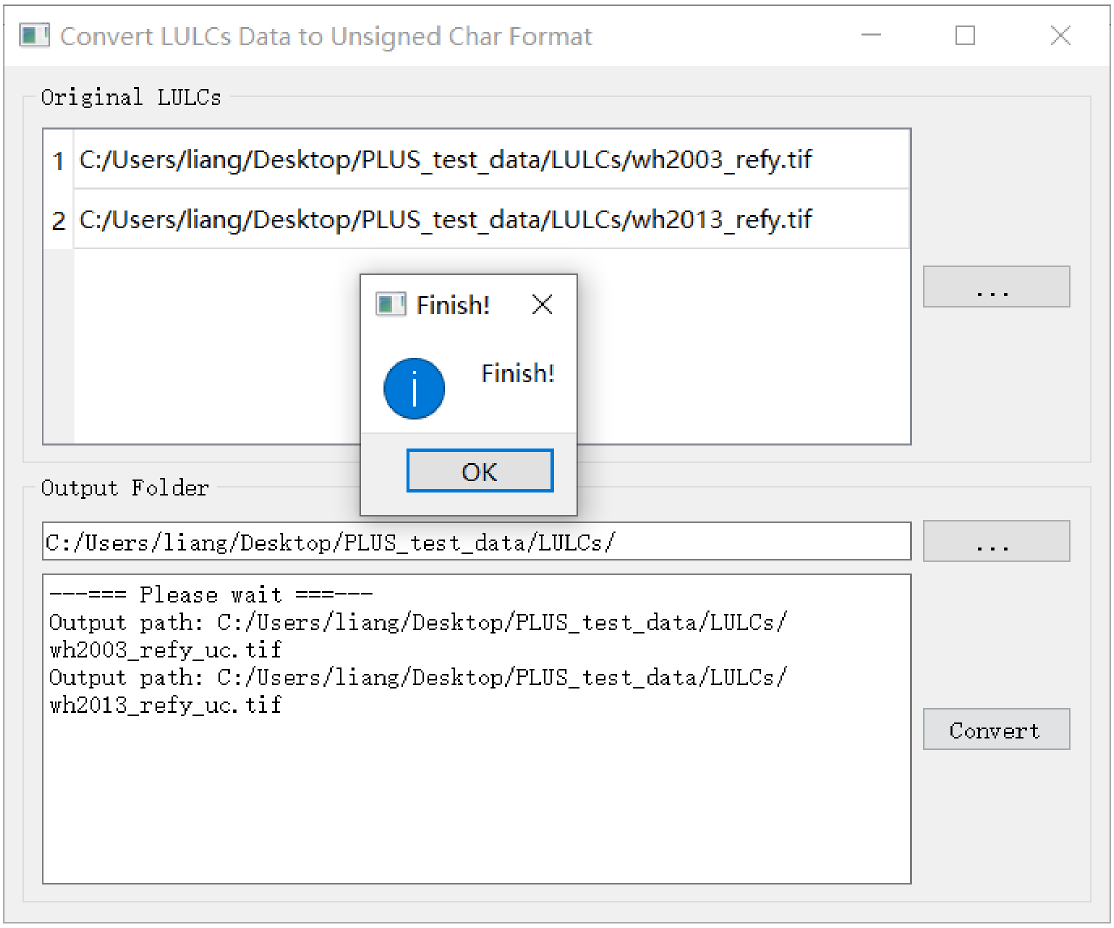


Fig. 2. The UI of the data conversion tool.

3.2. Extract Land expansion

The PLUS consist 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 corresponding module by selecting option from the "**PLUS**" button's pull-down menu on the main menu.

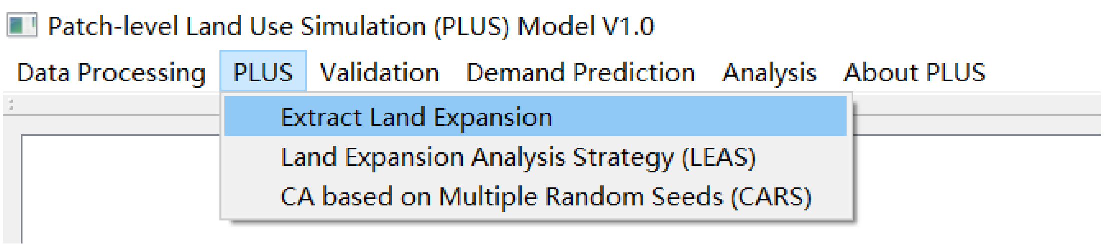
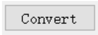


Fig. 3. Starting the 'Extract Land Expansion' module.

Click the 

button and select your land 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.

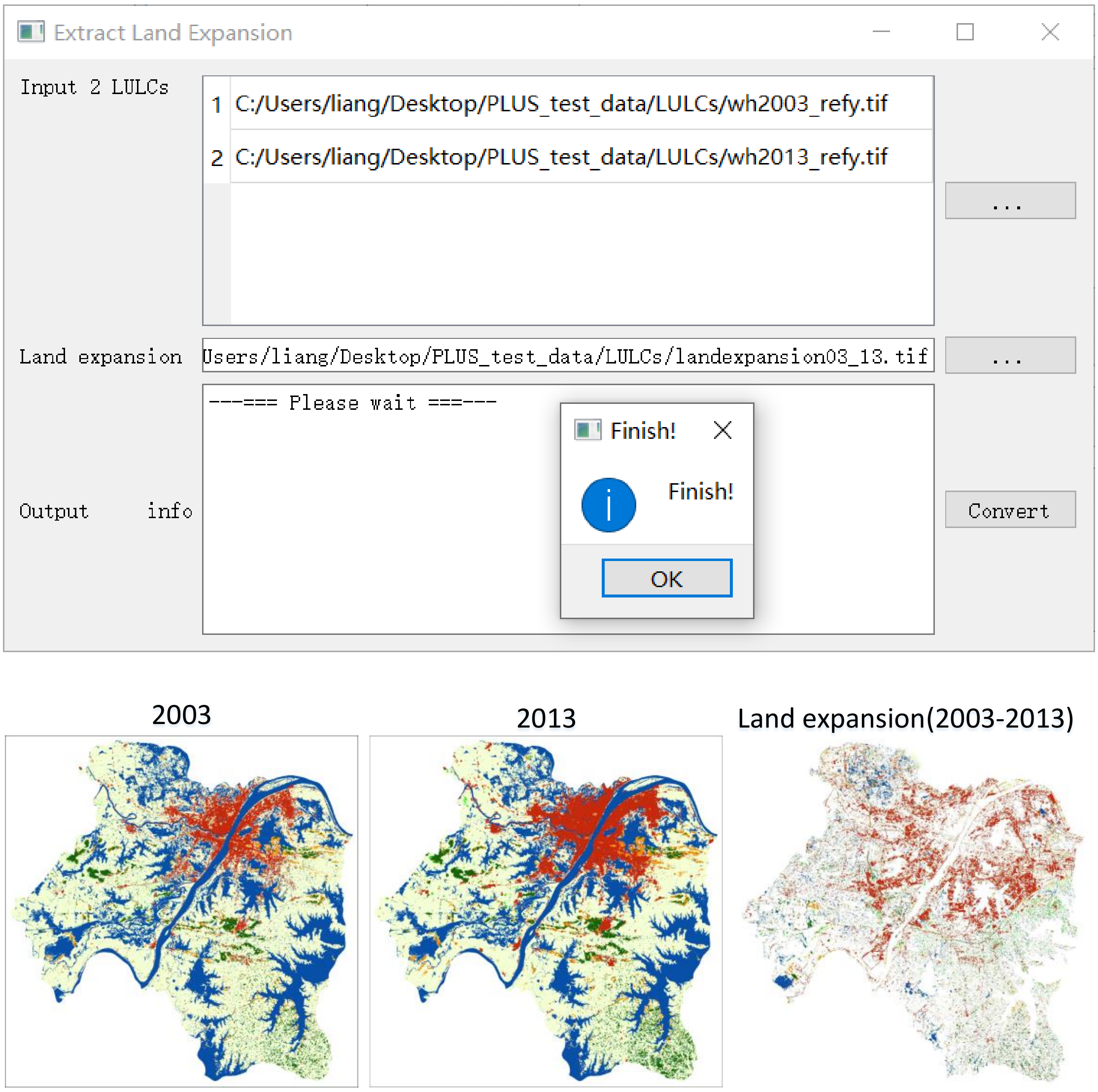


Fig. 4. Extracting land expansion from two periods of land use data.

3.3. Land expansion analysis strategy (LEAS)

3.3.1. Start the module

Click the “**Land Expansion Analysis Strategy (LEAS)**” item to start the module that has the same name as the item.

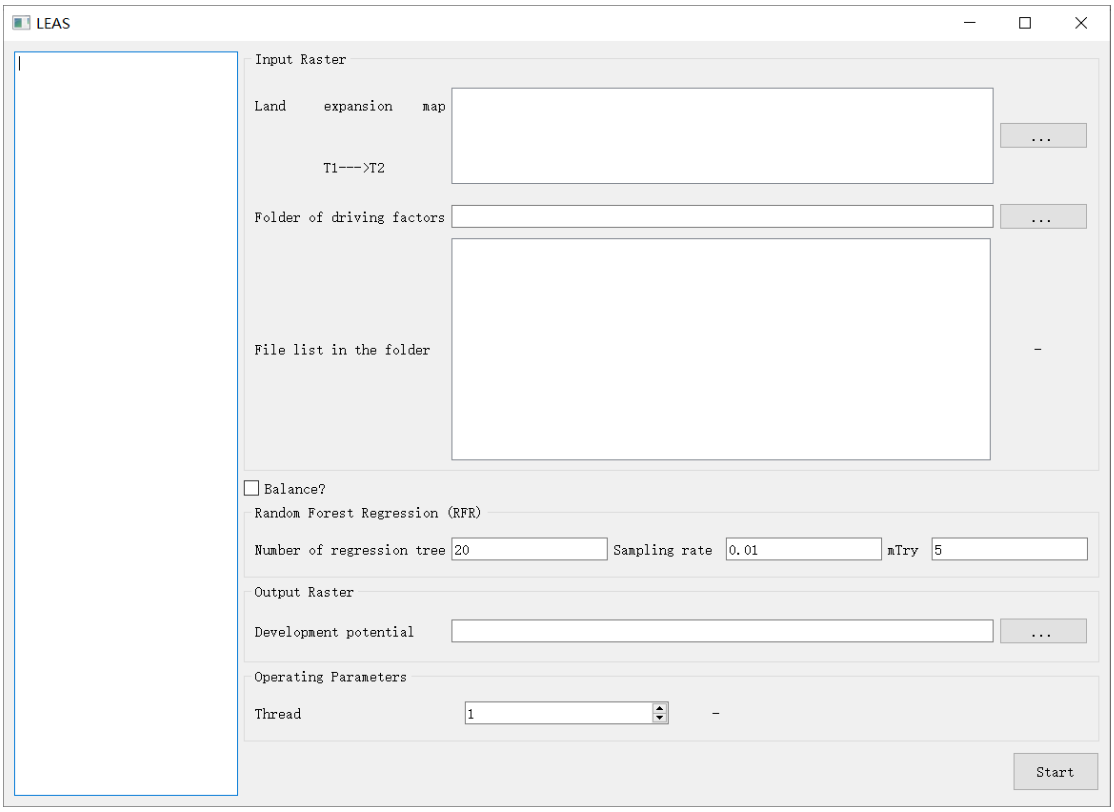


Fig. 5. UI of the 'Land expansion analysis strategy' module.

3.3.2. 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.

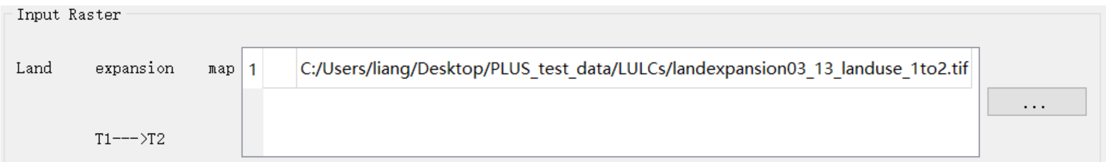


Fig. 6. Select land expansion map outputted in the last step.

3.3.3. 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. Our example data provide 16 driving factors for example run of PLUS model.

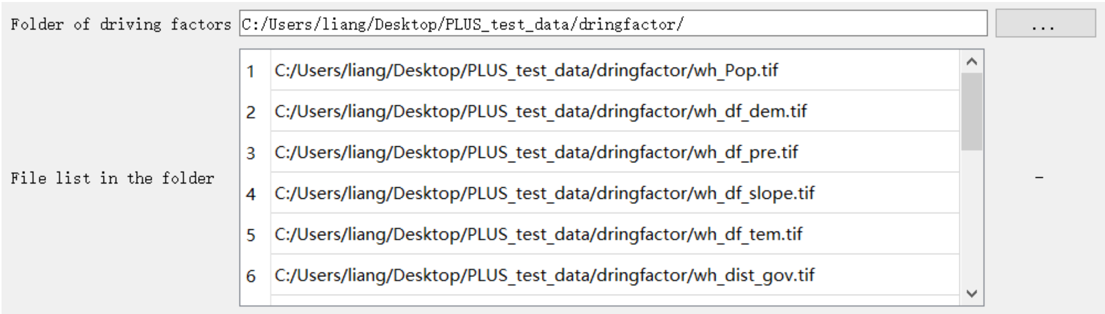
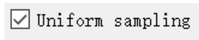


Fig. 7. Input all the driving factors by selecting one folder.

3.3.4. Setting the training parameters and output path

Users are allowed to set the parameters of **Random Forest Regression (RFR)**. Sampling rate is set to 0.01 by default, it means about one percent of pixels is selected for training. The number of regression tree 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 same numbers of sampling points for each land use type. If not, the sampling points will randomly scatter across the study area. Then, click the 

button in the “**Output Raster**” group box and input the save path of change probability in the pop-up dialog. The generated development potential of each land use type is named “devPotential.tif”. The users can increase the number of parallel internal threads to accelerate the running speed.

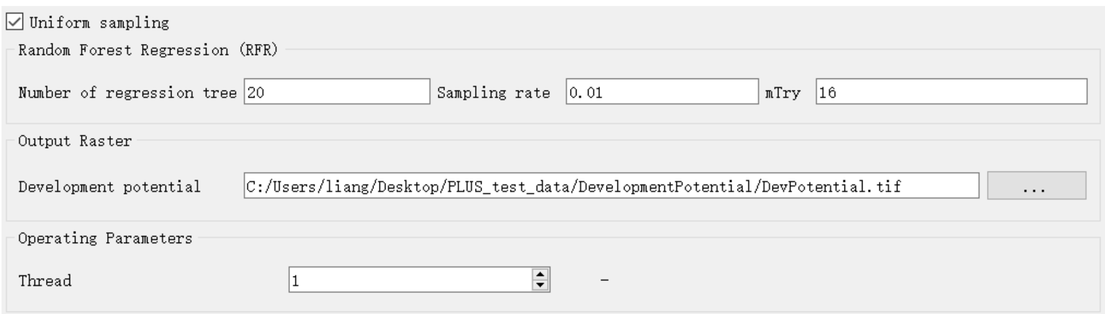
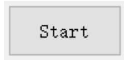


Fig. 8. Determining the training parameters and output path of the RFC.

3.3.5. Running program and examine the result

Once the setup mentioned above has completed, click the button 

for running the module. There will come out a message box as below when the model process run is complete.

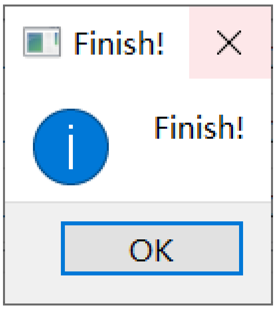


Fig. 9. A message that shows the program run successfully.

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

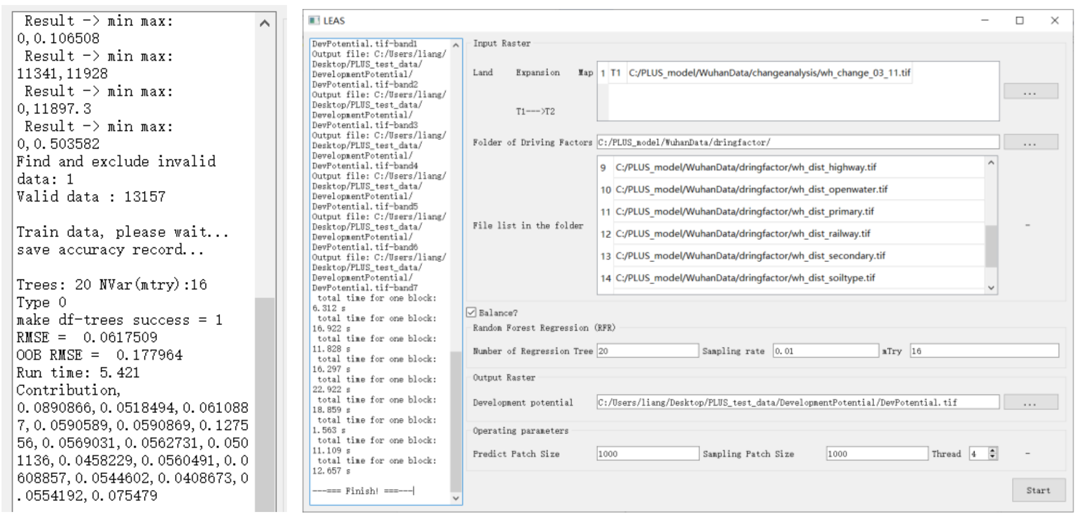


Fig. 10. Accuracy for each land use type.

What’s 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 which will be generated by the PLUS software:

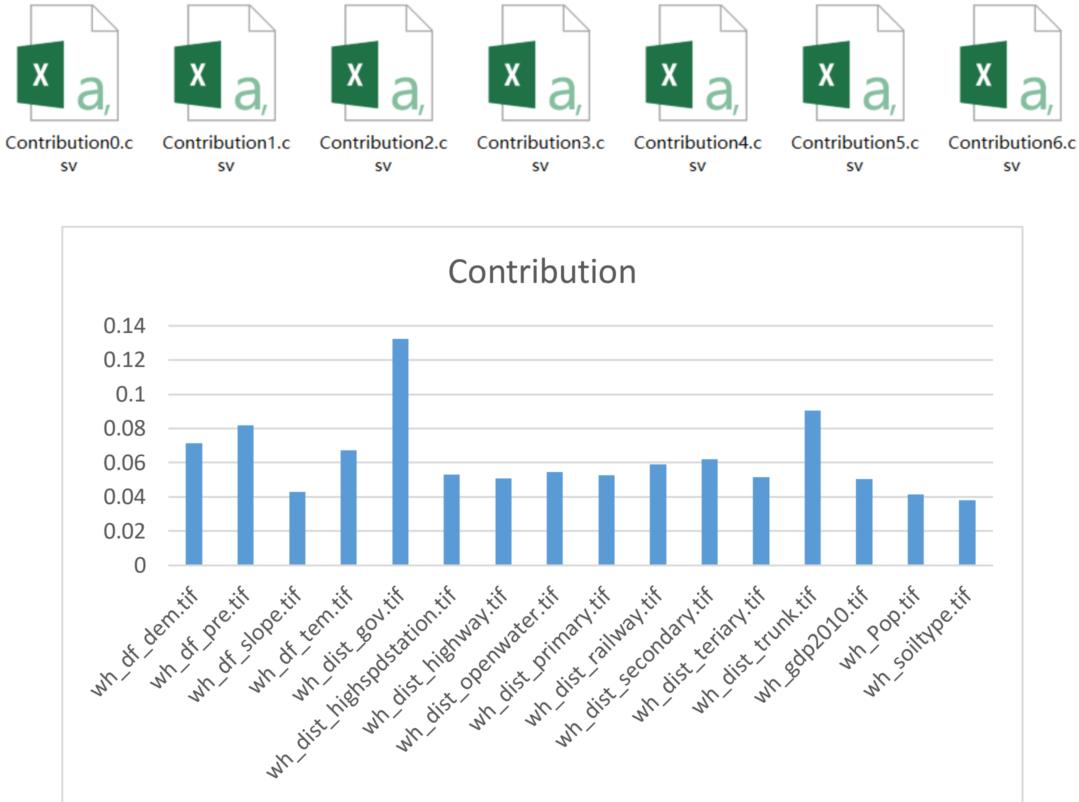


Fig. 11. Accuracy for each land use type.

For example, the file ‘contribution0.csv’ recorded the contribution of various driving factors on grass land. 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 at where is not strongly impacted by human activities, because the areas around governments are generally the most active places for human being.

After 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.

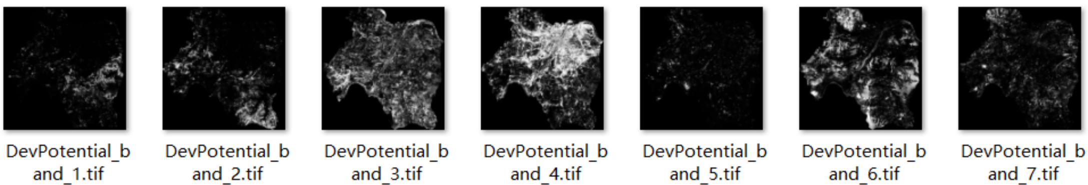
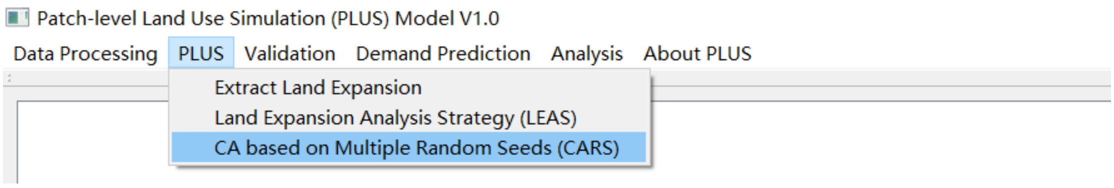


Fig. 12. Development potential map for each land use type.

3.4. Simulating patch-level land use change

3.4.1. Start the module

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



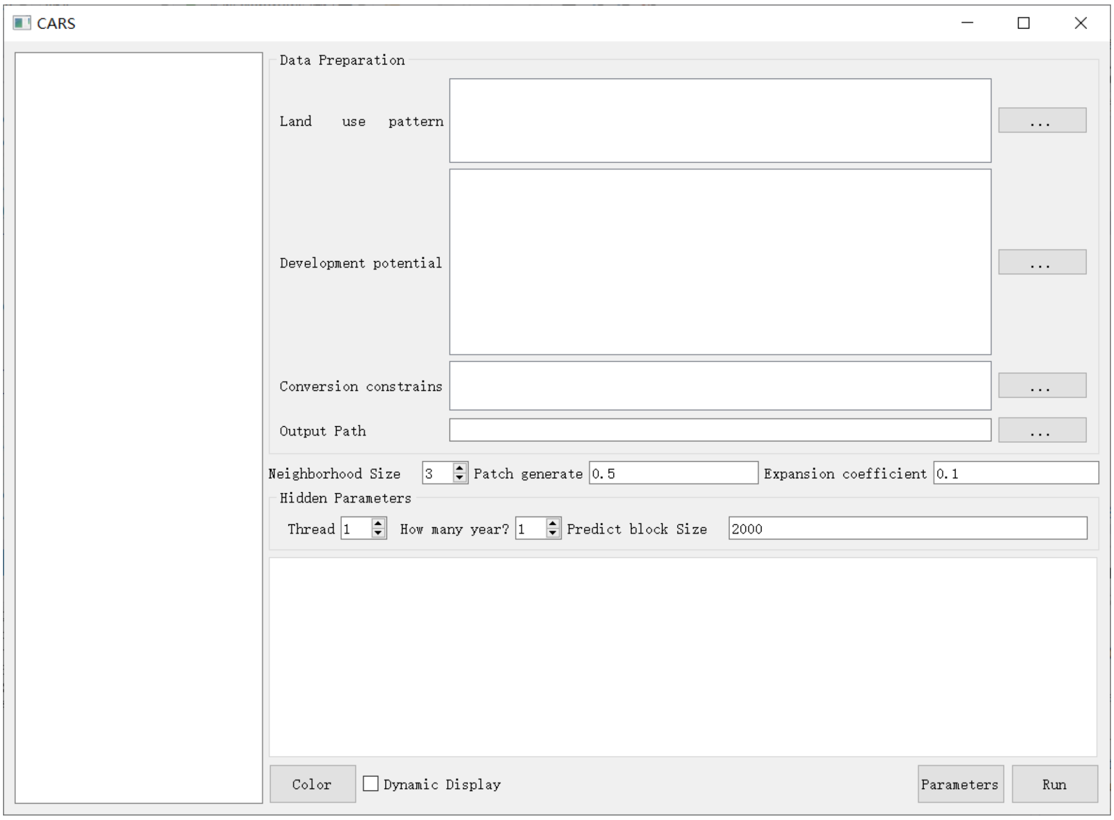


Fig. 13. The UI of ‘CARS’ module.

3.4.2. Input initial land use data

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

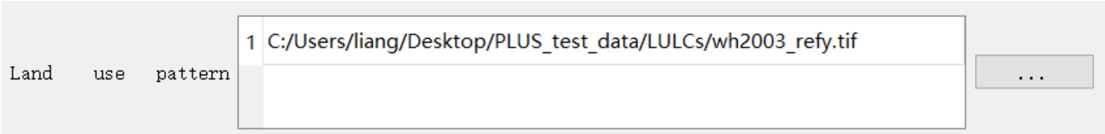


Fig. 14. Select the land use structure data at start year.

3.4.3. Input development potential maps

Input the development potential file outputted from the 'LEAS' module in the "**Development Potential**" group box.

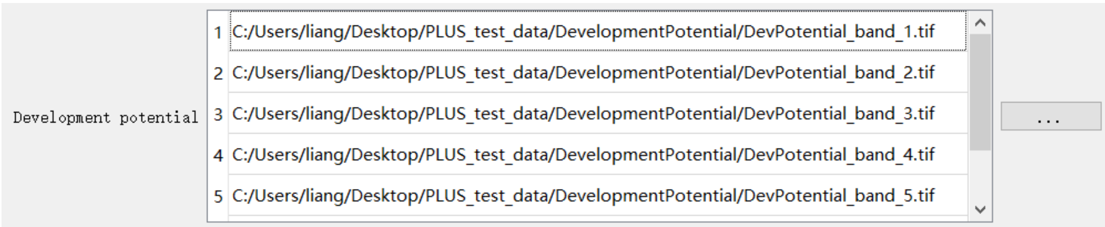


Fig. 15. Select the development potential maps outputted by the LEAS.

3.4.4. The constraint of open water (Spatial policies)

Some spatial policies restrict all land use change in designated areas, such as open water and some strictly protected nature reserves area. When spatial policies need to be considered into simulation, users should prepare a binary restricted area image that all the pixel are either 0 or 1. The value 0 means that grid cells with one land use type can never be converted to another type; 1 means that land use change is allowed in this region. We assume that open water (a sub-category of waterbody) is not allowed to be converted to other land use components.

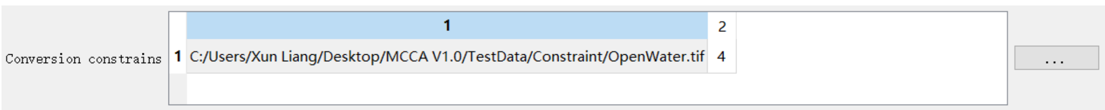


Fig. 16. Select the change probability data.

3.4.5. Set the path of saving simulation result

Input the save path of simulation result in the "**Output path**" group box.

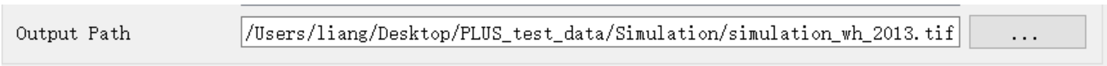
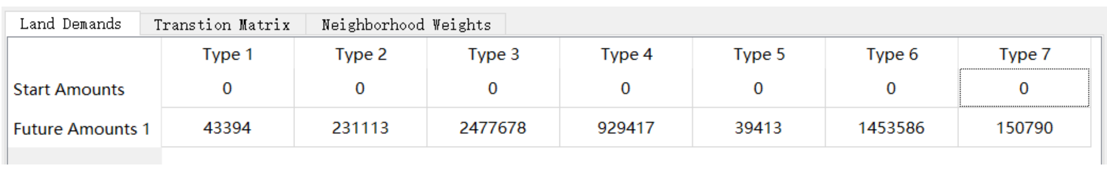


Fig. 17. Select the output path.

3.4.6. Simulation parameters

Simulation parameters include the land use demands, transition matrix and neighborhood weights. The land use demands are parameters of PLUS model, which need to be firstly set according to the actual situation of study region. So users need to firstly use external model to determine the land use demands of future scenario. **For our example run, the time span of the land use change simulation is from 2003 to 2013, so the land use demands are set to be the same as the actual area of various land use types in 2013.** <u>We also provide TWO simple tools for calculating the historical land demand and predict the 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 equal the demanded area for all land use types. In the tab widget, the second row "Future Amounts" in the "Land Demand" page is for inputting the future land use demand of each land use type. The initial pixel number will be automatically loaded after running the program. You can also input the future cell amount of each land use type in predicted year as what figure 13 shows.



	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 GeoSOS-FLUS is similar to traditional CA, the default value is 3. The **patch generate** factor is a decay threshold for generating new patches, ranging from 0 and 1.0. A higher decay threshold means a more conservative transition strategy, which makes the cells with lower overall probability are less likely to change. The **expansion coefficient** is a threshold to generate the new land use patches, ranging from 0 to 1. A higher expansion coefficient means a higher ability to generate new patches.

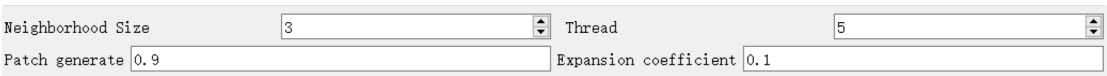


Fig. 19. The simulation parameters of PLUS model.

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 experimenter about study area. An example of cost matrix is provided 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.

And in the “**Neighborhood weights**” page, users need to set the neighborhood weights of different land use types, because the neighborhood effects may be different for different land use types in a study region. The value of the neighborhood weight for each land use type is determined according to expert knowledge and a series of model tests, ranging from 0-1. At this example, 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
Start Amounts	12742	169912	3005564	362518	132649	1506974	135893
Future Amounts 1	20133	43394	231113	2477678	929417	39413	1453586

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

3.4.7. Running the CARS

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

The visualization panel allowed the users to zoom in and observe the land use dynamics of local regions. The users can choose any land use type to show on the visualization panel by activating the corresponding radio buttons on the right. The horizontal axis of the dynamic chart is the number of iteration, and the vertical axis represents the product of the cover proportion and amount of cell.

The model will stop and save the simulation result in 2015 when the allocated area equal 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 loads the parameters that are recorded in the configuration files when click the Parameters button, which is convenient for the user to use to repeat their experiments.

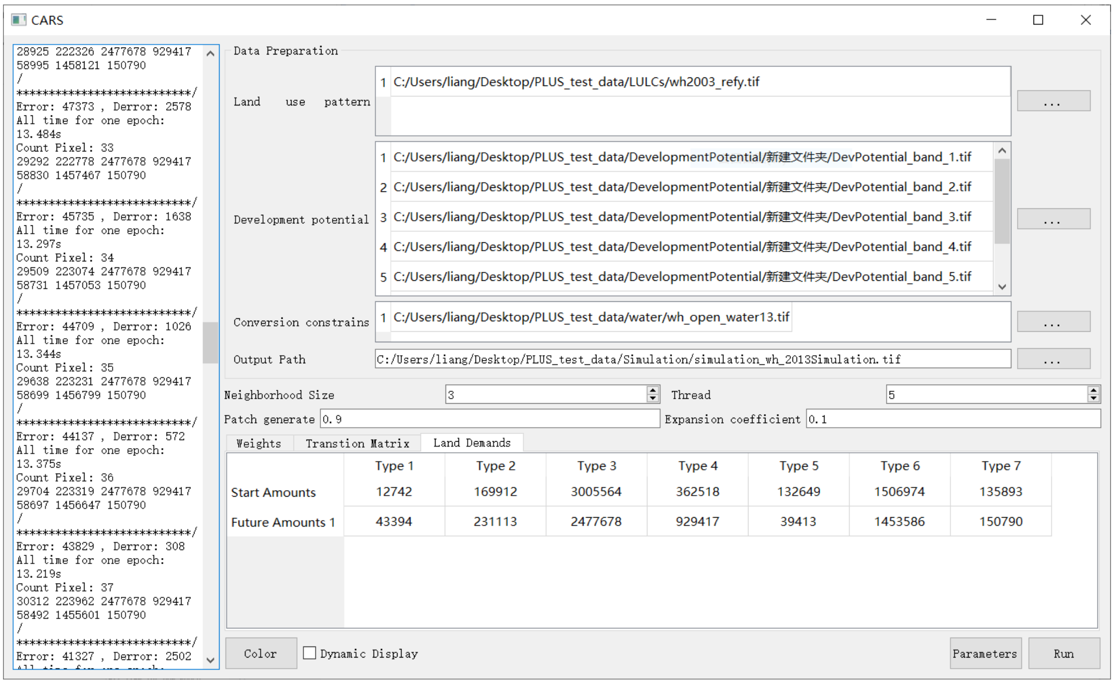


Fig. 22. The dynamic information of each iteration is show on left.

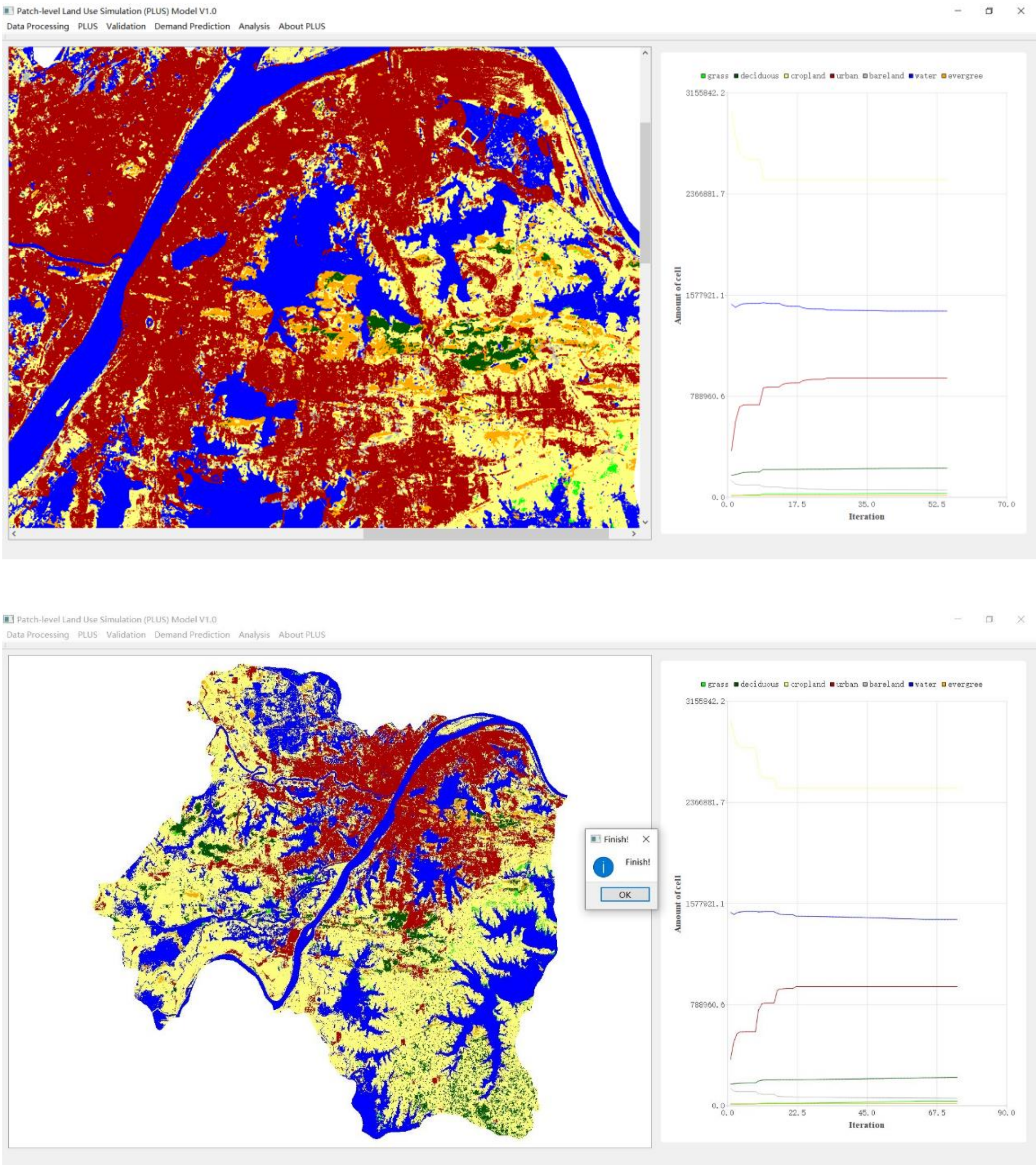


Fig. 23. The dynamic display of the simulation process on the main interface.

3.5. Accuracy validation of simulation

The PLUS software provides a kappa statistic tool and a FoM statistic tool for measuring agreement between simulation result and actual land use pattern.

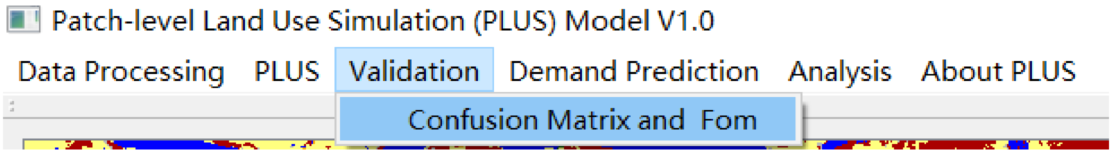


Fig. 24. Start the validation module.

3.5.1. Kappa statistic tool

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

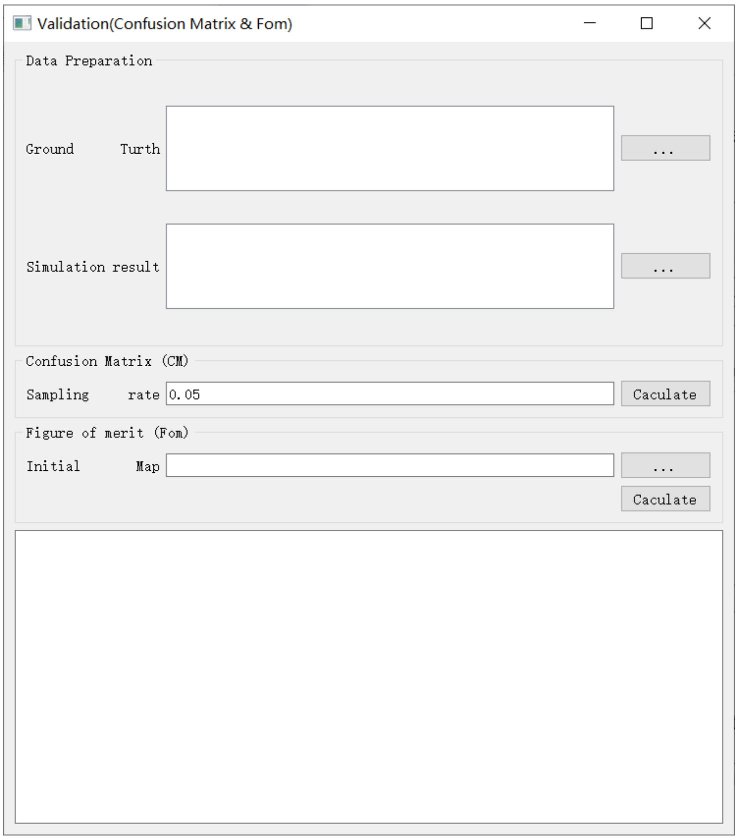


Fig. 25. UI of the validation module.

3.5.1.1. Load data and set the sampling rate

Click the  button in the "Ground Truth" row，to input the actual land use pattern named "wh2013_refy.tif" in 2013. Then click the  button in the "Simulation Result" row to input the simulated land use pattern. Then we set the sampling rate to 10%.

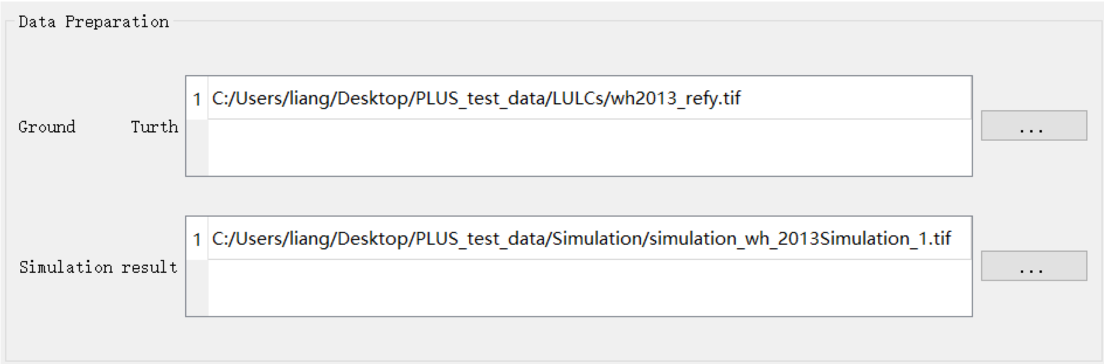
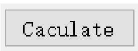


Fig. 26. Input the ground truth and simulation result.

3.5.1.2. Start calculating and save the result

Click the 

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

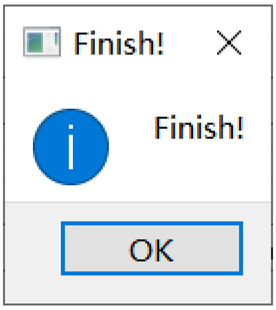


Fig. 27. The message box indicates that the calculation has finished.

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

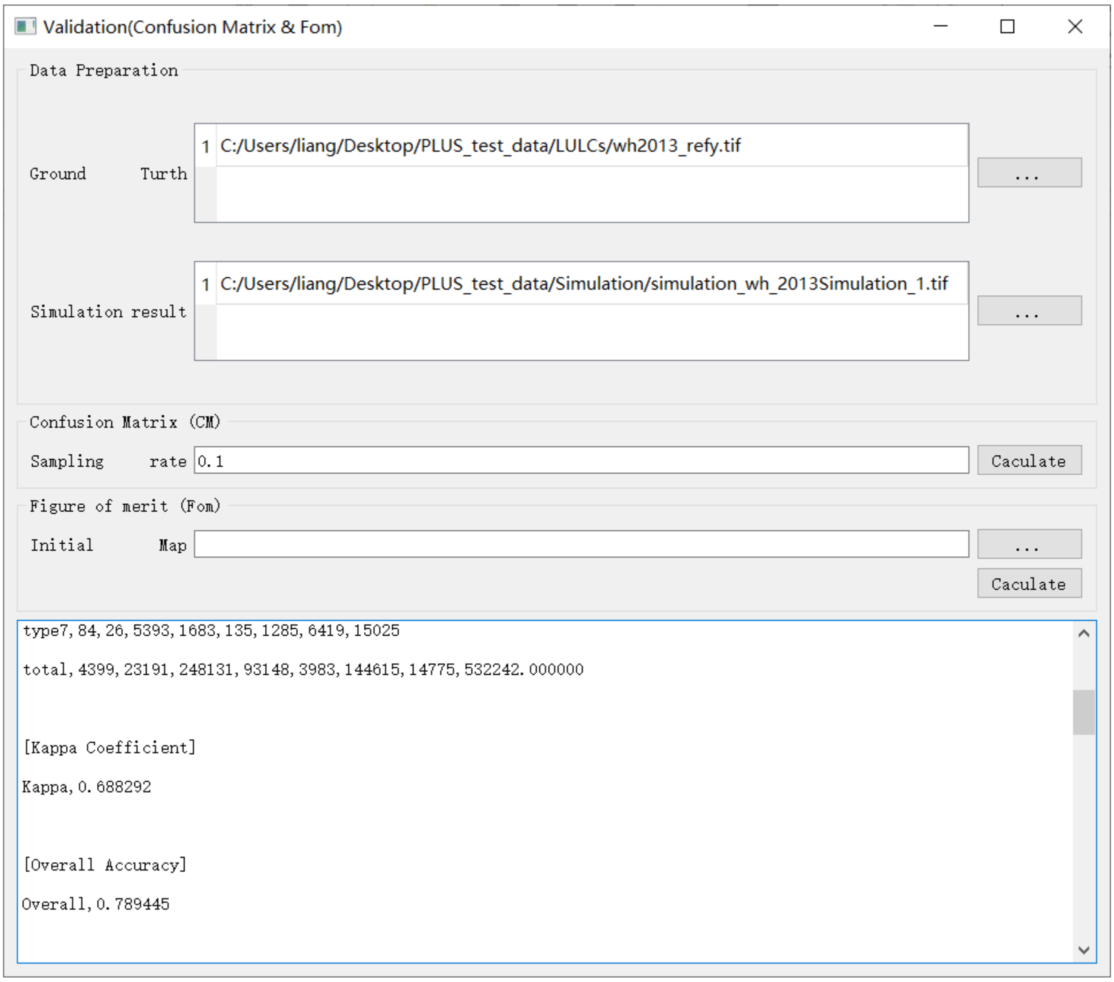
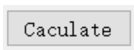


Fig. 28. Kappa coefficient and overall accuracy of the simulation result.

3.5.2. FoM statistic tool

Then click the  button in the "**Initial Map**" row, to input the start land use pattern named "wh2013_refy.tif" in 2003. Click the  button to start calculating FoM coefficient. A message box will pop up when the program completes the calculation.

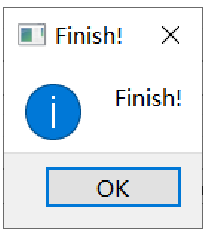


Fig. 29. The message box indicates that the calculation has finished.

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

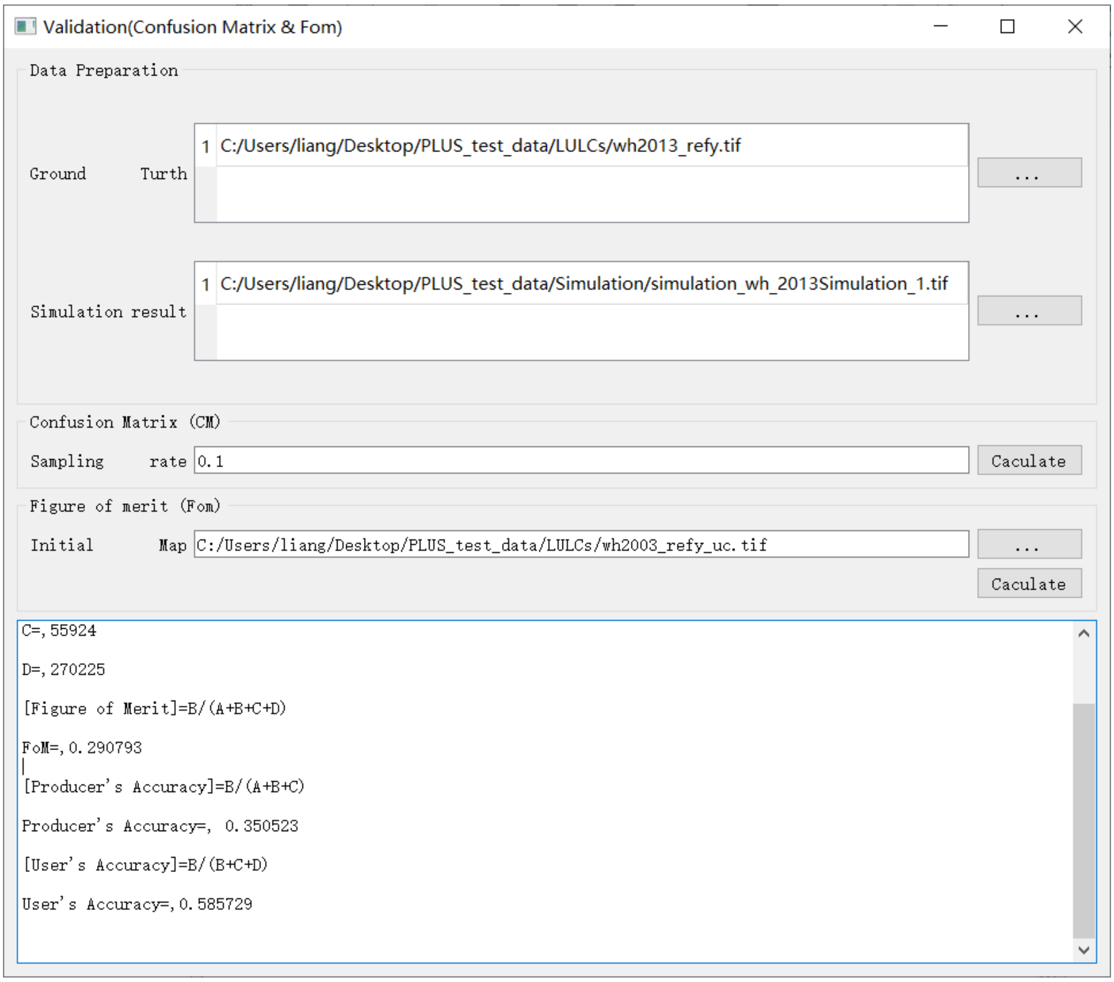


Fig. 29. Figure of merit of the simulation result.

3.6. Projecting future landuse demands

Future land use demands can be determined by many method, such as using expert experience, linear regression, Markov chains, the system dynamics model, or an integrated assessment model. This software provide linear regression method and Markov chain method. This experiment projects the future landuse demands based on the historical data.

3.6.1. Linear regression method

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

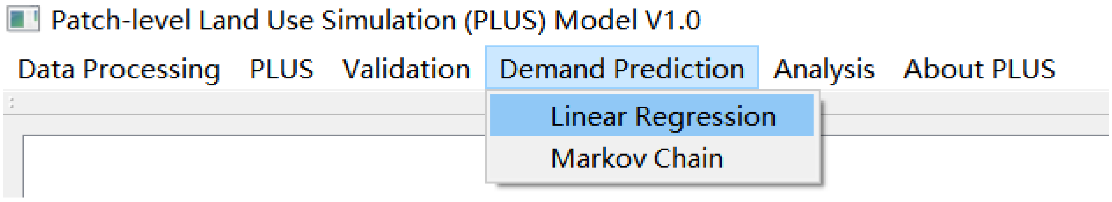


Fig. 30. Activating the land use demand projection module.

3.6.1.1. Input historical land use data

Click the  button, select the historical land use data in 2003, 2005, 2010 and 2013.

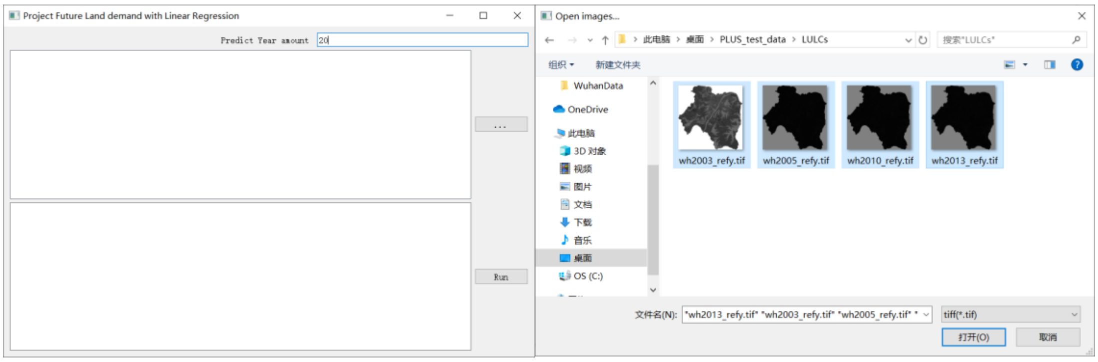


Fig. 31. Select the historical land use structure data.

What’s more, 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. Set the predict year amount as 20. Then the module will predict the future land use demand from 2013 to 2033 (2013+20).

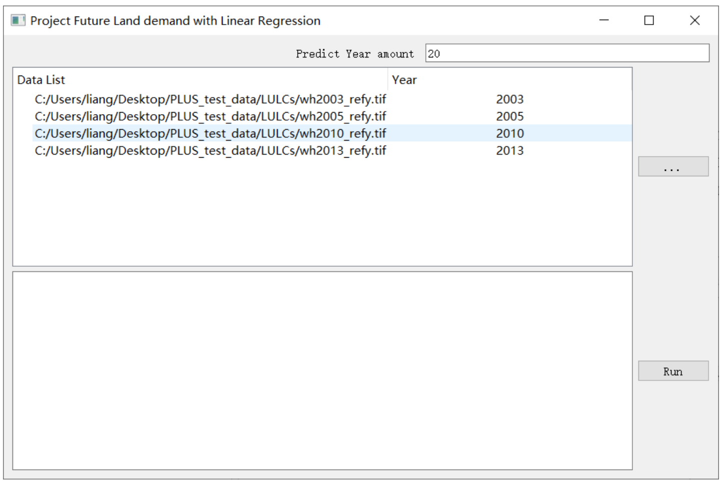
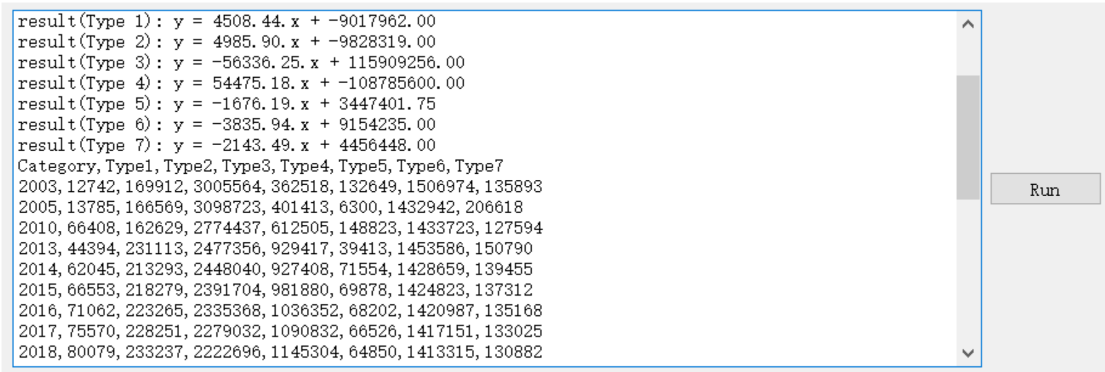


Fig. 32. Inputting the historical land use data to the Demand projection module.

3.6.1.2. 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 showed.



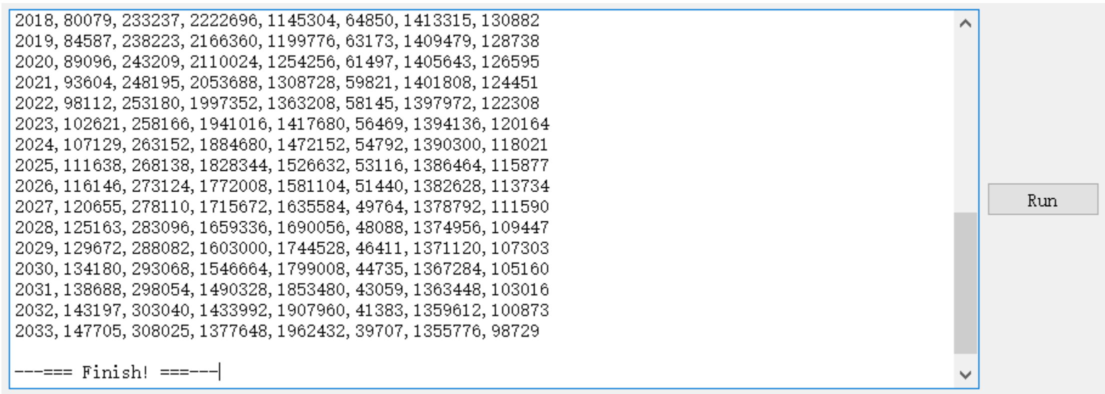


Fig. 33. Running the linear regression and predicting future land use demands.

3.6.2. Markov chain method

3.6.2.1. Input historical land use data

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

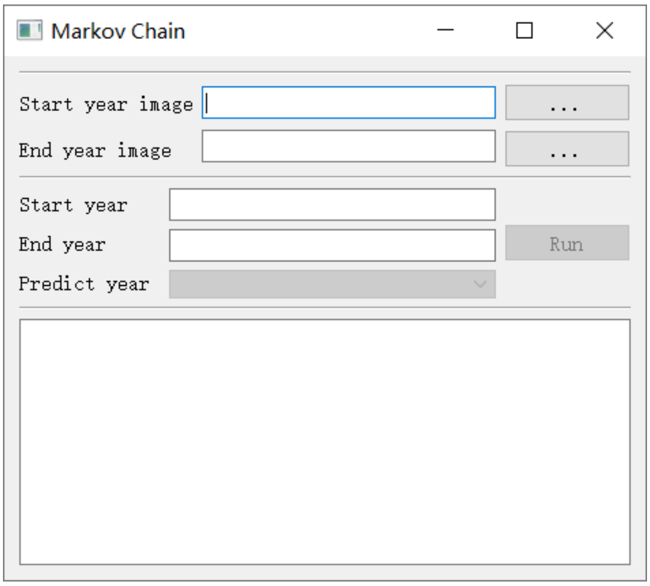


Fig. 34. The user interface for Markov chain.

Click the  button in the “**Start year image**” row to input the start year land use data (wh2003_refy.tif) in 2003. Then click the  button in the “**End year image**” row to input the start year land use data (wh2013_refy.tif) in 2013. Input the start year and end year for “Start year image” and “End year image”. Then click the text plain below and activated the “**Predict year**” pull-down menus. User can choose the future year they want to predict, for example 2033.

3.6.2.2. Predicting future land demands with Markov chian

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’.

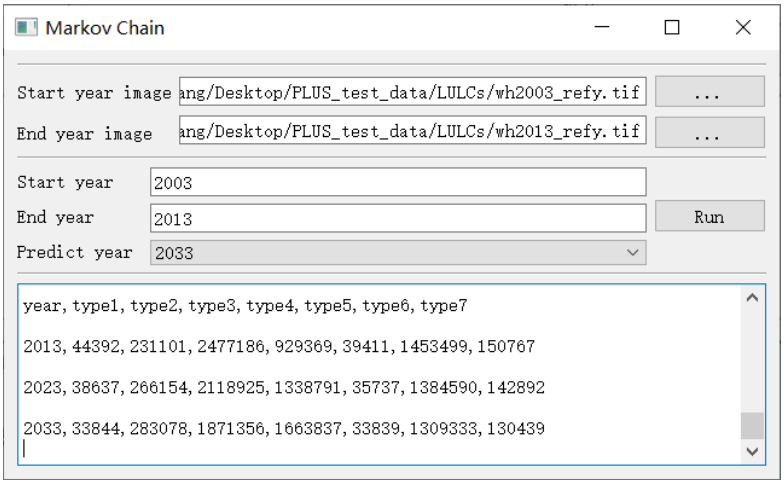


Fig. 35. Predict future land use amount with Markov chain.

3.6.3. Simulating future land use structure from 2013 to 2033

Click the  button in the “Land use pattern” text line and select the start year land use data “wh2013_refy.tif” in the pop-up dialog as the initial land use map. Then input the development potential file from 2003 to 2013 outputted by the ‘LEAS’ module in the “**Development Potential**” group box. Set 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) are keep the same as validation process (section 3.4.6). Then the users need to import the predicted land use demands in 2033. We analyze two future scenarios in this tutorial: 1) Linear regression scenario, and 2) Markov chain scenario. So the PLUS model should run twice with different land use demands and obtain two simulation results. Note that the final land use amounts will approximate but may not 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 “bottom-up” scale. The final land use amounts are final results of the interactions between “top-down” and “bottom-up” effects.

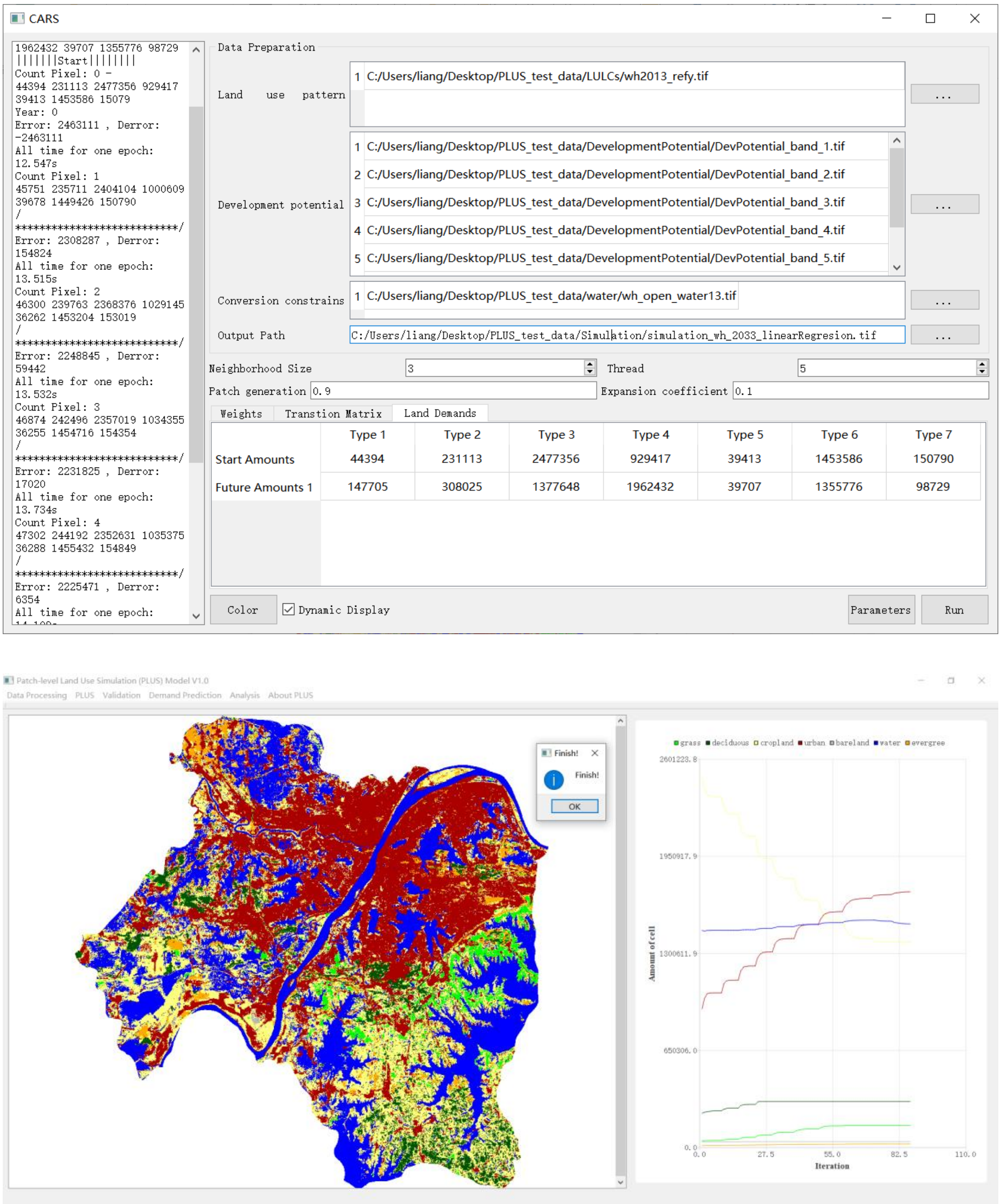


Fig. 36. Parameter setting and result for simulation from 2013 to 2033 under the Linear regression scenario.

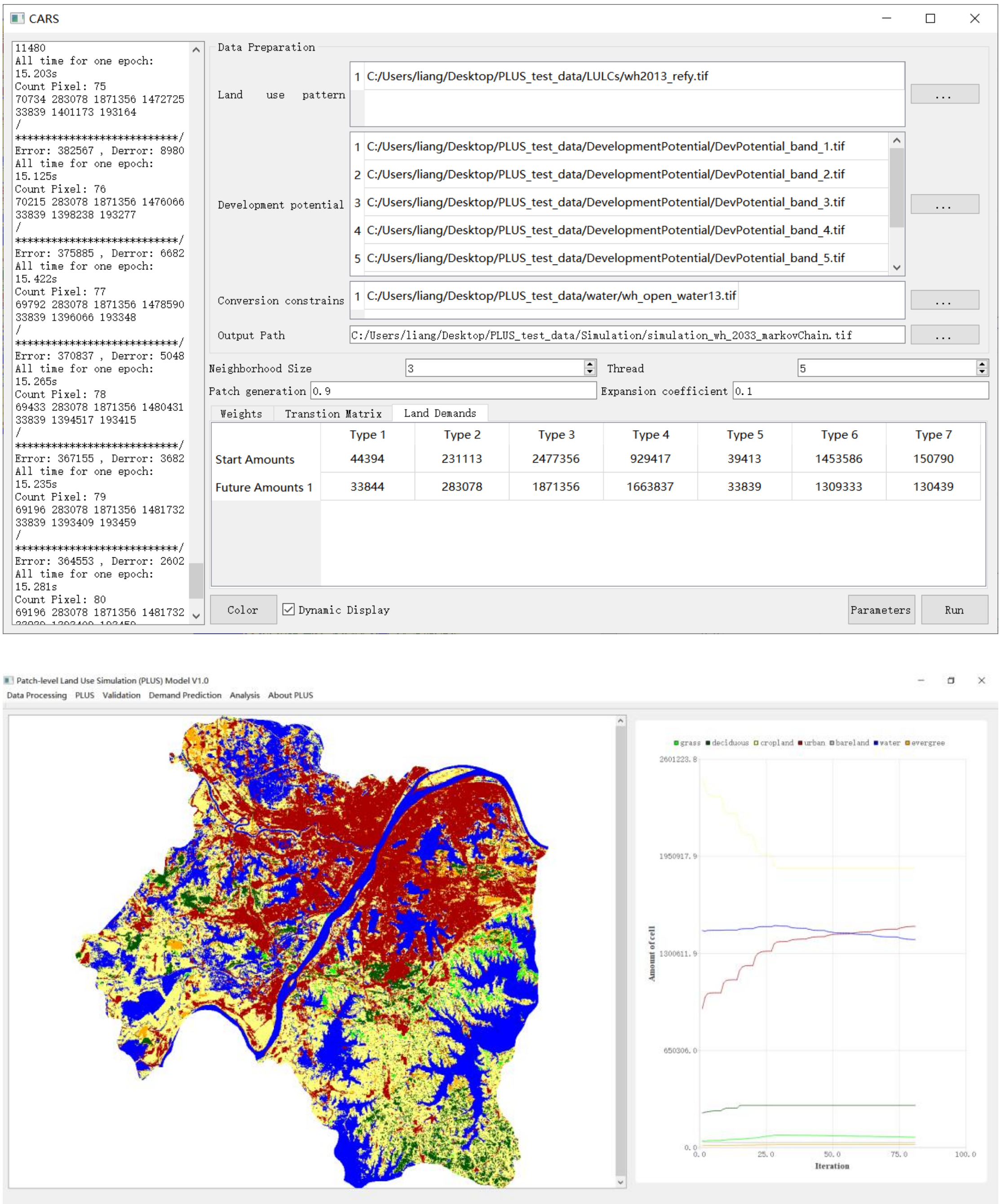


Fig. 37. Parameter setting and result for simulation from 2013 to 2033 under the Markov chain scenario.

The PLUS software will save all the parameters to two configuration files: “CARSparameters.tmp” in the ‘Parameterfile’ folder. The module will loads the parameters that are recorded in the configuration files when click the Parameters buttom, which is convenient for the user to use to repeat their experiments.

3.7. Calculation of scenario diversity

Scenario diversity shows pixels that are different between two or more scenarios. Areas that differ between scenarios are more sensitive to scenario or regional environment, while area with the same land use type regardless of scenario are less sensitive. Areas that differ between scenarios (3 scenario or more) also represent the hotspots of variability in study regions. The PLUS software provide a tool for calculating the scenario diversity.

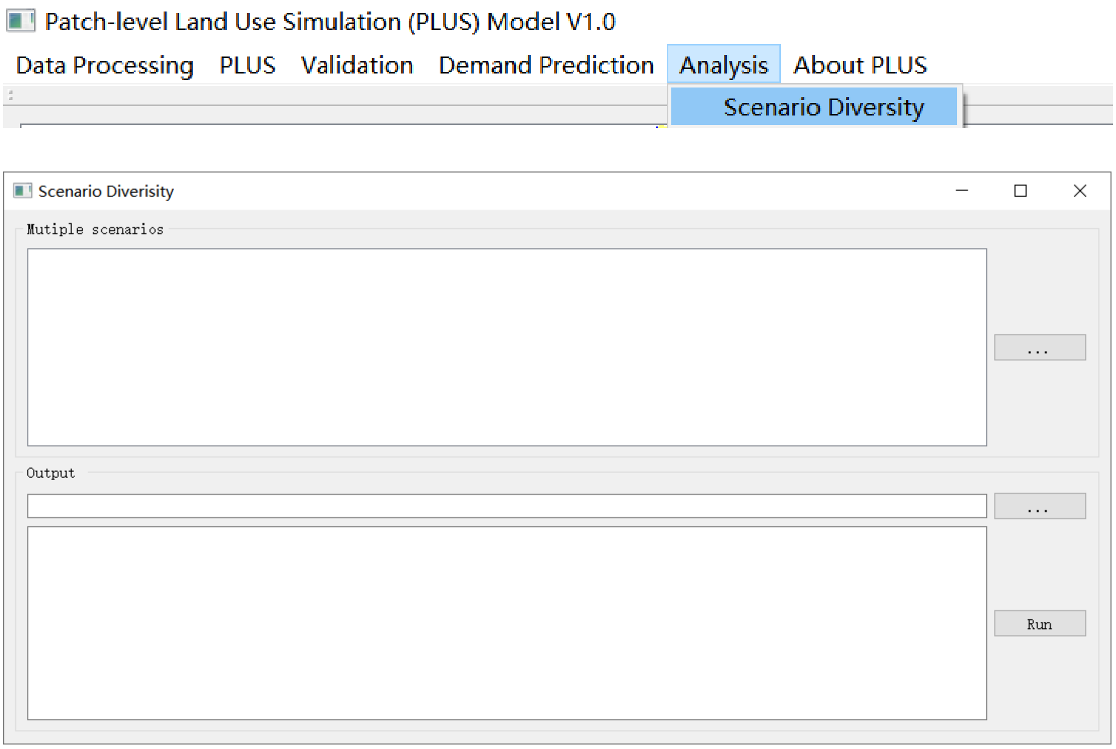


Fig. 38. Activating the ‘Scenario Diversity’ module.

3.7.1. 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.

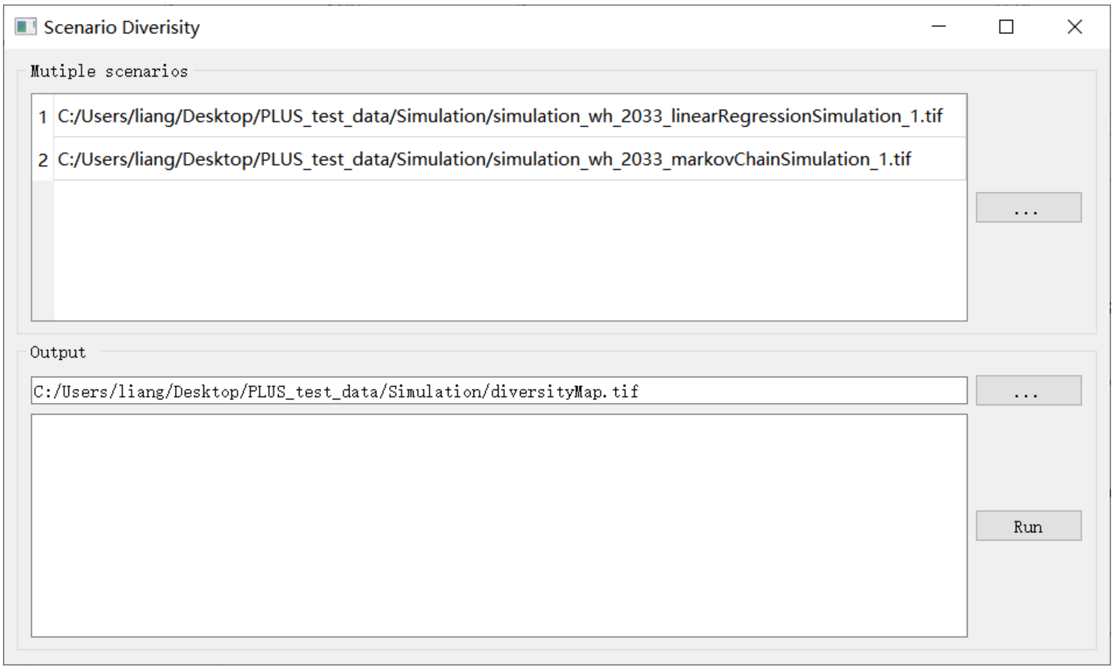
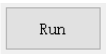


Fig. 39. Set the input and output paths for ‘Scenario Diversity’ module.

3.7.2. Calculating the land use mixture of each cell

Click the 

button to generate the scenario diversity map in 2033. Note that more scenario images can be imported to the ‘Scenario Diversity’ module. This tutorial only provide a simple case study.

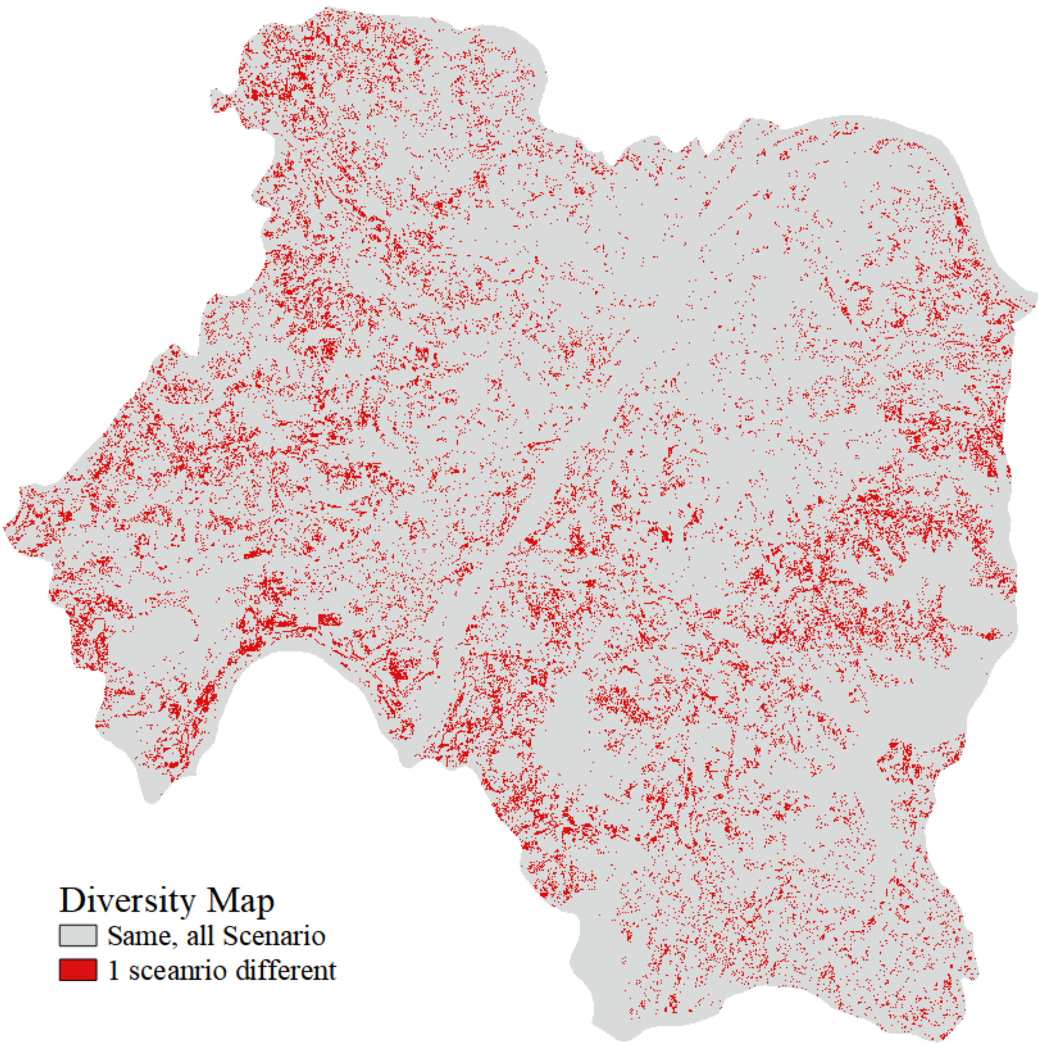


Fig. 40. Generating the scenario diversity map in 2033.

4. Some notes in using PLUS model

1) The input land use data must be in ‘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. But the coordinate or projection system of these images must be supported by GDAL. We recommend the World Geodetic System (WGS 84).

3) **In all the modules that involve two or more land use maps**, 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.

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