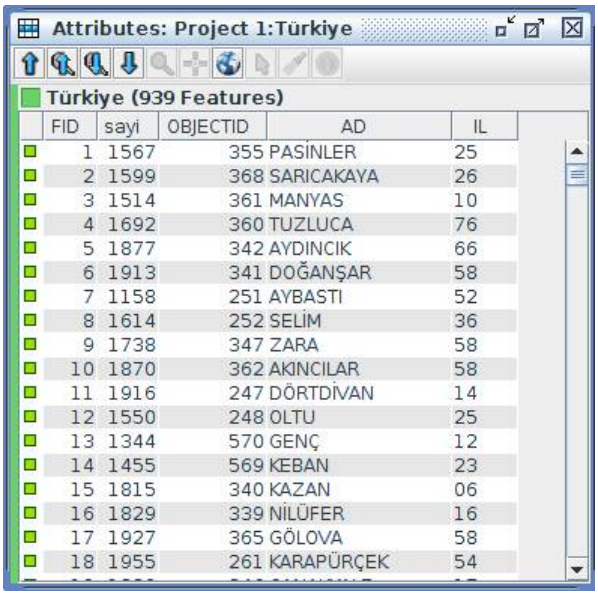
A Simple User's Guide for Functional Regionalisation Plugin

Burak Beyhan

1st Version
October 15, 2019

A Simple User's Guide for Functional Regionalisation Plugin (version 1.0.0)

Functional Regionalisation plugin is designed to reveal functional and planning regions on the base of flow databases. It is based on FRGIS algorithm developed by Beyhan (2019)[1]. In order to run the plugin, you should have two sets of databases: (1) a spatial database showing the basic spatial units (BSU) and (2) a flow database showing the magnitude and direction of linkages between BSUs. BSUs in these two databases will be interlinked to each other via an identity field. Thus, in the attribute table of the spatial database there should be a field showing a unique identity number for each BSU. In the flow database, the linkages between BSUs should also be organized by using the same set of identity numbers. Figure 1 shows a screenshot for a sample attribute table from OpenJUMP (in this table "sayi" shows the identity number for each BSU);



FID	sayi	OBJECTID	AD	IL
1	1567	355	PASİNLER	25
2	1599	368	SARICAKAYA	26
3	1514	361	MANYAS	10
4	1692	360	TUZLUCA	76
5	1877	342	AYDINCIK	66
6	1913	341	DOĞANŞAR	58
7	1158	251	AYBASTI	52
8	1614	252	SELİM	36
9	1738	347	ZARA	58
10	1870	362	AKINCILAR	58
11	1916	247	DÖRTDİVAN	14
12	1550	248	OLTU	25
13	1344	570	GENÇ	12
14	1455	569	KEBAN	23
15	1815	340	KAZAN	06
16	1829	339	NİLÜFER	16
17	1927	365	GÖLOVA	58
18	1955	261	KARAPÜRÇEK	54

Figure 1. A screenshot from the attribute table of the districts (NUTS4) in Turkey.

The most appropriate flow database that can be used in the delimitation of functional and planning regions is the commuting data. Nevertheless, other kinds of flow data can also be used for the delimitation of functional regions (such as migration data and commodity flows). In FRGIS, as explained in Beyhan (2019), it is assumed that flow database is organised in an edge data format showing the labels for origin BSU (OB) and destination BSU (DB) together with the shortest distance (SD) and the amount of commuting (AC) between them, and lastly the commuting level (CL) measured as the share of those commuting from OB to DB as a percentage of the total commuters residing in OB. In order to arrange the priorities assigned to each parameter that will be taken into consideration in the creation of functional regions (FR), the rows of the flow database should be ranked according to firstly CL in descending order, secondly SD in ascending order, and thirdly AC in descending order. For the set of formulas and algorithm used in the calculation of FR see Beyhan (2019).

[1] BEYHAN, B. (2019): The delimitation of planning regions on the basis of functional regions: An algorithm and its implementation in Turkey. Moravian Geographical Reports, 27(1): 15-30.

For the installation of the plugin, you should include FRGIS.jar under “ext” folder of OpenJUMP installation. You will also need dans-dbf-lib-1.0.0-beta-10.jar in order to process dbf files. After copying the respective files to “ext” folder, OpenJUMP can be initiated in order to run the plugin. The screenshots for FRGIS menu item (Figure 2) in the menu bar and tool item (Figure 3) in the toolbar can be seen below;

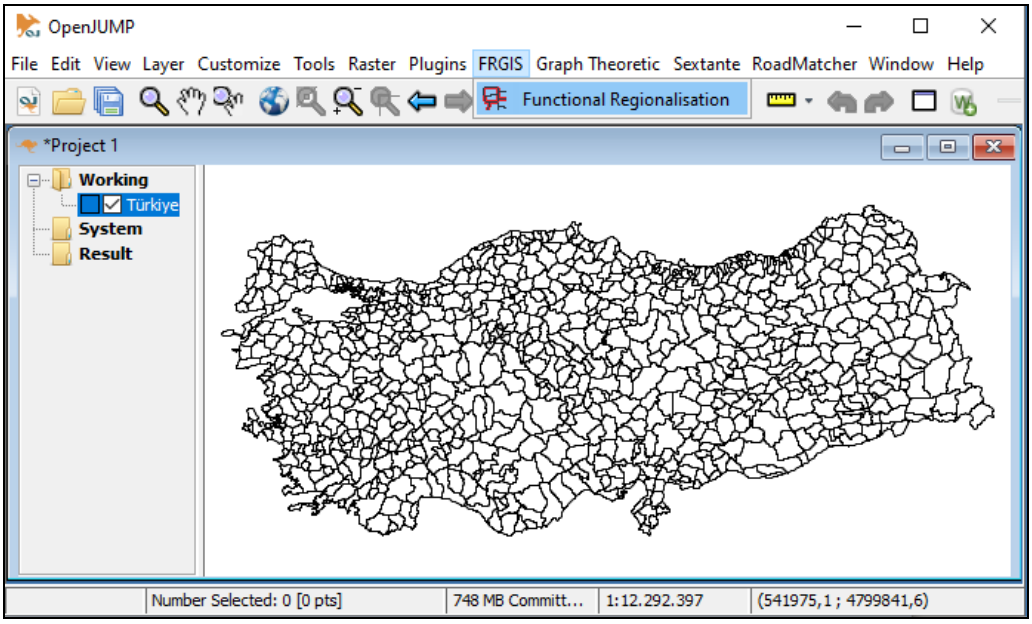


Figure 2. Functional Regionalisation plugin under FRGIS menu.

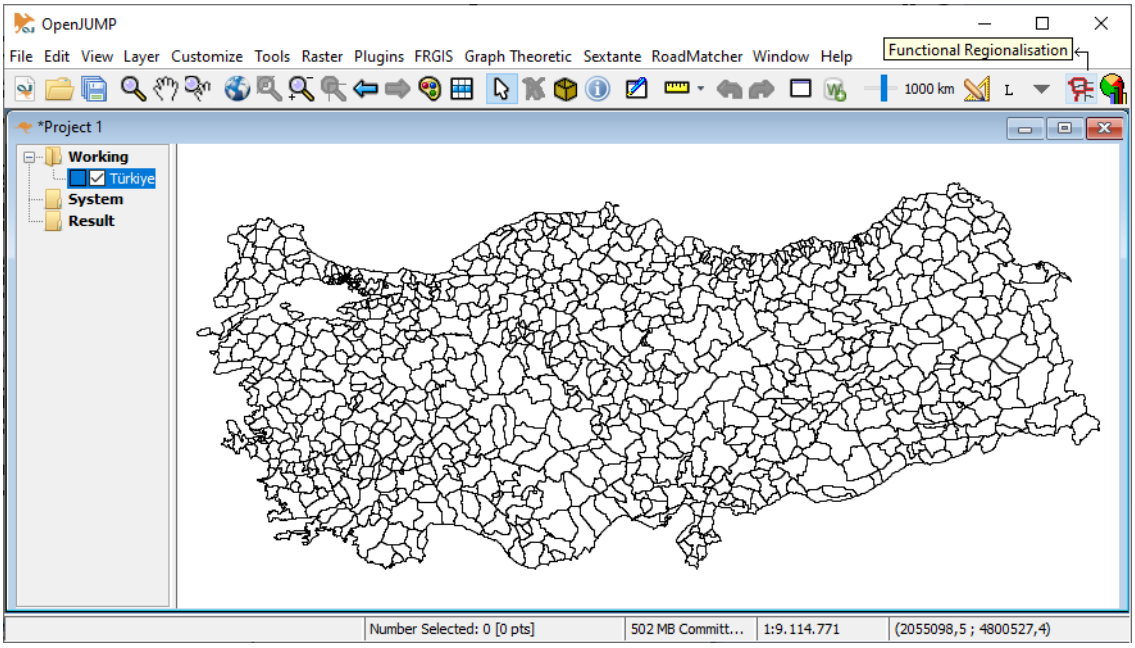


Figure 3. Functional Regionalisation plugin tool on the toolbar.

In order to activate the functional regionalisation plugin, you should either select FRGIS menu heading and then select the plugin item from the dropdown menu, or directly press FRGIS tool in the toolbar (- the symbol for the tool in the menu and toolbar). After activating the plugin following graphical user interface (GUI) (Figure 4) will be shown for the specification of; (1) the spatial and flow databases; (2) the fields for BSUs and parameters; (3) the thresholds for the parameters that will be taken into account in the delimitation of FRs; and (4) the FR options:

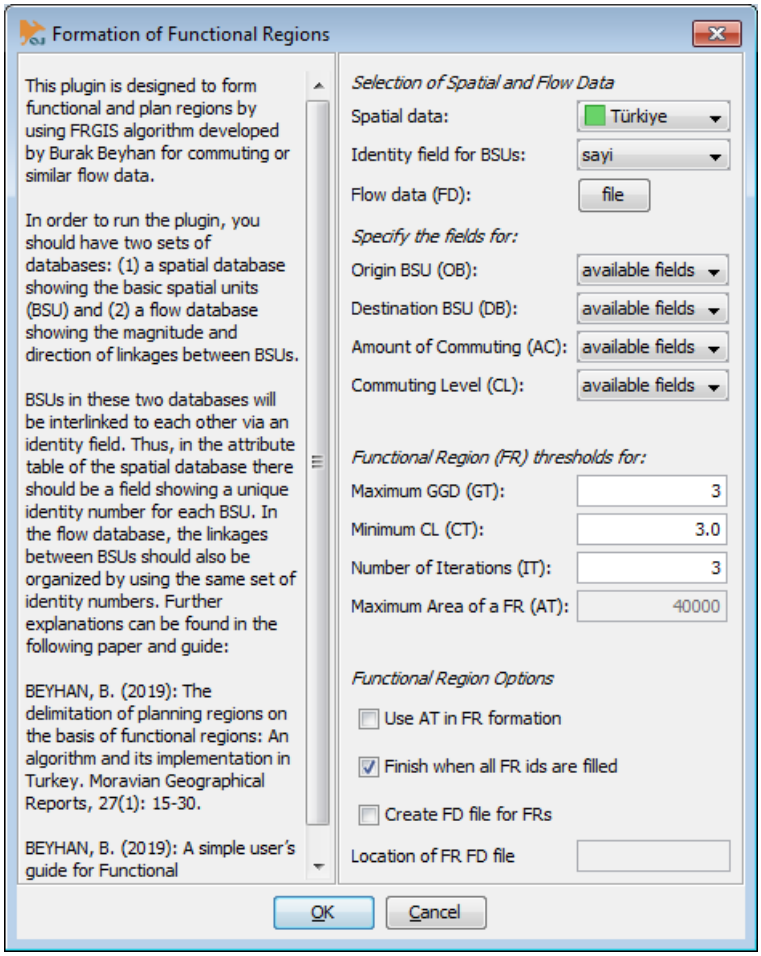


Figure 4. Graphical User Interface of Functional Regionalisation plugin.

For the specification of the spatial and flow databases (see "*Selection of Spatial and Flow Data*" section of GUI in Figure 5), the current selection in the layer panel of OpenJUMP will be automatically shown as the spatial database for the analysis. However, you can select another layer by using the dropdown menu in the dialog box created in the GUI of the plugin. Some editing operations can be required in polygon vector layer covering BSUs before running the plugin. Before the calculation of graph theoretical geodesic distance, physically disconnected BSUs (such as islands) separated from the rest of other BSUs because of the sea or straits should be connected to each other via narrow strips in order to prevent the exclusion of the BSUs concerned in the calculation of the geodesics.

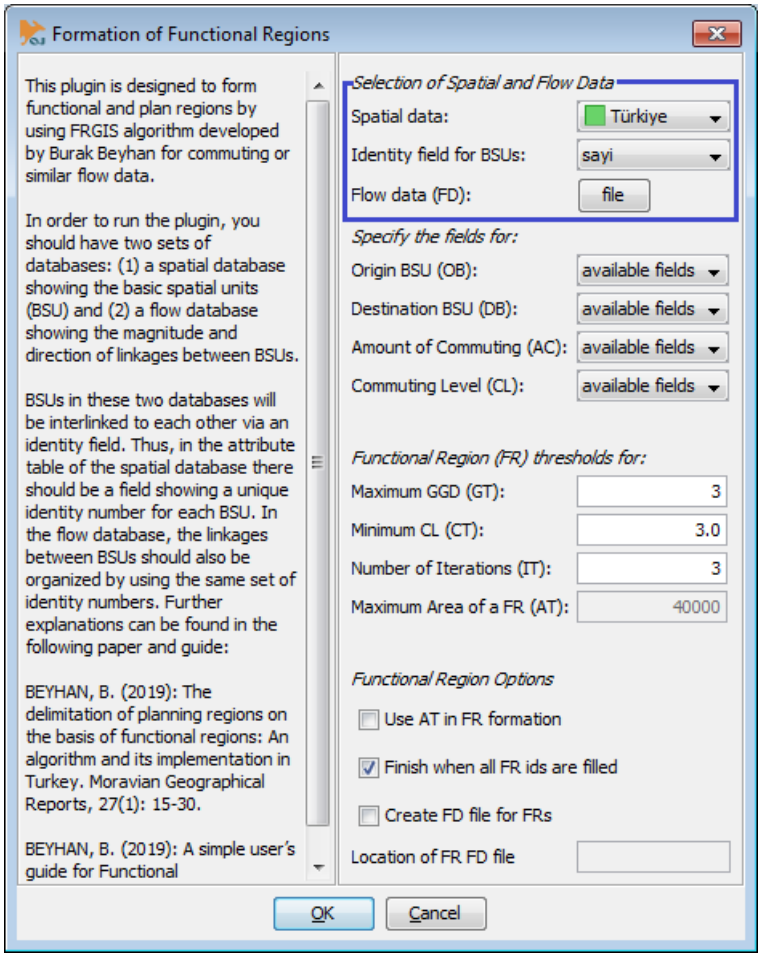


Figure 5. Specification of the spatial and flow databases (enclosed in a blue rectangle).

In order to define flow data, you should press *file* button that will active a file browser for you to select the file concerned (Figure 6):

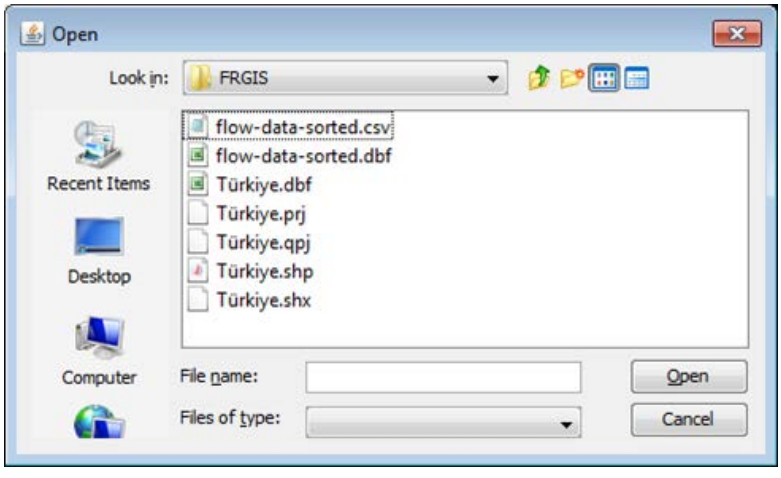


Figure 6. File browser activated to define the location of flow data file.

Two type of tabular data can be processed by the plugin: (1) character separated values (csv) and (2) database format (dbf). Once a flow file is chosen, the fields required to be defined under "*Specify the fields for:*" section of GUI will be populated according to the column headings available in the chosen csv (text) or dbf file. The chosen format of flow data file will also be treated as the template format for FR flow data file that can be created after formation of functional regions based upon the options defined by the user.

As explained above, before using the flow data in the analysis, you should organize the flow data in line with the prerequisites of FRGIS algorithm implemented in Functional Regionalisation plugin. Accordingly, the rows of the database should be sorted according to firstly Commuting Level (CL) in descending order, secondly Shortest Distance (SD) (if available) in ascending order, and thirdly Amount of Commuting (AC) in descending order (see Table 1 for an example).

Table 1. Part of a sample flow data showing the content in csv format and as a table.

Flow Data in csv format
OB;DB;AC;CL
1439;1439;20995;89.91435
1634;1634;17610;89.88363
1704;1704;33076;89.84625
…
1139;1844;265;15.61579
1819;1203;10201;15.60812
1303;2037;478;15.55483
1459;1318;77;15.49296
…

Flow Data			
OB	DB	AC	CL
1439	1439	20995	89.91435
1634	1634	17610	89.88363
1704	1704	33076	89.84625
…	…	…	…
1139	1844	265	15.61579
1819	1203	10201	15.60812
1303	2037	478	15.55483
1459	1318	77	15.49296
…	…	…	…

If the flow data is saved as a text file in csv format, user is prompted to define the separator used in the text in order to delimit the values after selection of the file concerned (Figure 7). A dropdown menu is available for the selection of the most widely used separators (such as semicolon [;], comma [,], tab, space and dash [-] characters). However, you may also directly enter the character used as a separator in the text.

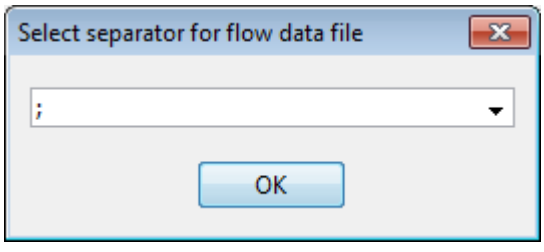


Figure 7. Selection of the character used as a separator in the text file.

After selecting the flow data file and, if necessary, defining the separator, you can define the fields for Origin BSU (OB), Destination BSU (DB), AC and CL by selecting the fields available in the flow data. For this purpose, in "*Specify the fields for:*" section of GUI, a dropdown menu is provided for each field (Figure 8).

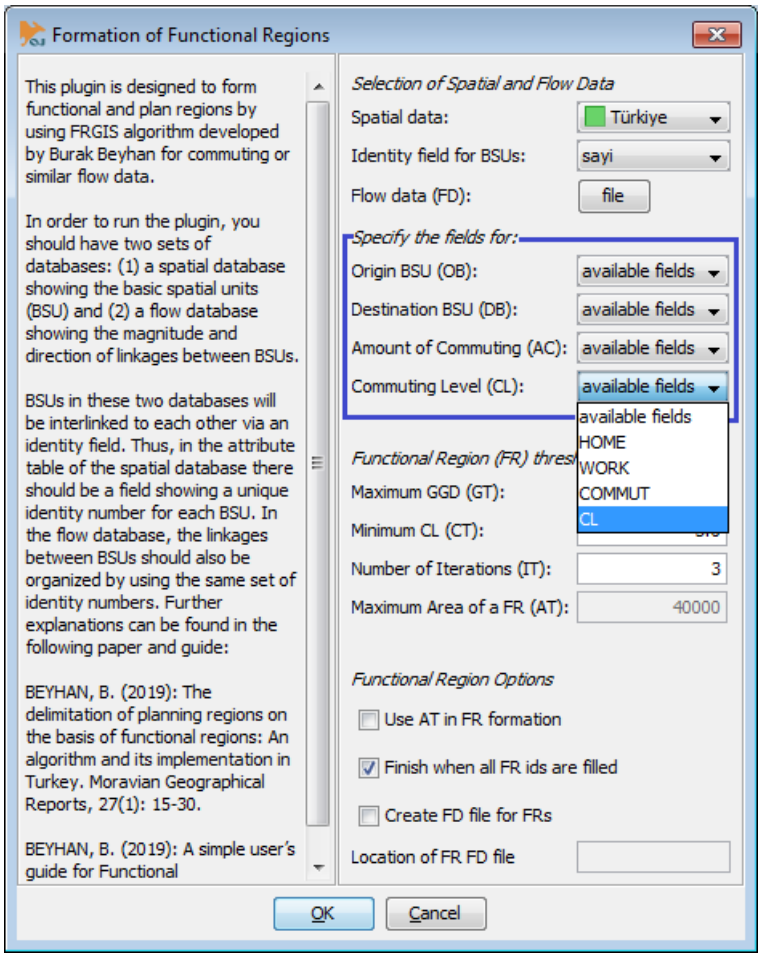


Figure 8. Selection of the fields for OB, DB, AC and CL in the GUI.

After defining the fields for OB, DB, AC and CL, threshold values for the parameters used in FRGIS should be defined by typing these values in the fields provided in the GUI (see "*Functional Region (FR) thresholds for:*" section of GUI in Figure 9). Further explanations for these parameters can be found in Beyhan (2019). What follows is a list of short descriptions for these parameters:

- GT is the threshold value for maximum GGD (Graph Theoretical Geodesic Distance) that will be allowed in the analysis for the fusion of BSUs in order to form FRs. By default, GT is set to 3, and needs to be at least 1 or above.

- CT is the minimum CL that will be considered in order to unite OB with DB. Any couple of OB and DB having a CL value below CT will not be taken into consideration in the formation of FRs if all of them can be assigned to a FR. CT is set to 3 by default.
- IT is the maximum number of iterations defined for the prevention of multi-polygon regions (MPR). IT is set to 3 by default. If necessary, it can be increased to higher values.
- AT is the threshold value (in km²) for the maximum area that can be occupied by a FR. Actually, AT is not required in the first phase of functional regionalisation process if the main intention is to designate the resulting FRs as PRs.

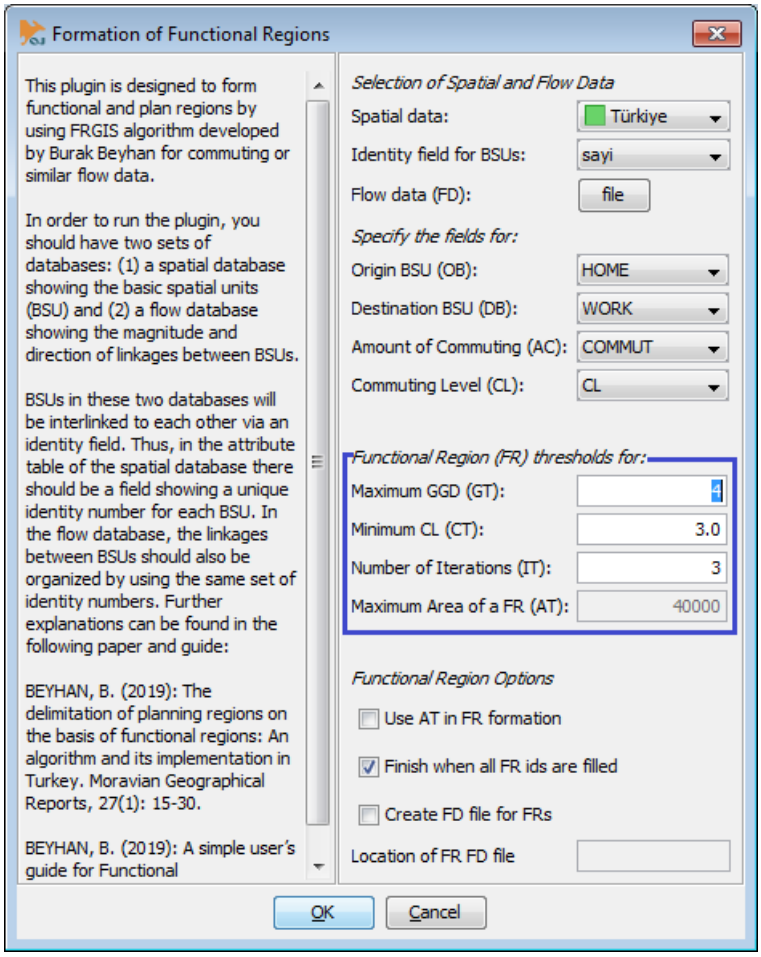


Figure 9. Entering values for GT, CT, IT and AT in the GUI.

There are also some functional region options in the GUI (see "*Functional Region Options*" section of GUI in Figure 10). The first option ("Use AT in FR formation") is not selected by default. It is assumed that area threshold (AT) (unit of measurement for this parameter is km²) will be used in the second phase of functional regionalisation process. In the second phase, formation of a FR having an area above AT is not allowed in order to prevent the formation of overly large FRs. The second option ("Finish when all FR ids are filled") is selected by default. Once this option is selected, FRGIS finishes the process of formation of FRs if there remains no

BSU without a FR id. If not, FRGIS continues to form FRs according to the predefined values for the parameters taken into consideration in the formation of FRs. The last option is required to specify whether or not a new flow data file will be created for the FRs formed after the analysis. Once this option is selected a file browser will be activated in order to define the name and location of the file concerned (you should type a file name to be created).

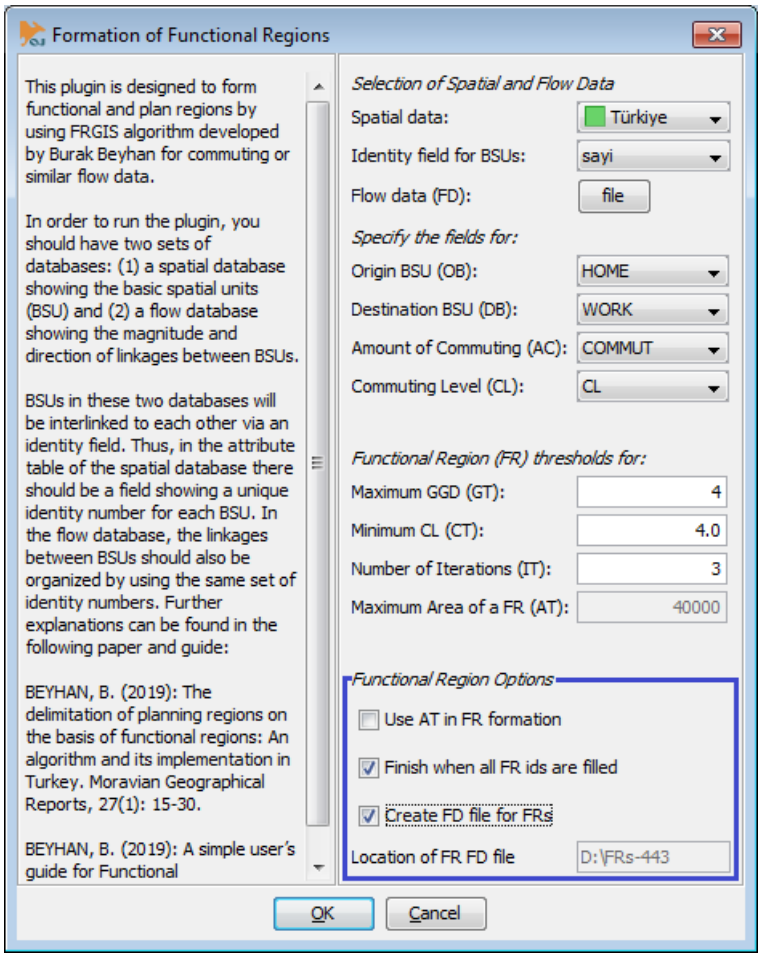


Figure 10. Selecting the options for functional region formation.

Once all the required layer, files, fields, parameter thresholds and options are defined, you can run the plugin. The current process performed by the plugin in the implementation of FRGIS and the progress of the process concerned are shown in the task monitor dialog box. Below are some sequential screenshots from the task monitor dialog box (Figure 11, 12, 13, 14, and 15):

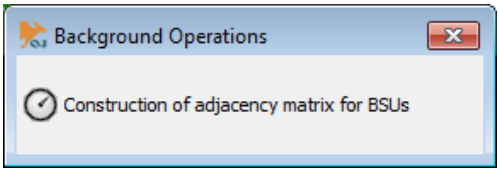


Figure 11. The process of construction of adjacency matrix (background operations).

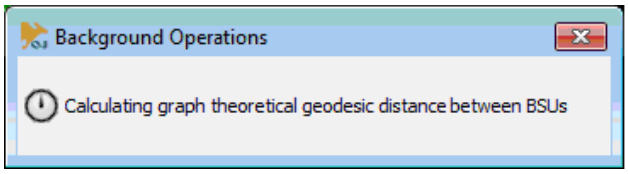


Figure 12. The process of calculation of graph theoretical geodesic distance between BSUs.

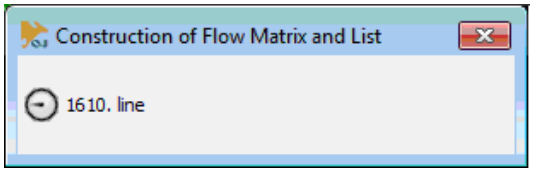


Figure 13. The process of construction of flow matrix and list used in the calculations.

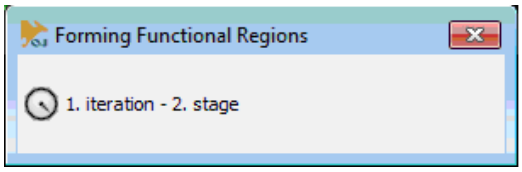


Figure 14. The current round of iteration and stage.

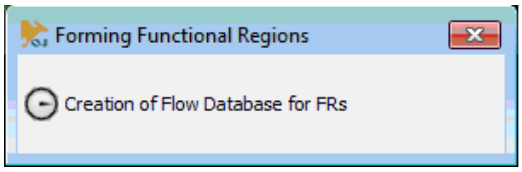


Figure 15. The process of creation of flow database for the resulting functional regions.

After the formation of FRs and, if selected, creation of flow data, a dialog box (Figure 16) informing the user about the results of the functional regionalisation analysis is shown. Accordingly, statistics for the total number of FRs formed, minimum CL (MCL) taken into consideration within the given limits of thresholds, the number of rows read to satisfy CT (last line read - Lline), MCL for those OB and DB that cannot be assigned to a FR within the given limits of CT (MCL without threshold – MCL_out), the number of rows read in order to assign those BSUs involved in the couples of OB and DB having a CL value below CT to a FR (last line read for minimum CL without threshold – Lline_out), the number of multi-polygon regions (MPR), the number of polygons in MPRs (total number of disconnected parts (components) in MPRs), the area of the largest FR, the area of the smallest FR, the area of the median FR, the average area of a FR, the maximum level of self-containment (LSC) observed for a FR, the minimum LSC observed for a FR, the median LSC observed for a FR, and the average LSC observed for a FR are reported in the dialog box.

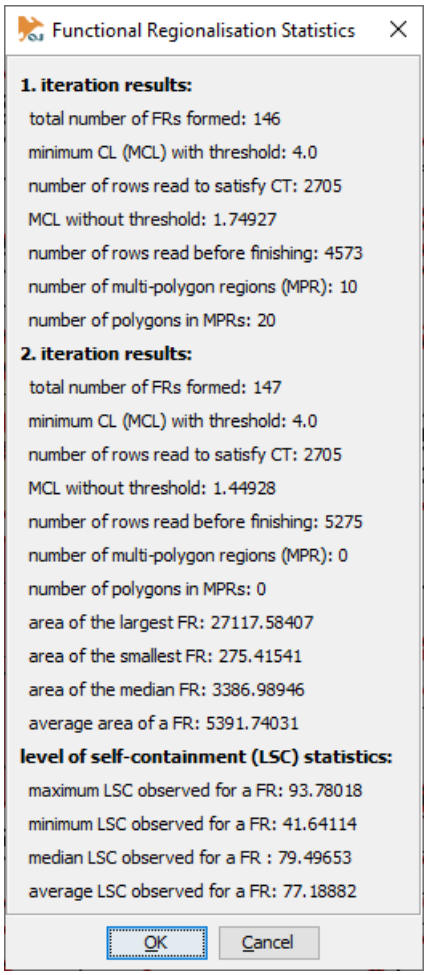


Figure 16. A sample dialog box showing the statistics for functional regionalization process.

In addition to this dialog box, after each run, the result of the analysis is also logged to a file (stats.csv) in the folder where the flow data is located. In the respective file, all the statistics shown in Figure 16 are registered to stats.csv together with the threshold values of the parameters used in the functional regionalisation process. After finishing your analysis by running the plugin for different configuration of the values defined for the parameters used in the analysis, you can open stats.csv (Table 2) in a spread sheet software program and analyze the results in order to select the best run of the plugin for the designation of final FRs.

Table 2. A sample table showing the content of the log file (stats.csv).

GT	CT	NI	AT	FRs	MCL	Lline	MCL_out	Lline_out	MFRs	Parts	max_area	min_area	med_area	mean_area	max_LSC	min_LSC	med_LSC	mean_LSC
3	3	1	-	147	3.002	3174	0.51107	13037	9	18	27609.94	27.44	3386.99	5391.74	-	-	-	-
3	3	2	-	146	3.002	3174	0.51107	13037	0	0	27117.58	27.44	3388.59	5428.67	93.79	24.36	79.50	76.99
4	4	1	-	146	4	2705	1.74927	4573	10	20	27609.94	287.94	3422.36	5428.67	-	-	-	-
4	4	2	-	147	4	2705	1.44928	5275	0	0	27117.58	275.42	3386.99	5391.74	93.78	41.64	79.50	77.19
1	0	1	40000	26	0.1729	2546	--	0	0	0	39976.60	7365.58	35709.76	30484.07	96.89	68.91	87.81	86.16

A detailed explanation is available in Beyhan (2019) for the column headings given in the above table. As explained before, GT stands for the threshold value for maximum GGD, CT stands for the minimum CL that will be taken into account in order to unite OB with DB, and AT stands for the threshold value (in km²) for the maximum area that can be occupied by a FR. NI is the number of iteration; FRs is the number of FRs formed; MCL is the minimum commuting level used in the formation of FRs; MFRs is the number of MPRs; Parts are the number of components in MPRs concerned. Column headings ending with 'area' and 'LSC' (level of self-containment) show statistics regarding these attributes of the FRs formed. Accordingly, column headings having a prefix of max (maximum), min (minimum), med (median), or mean (average) show the observed statistics for these values. In addition to the columns given in this table, a series of statistics is also reported in additional columns for max_res (maximum number of people residing in a FR), min_res (minimum number of people residing in a FR), med_res (median number of people residing in a FR) mean_res (average number of people residing in a FR), max_work (maximum number of people working in a FR), min_work (minimum number of people working in a FR), med_work (median number of people working in a FR), mean_work (average number of people working in a FR).

When the plugin is run, initially, the fill colour for the polygon layer showing the boundaries of BSUs is made transparent, and the width and colour of the boundaries of BSU are also changed (to thinner black lines). After each run of the plugin, a series of layers are also created in relation to the resulting FRs and the iterations performed by the plugin during the functional regionalisation process (Figure 17). Accordingly, for each iteration user can see the functional regions and multi-polygon regions (MPRs) (with a suffix showing the number of iteration) separately. For the final iteration leading to the formation of FRs (Functional Regions) without any MPR, number of iterations is not added as a suffix to the end of the layer name. The result of this final iteration is presented in a layer labelled as "Functional Regions" whose boundaries are coloured in red and thicker than original layer (Figure 17). MPRs layers are coloured according to the attribute name showing the FR identity (FRID) number. This helps users see MPRs clearly.

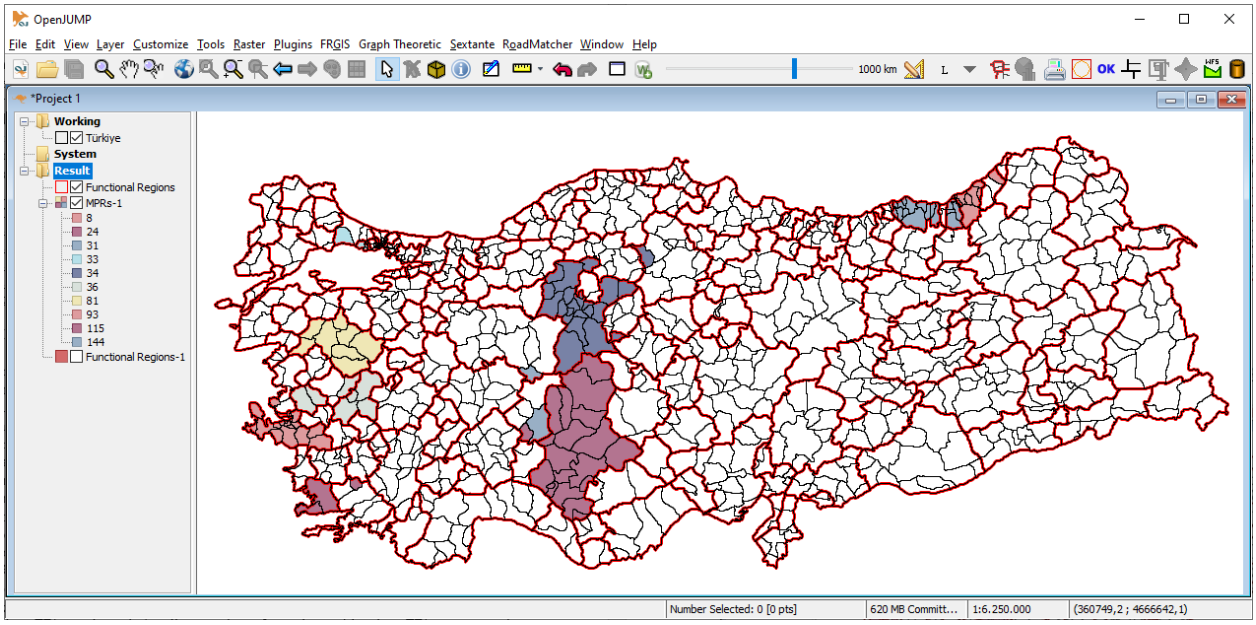
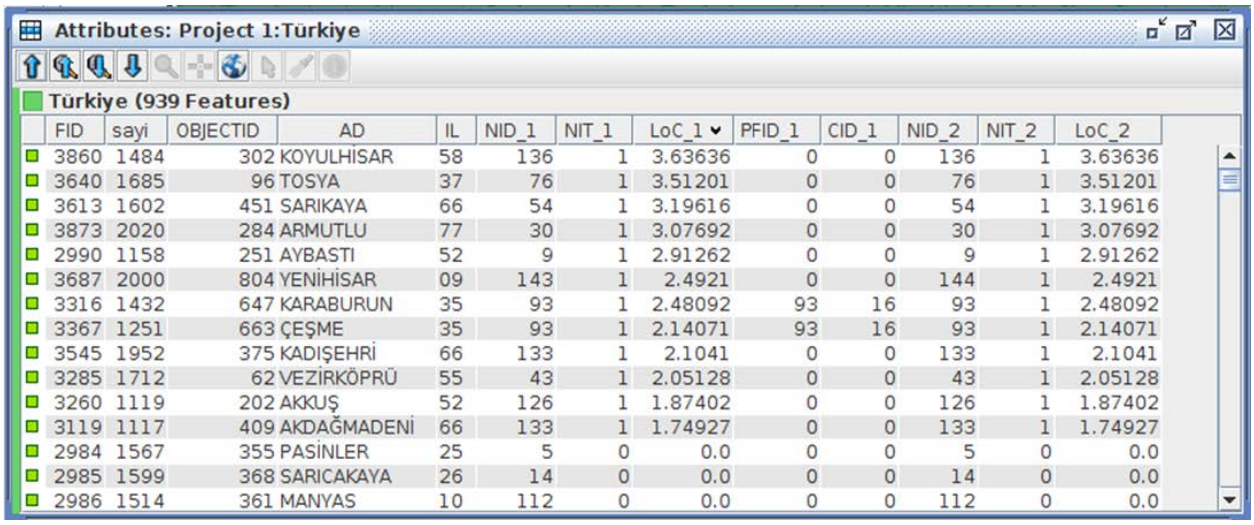


Figure 17. A sample screenshot from the results of analysis showing the layers created.

Some additional statistics (mainly for those BSUs that cannot be assigned to a FR within the given limits of thresholds or those BSU that are involved in MPRs) are also registered to the attribute tables of both original map showing the BSUs and maps of the resulting Functional Regions and MPRs. In the attribute table of the original map, you will see that following fields are added to the table with a suffix showing the number of iteration (Figure 18);

- *NID* stands for the new identity number showing the resulting FR that covers the BSU concerned,
- *NIT* stands for the round of iteration of commuting database when a FR id is assigned to the BSU having no FR id after the first stage (a value of 0 shows that a FR id is assigned to the respective BSU in the first stage within the given limits defined for GT and CT),
- *LoC* stands for the level of commuting observed between the BSU previously having no FR id and the BSU whose FR id is assigned to the one having no FR id (a value of 0 shows that the BSU concerned is involved within a FR within the given limits defined for GT and CT),
- *PFID* stands for the previous FR id assigned to those BSUs involved in an MPR in the previous iteration of FR formation (a value of 0 shows that the BSU concerned is not involved in any MPR),
- *CID* stands for the component id of the BSU involved in the MPR concerned. An MRP may involve several components (disconnected parts). Thus, CID shows the component encapsulating the BSU concerned (a value of 0 shows that the BSU concerned is not involved in any MPR).



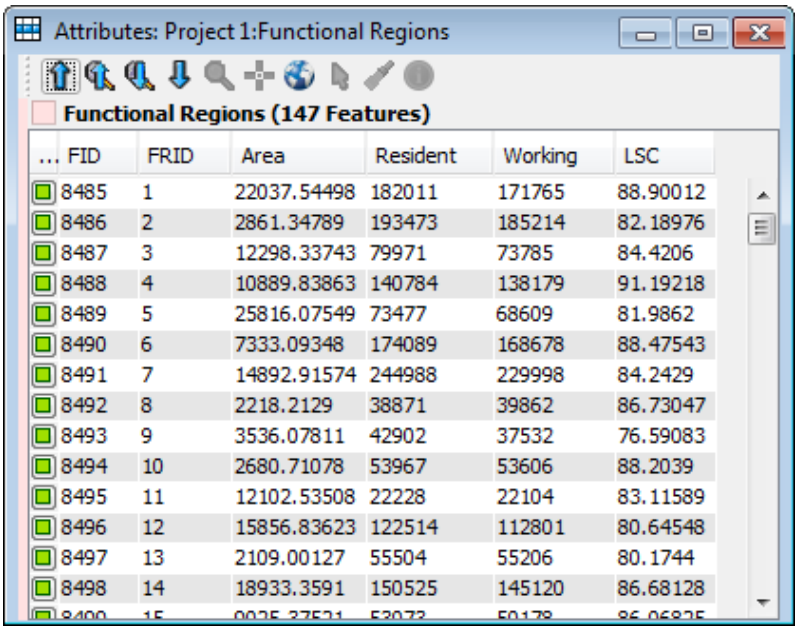
FID	sayi	OBJECTID	AD	IL	NID_1	NIT_1	LoC_1	PFID_1	CID_1	NID_2	NIT_2	LoC_2
3860	1484	302	KOYULHİSAR	58	136	1	3.63636	0	0	136	1	3.63636
3640	1685	96	TOSYA	37	76	1	3.51201	0	0	76	1	3.51201
3613	1602	451	SARIKAYA	66	54	1	3.19616	0	0	54	1	3.19616
3873	2020	284	ARMUTLU	77	30	1	3.07692	0	0	30	1	3.07692
2990	1158	251	AYBASTI	52	9	1	2.91262	0	0	9	1	2.91262
3687	2000	804	YENİHİSAR	09	143	1	2.4921	0	0	144	1	2.4921
3316	1432	647	KARABURUN	35	93	1	2.48092	93	16	93	1	2.48092
3367	1251	663	ÇEŞME	35	93	1	2.14071	93	16	93	1	2.14071
3545	1952	375	KADIŞEHRİ	66	133	1	2.1041	0	0	133	1	2.1041
3285	1712	62	VEZİRKÖPRÜ	55	43	1	2.05128	0	0	43	1	2.05128
3260	1119	202	AKKUŞ	52	126	1	1.87402	0	0	126	1	1.87402
3119	1117	409	AKDAĞMADENİ	66	133	1	1.74927	0	0	133	1	1.74927
2984	1567	355	PASİNLER	25	5	0	0.0	0	0	5	0	0.0
2985	1599	368	SARICAKAYA	26	14	0	0.0	0	0	14	0	0.0
2986	1514	361	MANYAS	10	112	0	0.0	0	0	112	0	0.0

Figure 18. A sample screenshot from the attribute table of original spatial database after analysis (attribute table is ranked according to LoC_2).

By analysing the values presented in the attribute table of the original map, the thresholds defined for GT and CT can be readjusted in order to get better results regarding higher scores of commuting level and prevention of MPRs. In this respect, further information can also be found

in the attribute table of resulting Functional Regions. The attribute table (Figure 19) of the resulting functional regions includes following fields;

- *FRID* stands for the identity number of the FR concerned,
- *Area* shows the area occupied by the resulting FR,
- *Resident* shows the total number of commuters residing in the FR concerned,
- *Working* shows the total number of commuters working in the FR concerned,
- *LSC* shows the level of self-containment for the FR concerned,



...	FID	FRID	Area	Resident	Working	LSC
	8485	1	22037.54498	182011	171765	88.90012
	8486	2	2861.34789	193473	185214	82.18976
	8487	3	12298.33743	79971	73785	84.4206
	8488	4	10889.83863	140784	138179	91.19218
	8489	5	25816.07549	73477	68609	81.9862
	8490	6	7333.09348	174089	168678	88.47543
	8491	7	14892.91574	244988	229998	84.2429
	8492	8	2218.2129	38871	39862	86.73047
	8493	9	3536.07811	42902	37532	76.59083
	8494	10	2680.71078	53967	53606	88.2039
	8495	11	12102.53508	22228	22104	83.11589
	8496	12	15856.83623	122514	112801	80.64548
	8497	13	2109.00127	55504	55206	80.1744
	8498	14	18933.3591	150525	145120	86.68128

Figure 19. A sample screenshot from the attribute table of resulting Functional Regions.