Digital Soil Mapping:
The SoLIM Approach
A Procedure Manual

Contents

1. Introduction

This manual is based on the materials compiled for the various workshops given on SoLIM using SoLIMSolutions. This manual is different from the Functionality Manual which comes with the SoLIMSolution distribution in that this manual provide a detailed description of procedures for digital soil mapping using the SoLIM method. The Functionality Manual provides a detail account of the functions available in SoLIMSolution. Users are recommended to use these two manuals together in actual operation of SoLIMSolutions.

2. DSM Using SoLIM Solutions 2015

SoLIM Solutions 2015 is the latest version of software employing SoLIM (Soil Land Inference Model). SoLIM is a new technology for soil mapping based on recent developments in Geographic Information Sciences (GIS), Artificial Intelligence (AI), and information representation theory. SoLIM was designed to overcome the limitations of existing soil survey methods and to improve the efficiency and accuracy of soil survey.

Case studies have shown that SoLIM is more efficient and accurate than traditional soil survey methods, in that it generates a range of products which the traditional approaches couldn't provide, and it can be employed in a production mode of soil survey. For more information about SoLIM, please visit the SoLIM website: http://solim.geography.wisc.edu/index.htm.

In this training, you need download the file and save it on C or D Disk in your computer. Please move/copy “Workshop_Data” folder to C or D Disk. In this handout, we assume that you save these two folders on D disk, such as *D:\Workshop_Data and D:\SoLIMSolutions2015*.

2.1 Covariate Extraction Based on Dynamic Patterns from Remote Sensing

- **Case study:**

Carman, Manitoba, Canada

- **Data:**

MODIS band1-7 data (resolution resampled to 250m) after a major rainfall event

- **Learning Goals:**

Create additional soil covariates from remote sensing data

- **Key References:**

Zhu, A.X., F. Liu*, B.L. Li, T. Pei, C.Z. Qin, G.H. Liu, Y.J. Wang, Y.N. Chen, X.W. Ma, F. Qi, C.H. Zhou, 2010. "Differentiation of soil conditions over flat areas using land surface feedback dynamic patterns extracted from MODIS", *Soil Science Society of America Journal*. 74(3), 861-869.

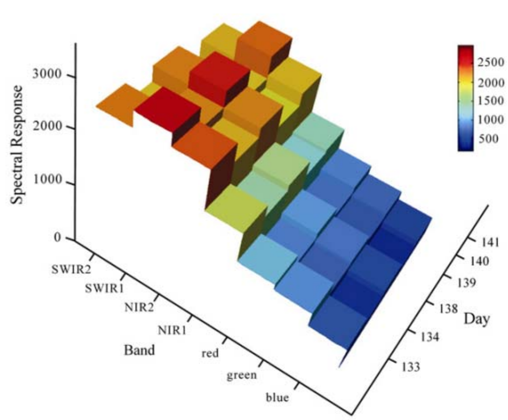
Liu, F., X. Geng, A.X. Zhu*, W. Fraser, 2012. "Digital soil mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS", *Geoderma*, 171–172, 44–52.

Guo, Shanxin, A-Xing Zhu*, Lingkui Meng, James E. Burt, Fei Du, Jing Liu, Guiming Zhang, (Accepted). "Unification of soil feedback patterns under different evaporation conditions to improve soil differentiation over flat area", *International Journal of Applied Earth Observation & Geoinformation*, Vol. 49, pp. 126-137.

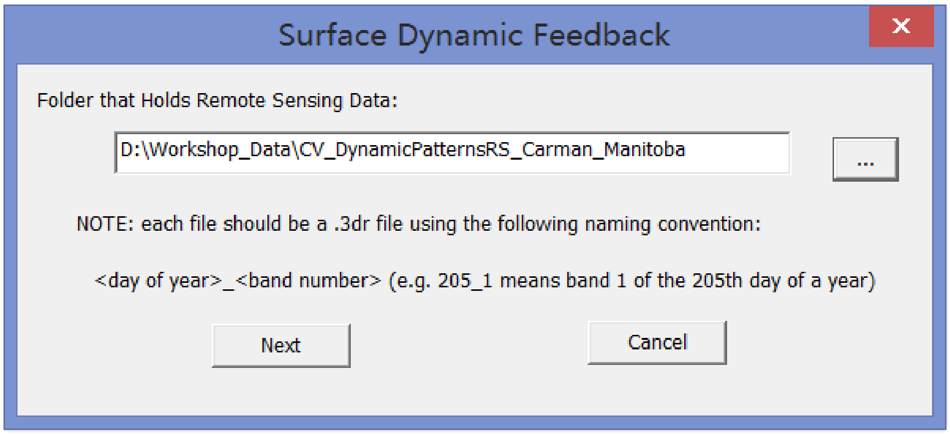
- **Operation Procedures:**

For the areas where the spatial variation of soil cannot be effectively indicated by the commonly-used covariates (e.g. terrain), land surface dynamic feedback captured by high temporal resolution remote sensing can be used to distinguish soils.

The multi-band, multiple day remote sensing data captured immediately after a major rainfall event are used to differentiate different soil types /properties. Each location (pixel) has a spatial-temporal response surface along band axis and temporal axis. Wavelet analysis is applied to summarize this surface for each location.



Choose “Data Preparation -> Remote Sensing Analysis”. Specify the folder that holds remote sensing data（\Workshop_Data\CV_DynamicPatternsRS_Carman_Manitoba）.

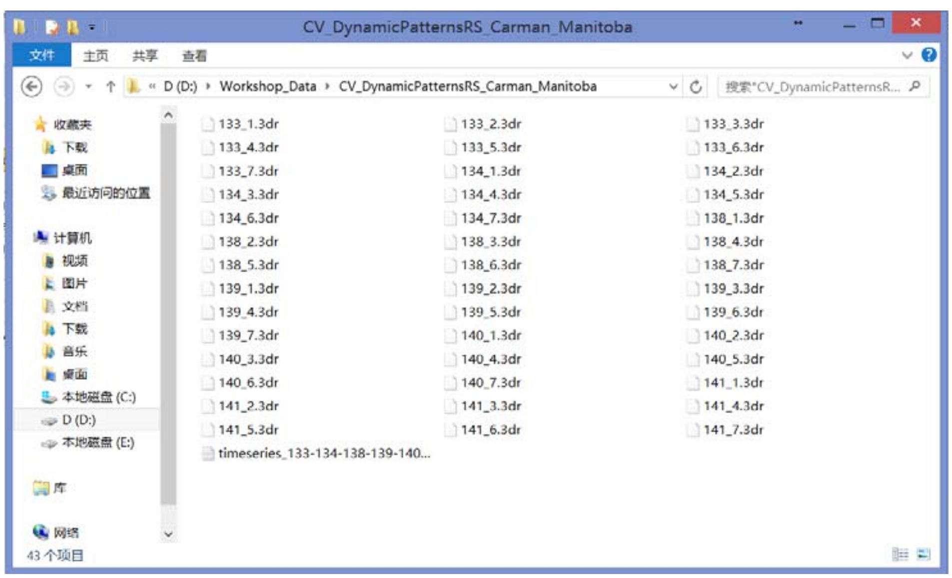


In this folder, each file should be a .3dr file with the following naming conventions:

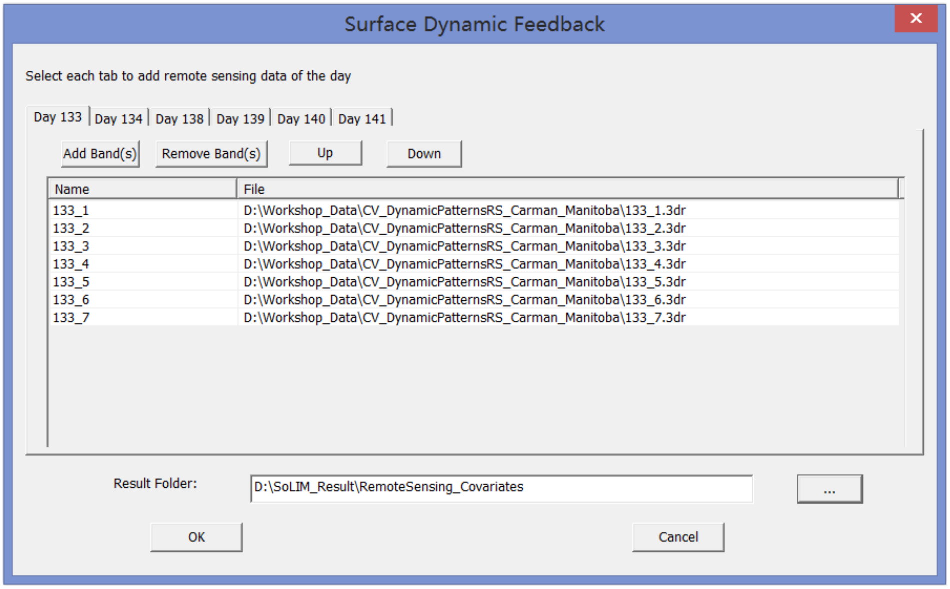
<day of year>_<band number>.3dr

For example, 134_3.3dr is the file that records band 3 of 134th day of a year.

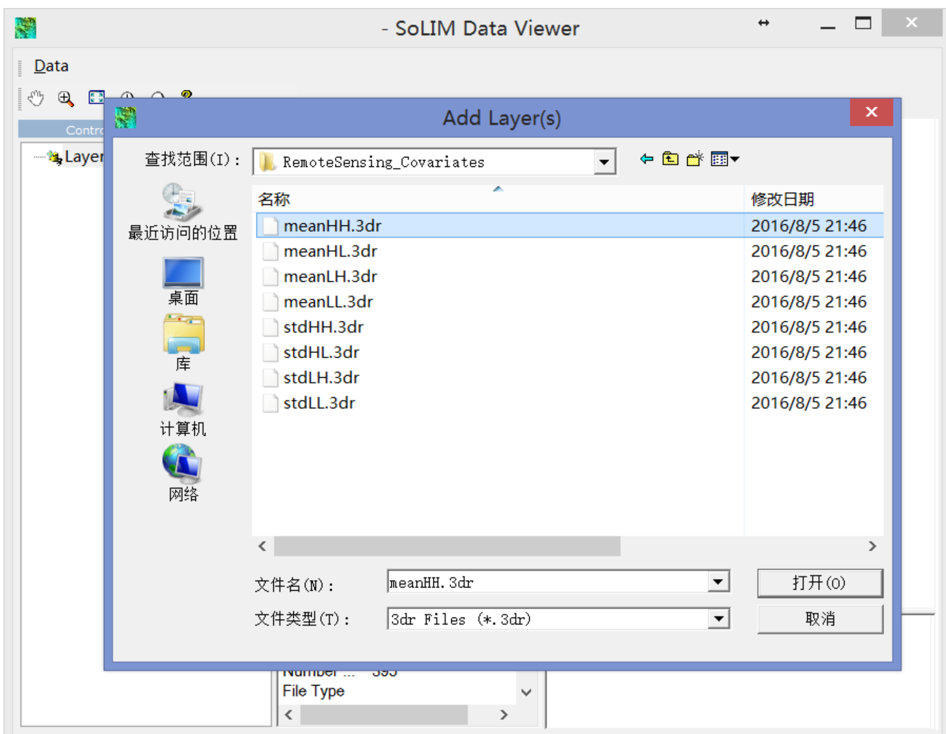
The structure of the provided folder for this workshop is shown below:



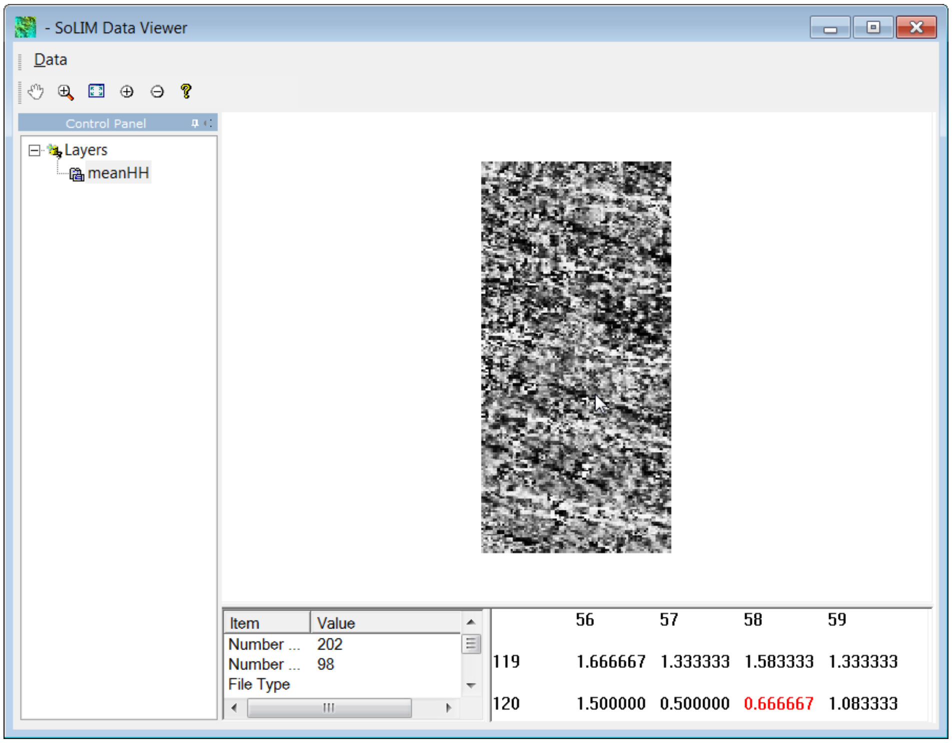
Click “Next”, SoLIM Solutions will parse the files in the folder and organize them based on days.



Specify the result folder and click “OK”. The generated covariates will be put into the result folder.



In the result folder, there will be eight new files generated from wavelet analysis. They are used to summarize the spatial-temporal response surface along different directions (e.g. horizontal direction, diagonal direction).You can visualize them in SoLIM Data Viewer.


- SoLIM Data Viewer
Data
Control Panel
Layers
meanHH
Item
Value
Number ... 202
Number ... 98
File Type
56 57 58 59
119 1.666667 1.333333 1.583333 1.333333
120 1.500000 0.500000 0.666667 1.083333

2.2 Rule-based Soil Mapping

2.2.1 Soil Mapping Based on Knowledge from Expert

- **Case Study:**

  Pleasant Valley, Wisconsin, USA

- **Data:**

  Knowledge of soil-landscape relationships from soil experts

  Environmental data (Elevation, Profile curvature, Slope gradient)

- **Learning Goal:**

  Encode expert knowledge to fuzzy rules

  Soil inference using fuzzy rules

  Generate hardened map

  Generate soil property map

- **References:**

Zhu, A.X., B. Hudson, J. Burt, K. Lubich, and D. Simonson, 2001. “Soil Mapping Using GIS, Expert Knowledge, and Fuzzy Logic”, *Soil Science Society of America Journal*. 65: 1563-1472

Zhu, A.X., Lawrence E. Band, Barry Dutton, Thomas J.Nimlos, 1996. “Automated soil inference under fuzzy logic”, *Ecological Modelling*. 90: 123-145.

- **Operation Procedure:**

  There are four soil types in Pleasant Valley: Basco, Elkmound, Council and Orion. You can see their distribution along a slope in Figure1. The environmental covariates you will use to predict soil types include slope gradient, elevation and profile curvature. In order to map soil types, you need to know the relationship between soil and the environmental variables. The relationship is usually expressed as fuzzy rules.

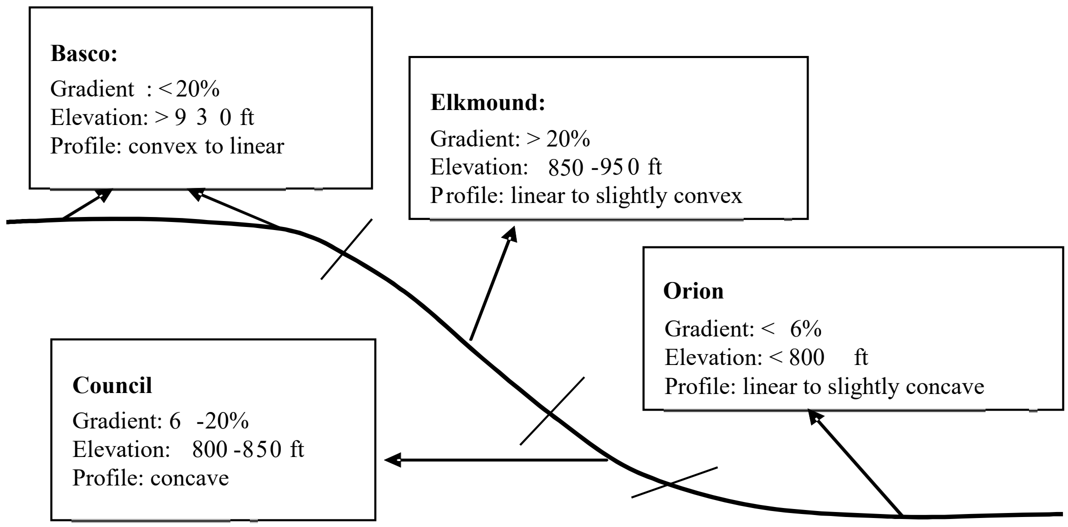


Figure 1: Knowledge of Soil-Landscape relationships for the Pleasant Valley Area

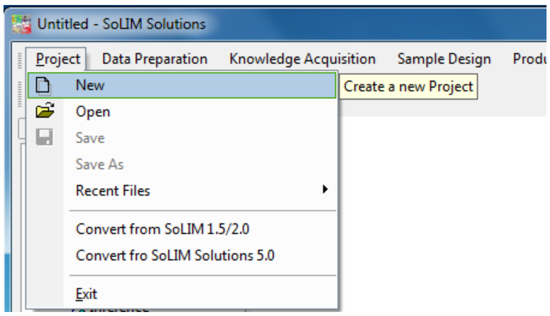
Development of fuzzy membership functions is a nontrivial process. For the purpose of this exercise, we have converted the rules in Figure 1 into key parameters that are needed in defining the fuzzy membership functions (bell-shaped, Z-shaped, and S-shaped curves) for each soil types in Table 1.

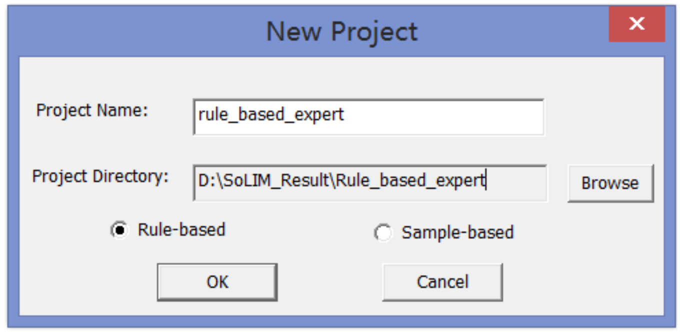
Table 1

Soil Types	Environmental Covariates	Shape	Lower crossover	Lower Optimal Value	Upper Optimal Value	Upper Crossover
Basco	Slope	Z-shape			0.15	0.2
	Elevation	S-shape	930	960		
	Profile	Z-shape			-0.002	0.003
Elkmound	Slope	S-shape	0.20	0.25		
	Elevation	bell-shape	760	870	870	980
	Profile	bell-shape	-0.0008	0	0.0003	0.0013
Council	Slope	bell-shape	0.06	0.10	0.16	0.2
	Elevation	Z-shape			820	880
	Profile	S-shape	0.00048	0.001		
Orion	Slope	Z-shape			0.03	0.05
	Elevation	Z-shape			780	800
	Profile	bell-shape	-0.0005	0	0.0003	0.0013

Step 1: Create a new project

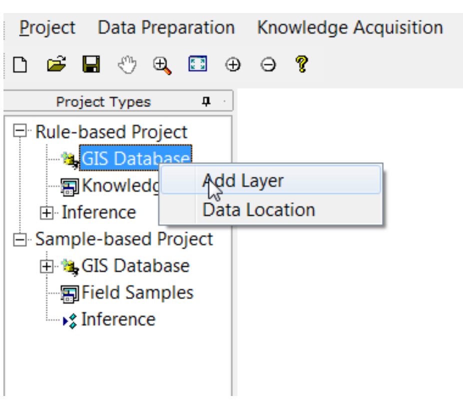
To conduct fuzzy soil mapping in SoLIM Solutions, you need to create a new project. Choose "Project -> new" on the main menu. Type the name of the project (rule_based_expert) and choose a directory to hold the project (D:\SoLIM_Result\Rule_based_expert). Keep other settings unchanged and click "OK".

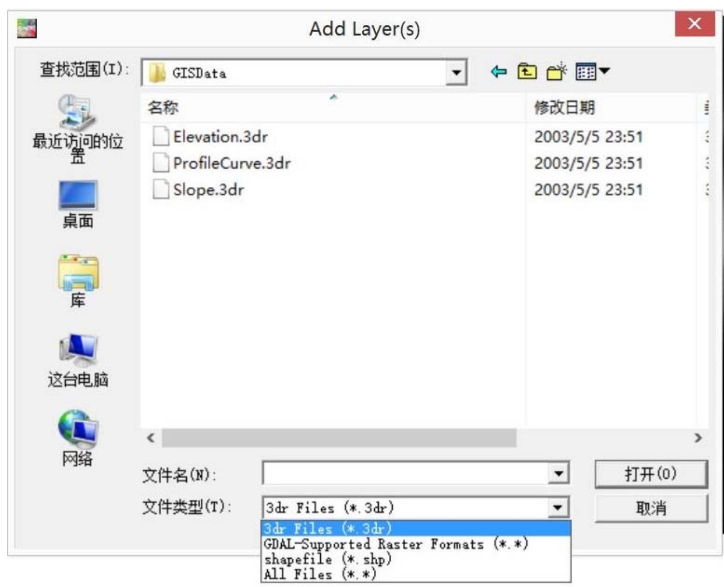




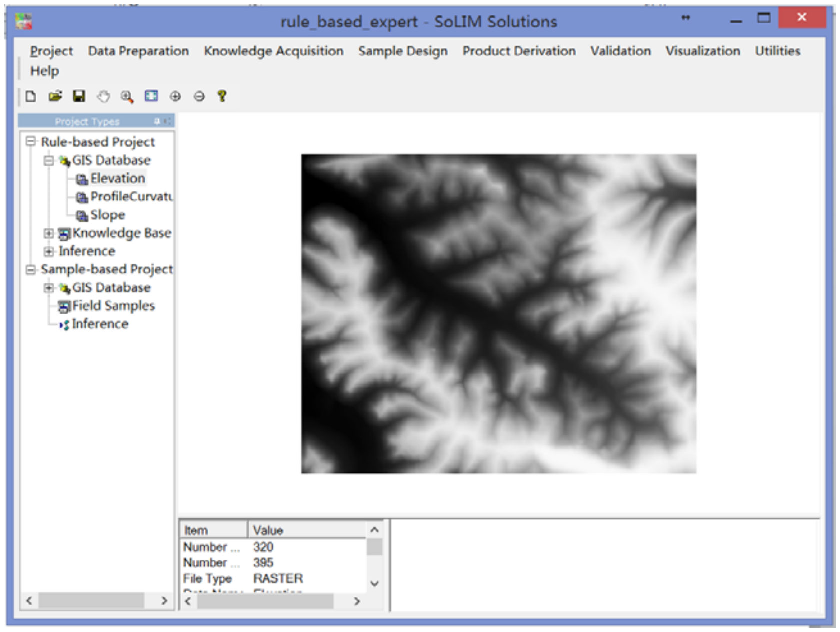
Step 2: Add GIS Data layers

Right click "GIS Database" under "Rule-based project" and select "Add Layer". Select all the environmental layers in the folder "\Workshop_Data\RB_ExpertKnowledge_PleasantValley\GISData\" and load them into the project.



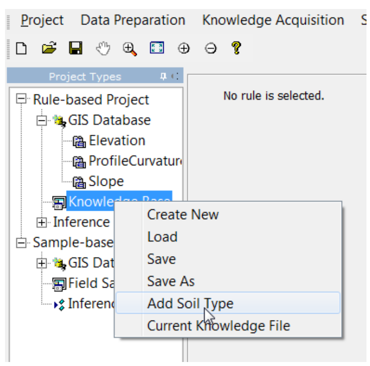


There are three types of file you can select, ".3dr", ".sdat", ".shp", respectively. If you select "GDAL-Supported Raster Formats" as input file type, you can choose a Saga file (the extension of the file should be ".sdat"), then click "open" button. In this exercise, you should select "3dr File" as input file type. Click on the name of an environmental layer under "GIS Database" node and view it in the panel on the right.

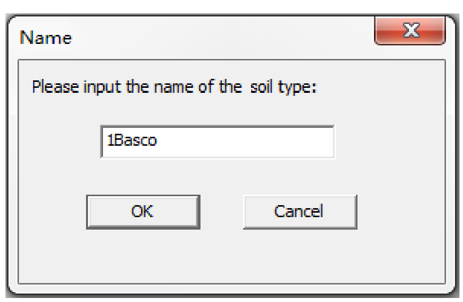


Step 3: Define fuzzy membership curves

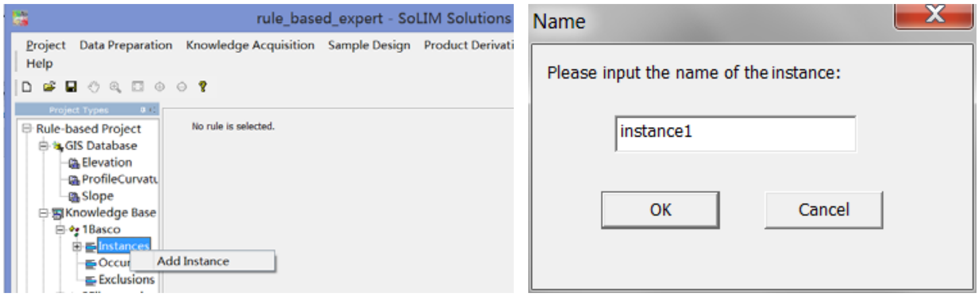
Right click on the "Knowledge Base" under "Rule-based Project" and choose "Add Soil Type".



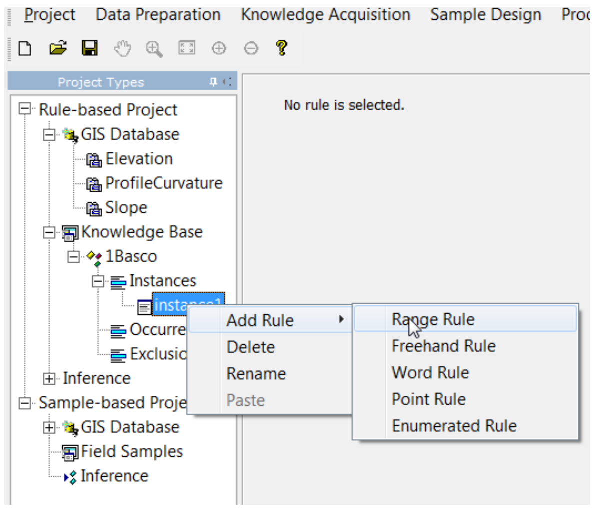
Type soil name "1Basco" in which 1 is the ID assigned to Basco soil. This is a naming convention of SoLIM Solutions. For other soil types, use different IDs. For example: 2Elkmound, 3Council, 4Orion.



Click "OK". Right click "Instances" under "1Basco" and select "Add Instance". Type "instance1" and click "OK".

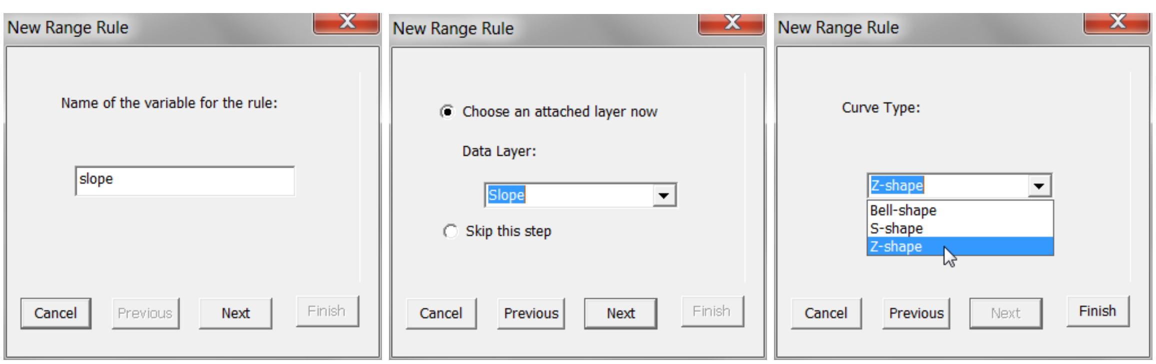


Now we can start to define rules for Basco. Right click on “instance1” and select “Add Rule->Range Rule”.

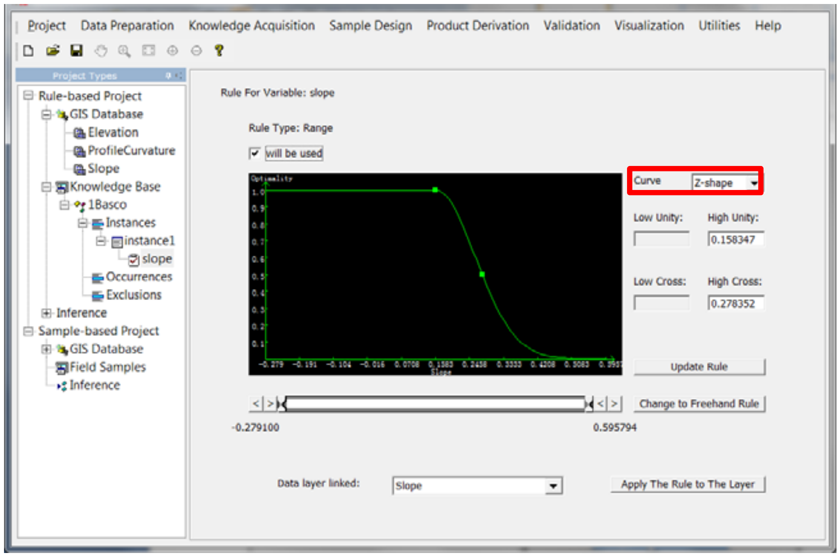


Follow the wizard to define the rule. Take slope rule for Basco as an example, as shown in Table 1, it is a Z-shaped curve with optimal value at 0.15 and right crossover point at 0.2. Specify the name of the rule, the attached layer and curve type.

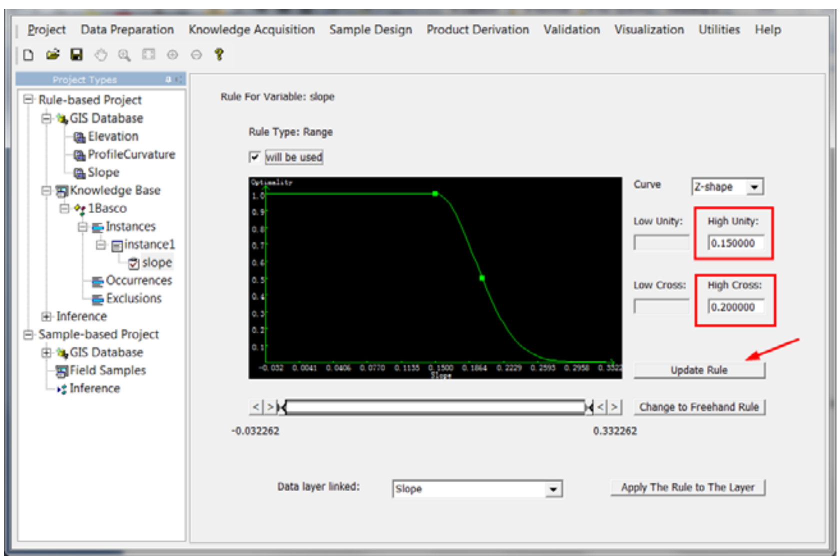
Caution: You must click on the button in the dialog rather than click “Enter” on the keyboard.



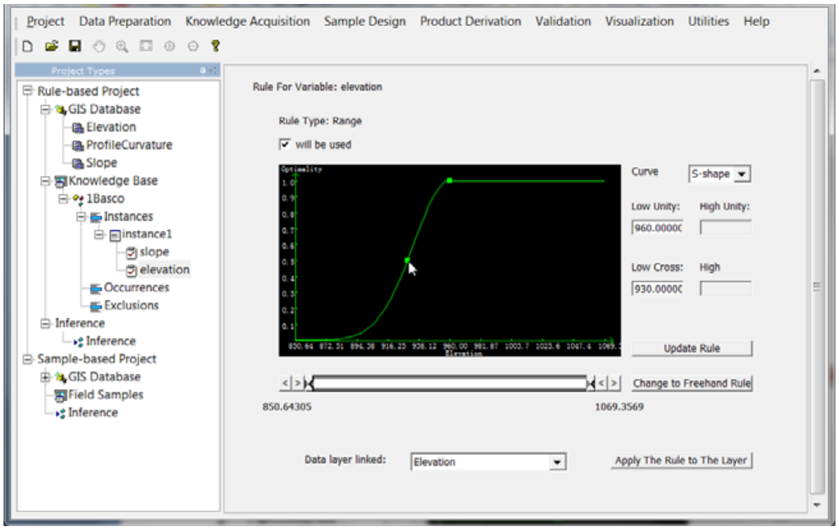
A default fuzzy membership curve will show up in the interface. There is a dropdown list you can change the curve shape when your selection of curve type is wrong in the previous step.

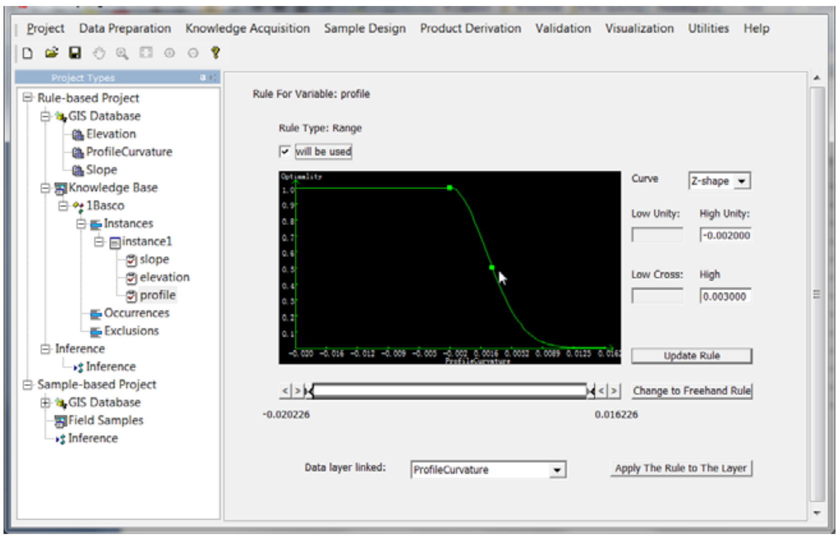


Adjust the optimal value (unity value) and crossover point by typing the values into the corresponding text boxes and click "Update Rule". You can see the changes in the graphical area after you update the rule.



Similarly, you can define fuzzy membership curves of Elevation and Profile curvature for 1Basco.

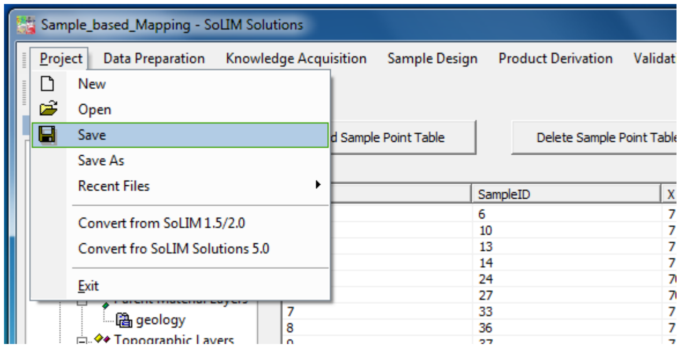




After you finish defining rules for Basco, you could define the rules for Elkmound, Council and Orion following similar procedures.

Step 4: Save the Project

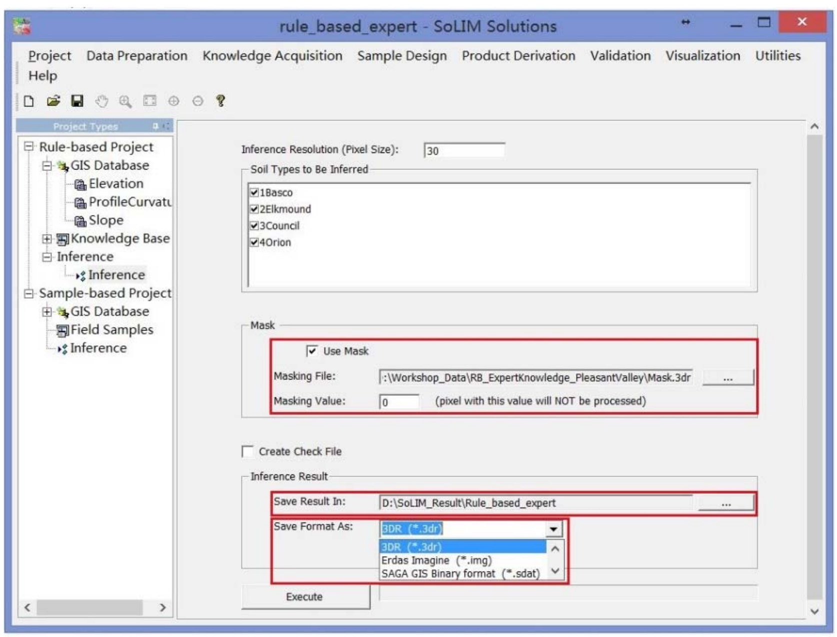
The project can be saved through “Project->Save”.



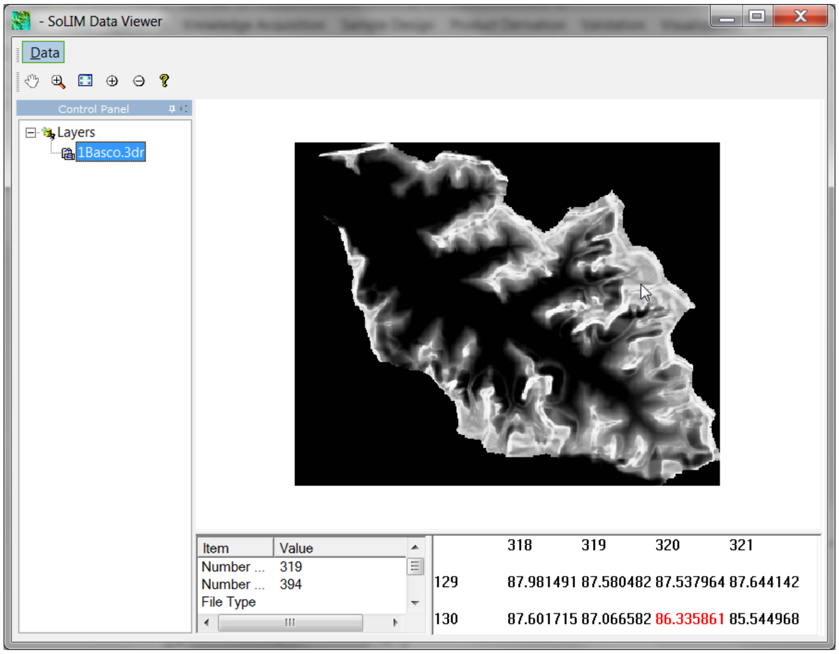
The project configuration and field samples will be saved into “rule_based_expert.sip” in the project directory. Then you can open a project through “Project -> Open”. The “GIS Dataset” and “Knowledge Base” that have already been saved in the .sip file will be opened.

Step 5: Generate fuzzy membership maps

Now you can start to make fuzzy soil maps. Switch to “Inference” panel. Select the soil types to be inferred and specify where to save the result and the output format. In this exercise, we select the masking file “Mask.3dr” and the masking value “0”. Pixels with the masking value will not be processed. Then click on “Execute”.

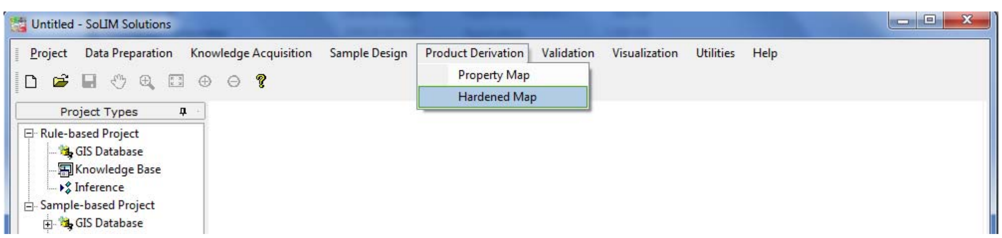


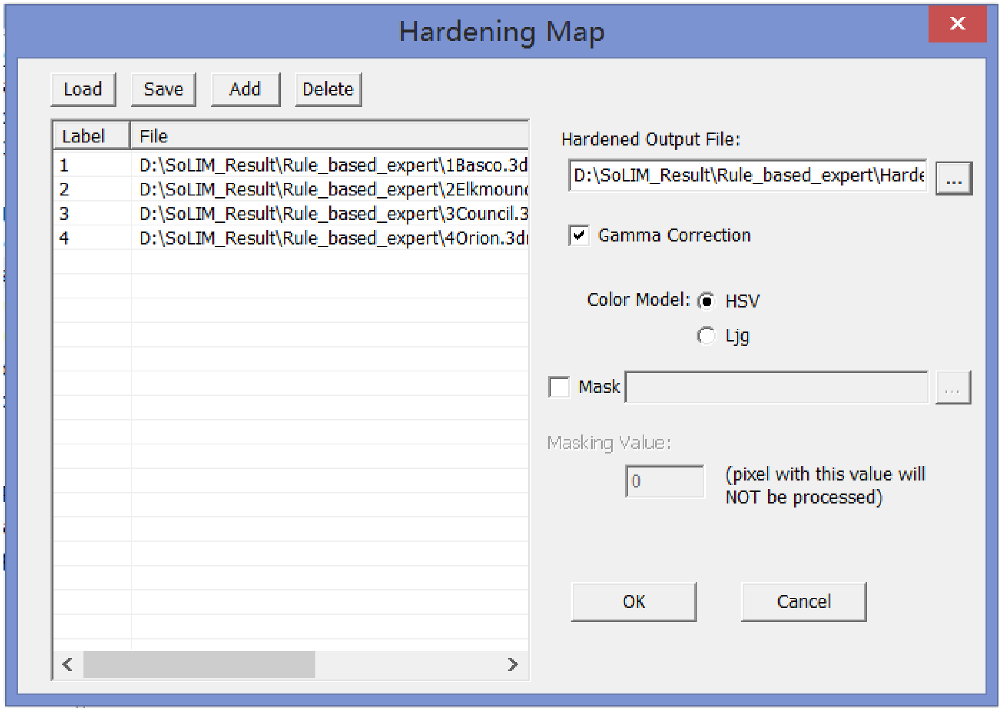
If the execution is successful, you can see four files in the result folder (e.g. a.3dr). Those “.3dr” files are fuzzy membership maps for the four soil types. You can use the Data viewer in SoLIM Solutions to explore them. Choose “Visualization->2D” to start it. In SoLIM Data viewer, choose “Data->Add Layer” to add .3dr files.



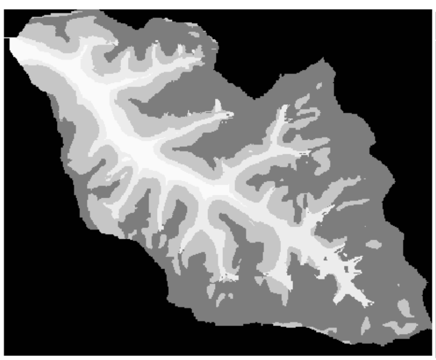
Step 6: Generate hardened soil map

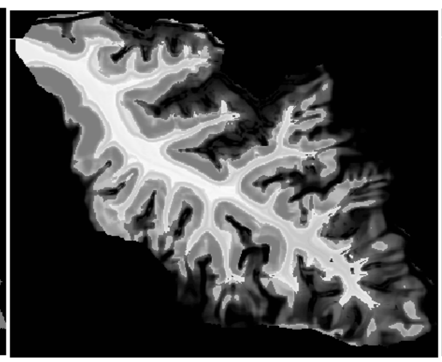
You can create a hardened soil map based on the fuzzy membership maps for the four soil types. Go to "Product Deviation -> Hardened Map" on the main menu. Add the fuzzy membership maps and specify the output location. By hardening each location will be assigned the soil types to which the location has the maximum membership.

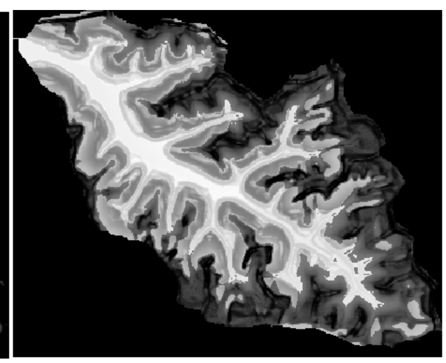




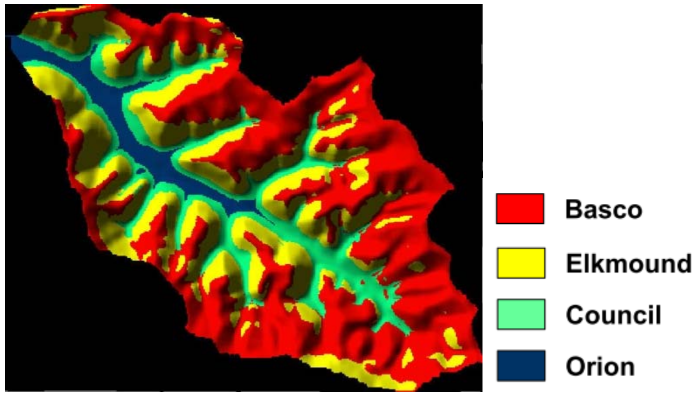
If the hardening process is finished successfully, three files should be created in the same directory. One is the hardened map, the other two files are entropy map and exaggerated uncertainty map, respectively.


HardenUnMap.3dr


HardenUnMapEnt.3dr


HardenUnMapExg.3dr

If you want to express the hardened map in the following way, please contact jeburt@wisc.edu to get the Authorization Number of 3dMapper software. The 3dMapper Help provides a brief introduction to 3-d landscape visualization and mapping. The final hardened soil map would look like this:



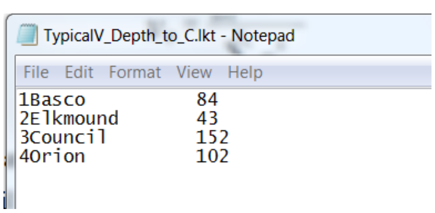
Step 7: Generate soil property map

Another product that can be derived from the fuzzy membership maps is soil property map. A look-up table that lists the typical soil property of each soil type should be prepared first. A weighted average approach is used to get the final soil property for each location using the following equation:

$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} V^{k}}{\sum_{k=1}^{n} S_{ij}^{k}}$$

where v_{ij} is the property at site (i, j); v^k is the typical value of a given soil property of soil type k; s_{ij}^k is the fuzzy membership of soil type k at site (i, j); and n is the total number of prescribed soil type in the area.

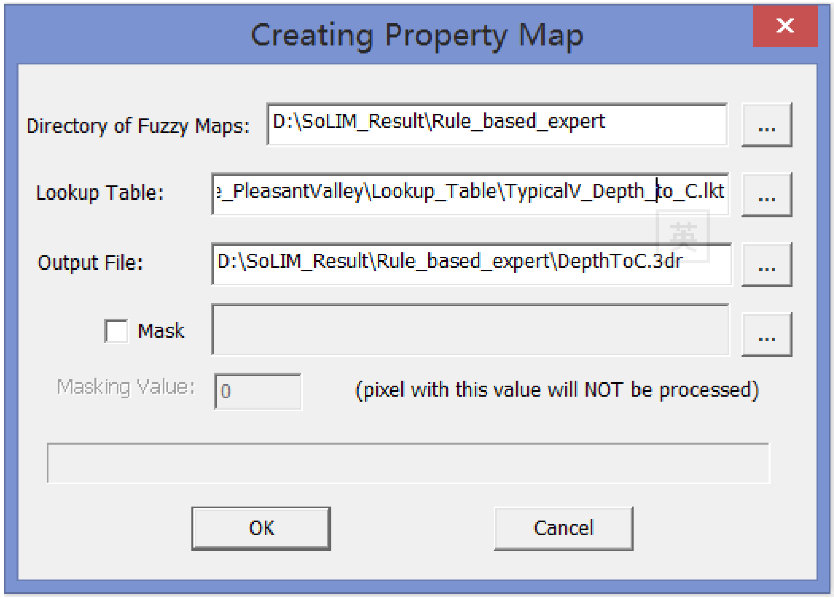
In this exercise, we will create the depth to soil C horizon map for Pleasant Valley. You may want to open the provided look-up table to take a look at its structure first. The first column is the soil fuzzy membership name and the second column is the property value (for the soil type).


TypicalV_Depth_to_C.lkt - Notepad

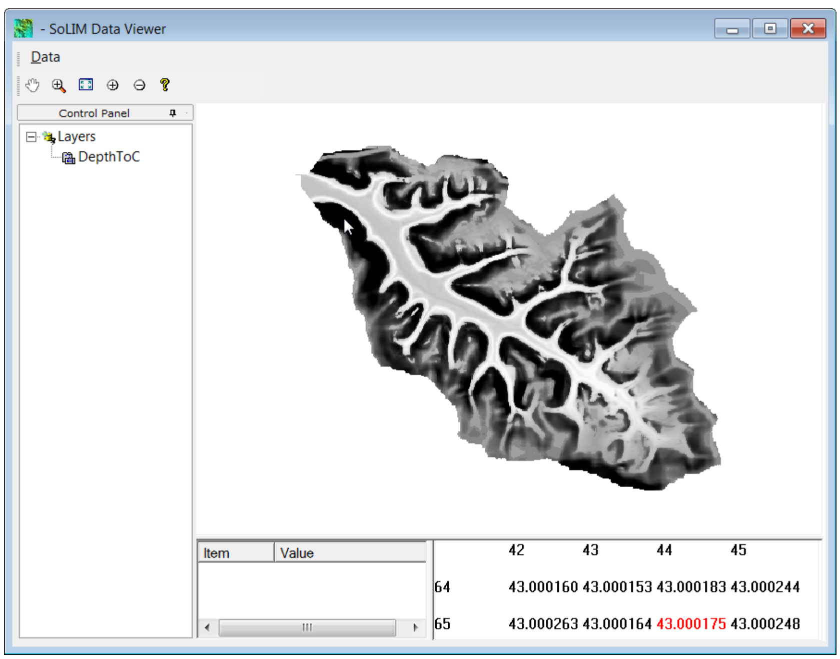
File Edit Format View Help

```
1Basco          84
2Elkmound       43
3Council        152
4Orion          102
```

Choose "Product Derivation -> Property Map". The directory of fuzzy maps is the result folder that holds the fuzzy membership maps. Specify the path to the look-up table which is provided to you. Specify the output file name and click "OK".



Once the calculation process is finished successfully, you can view the result file using SoLIM Data Viewer.



You can also use "Utilities->Data Format Conversion -> 3dr to Grid Ascii" to convert the property map (.3dr) to .asc (ASCII file).

Step 8: Validation of results

SoLIMSolutions provides validation of two products as derived above (soil class map and soil property map). Validation of soil property map validation can be done using the step 6 in **2.3 Sample-based**

Soil Mapping. Validation of soil class map produced in Step 6 above, also requires you to provide a set of independent validation samples. At each sample soil class type will be needed to be reported. The validation samples are stored in a text file in one of the predefined formats (see the Functionality Manual for details). Validation of soil class map is done through the “Type Validation” under the “Validation” menu.

2.2.2 Soil Mapping Using Rules Extracted from Soil Maps

- **Case Study:**

  Raffelson, Wisconsin, USA

- **Data:**

  NRCS SSURGO soil map

  Environmental data (Elevation, Slope gradient, Profile curvature, Planform curvature)

- **Learning Goals:**

  Generate and modify knowledge curves (probability density functions)

  Import curves into SoLIM Solutions for soil mapping (soil types and soil property)

- **References:**

Qi, F. and A.X. Zhu, 2003. "Knowledge discovery from soil maps using inductive learning", *International Journal of Geographic Information Science.* Vol. 17, No. 8, pp. 771–795.

Qi, F., Zhu, A-X., Pei, T., Qin, C., and Burt, J.E., 2008. "Knowledge discovery from area-class resource maps: capturing prototype effects", *Cartography and Geographic Information Science*, Vol. 35, No. 4, pp. 223-237.

Du, F., A-Xing Zhu, Lawrence Band, J. Liu, 2015. "Soil property variation mapping through data mining of soil category maps", *Hydrological Processes*, 29, 2491–2503.

- **Operation Procedure:**

Step 0: File Preparation (done in this exercise)

1) Environmental data layers

Environmental data layers such as elevation, slope gradient, slope aspect, planform curvature, profile curvature, topographic wetness index, etc., are required in .3dr format. You may find environmental layers we need have been prepared for this exercise in the folder "\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data". The variables used are selected by the soil scientist according to the belief regarding what variables are likely to be helpful in separating soil classes from one another.

2) Soil survey file

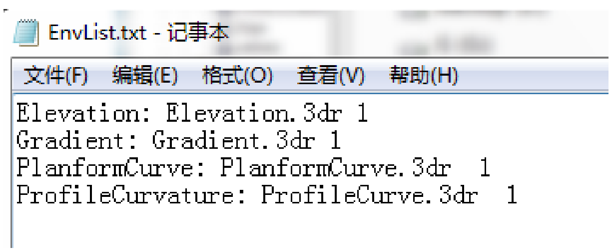
The existing soil survey needs to be in polygon shape file format and the attribute table should contain at least two fields: Polygon ID and Map Unit Key. Polygon ID is a unique identifier of each soil polygon and Map Unit Key is a unique identification of each map unit name (i.e. soil type name). Polygon ID must be integer and Map Unit Key can be integer or string (no spaces). Please note that every polygon must be assigned a map unit. In this exercise, you may use the SSURGO soil map of Raffelson area produced by NRCS in the folder "\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data\nrcs_soil.shp"

3) Environmental data layer list file

This file lists the environmental data layers to be used to perform the analysis. It is a plain text file and should follow the format shown below:

Variable:	FileName	DataType
Variable:	FileName	DataType
… …		

"Variable" is the label or tag for a data layer, "FileName" contains the file name of this data layer, and "DataType" indicates the data type of this data layer. "DataType" needs to be "1" (for continuous data type) or "2" (for categorical data type). "Variable", "FileName", and "DataType" are delimited using space or "Tab". In this exercise, this file has been prepared for you. It is located in the folder "\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data" with name "EnvList.txt".




```
Elevation: Elevation.3dr 1
Gradient: Gradient.3dr 1
PlanformCurve: PlanformCurve.3dr  1
ProfileCurvature: ProfileCurve.3dr  1
```

4) Map unit list file

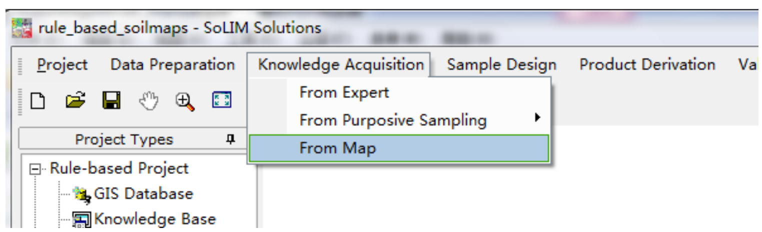
This file lists the map units used to perform the analysis. It's a plain text file and can be created through "File -> New Map Unit List …". The file format is:

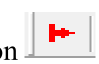
MUKey	SoilLabel
MUKey	SoilLabel
… …	

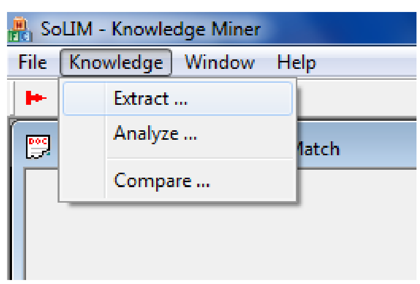
"MUKey" is the key for a map unit and "SoilLabel" is the corresponding label for the map unit. "MUKey" and "SoilLabel" are delimited by space or "Tab". This file also has been prepared for you. It is located in the folder "\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data" with name "mapUnitList.txt".

Step 1: Knowledge extraction

On the main menu of SoLIM Solutions 2015, select "Knowledge Acquisition->From Map" to start SoLIM-Knowledge Miner.

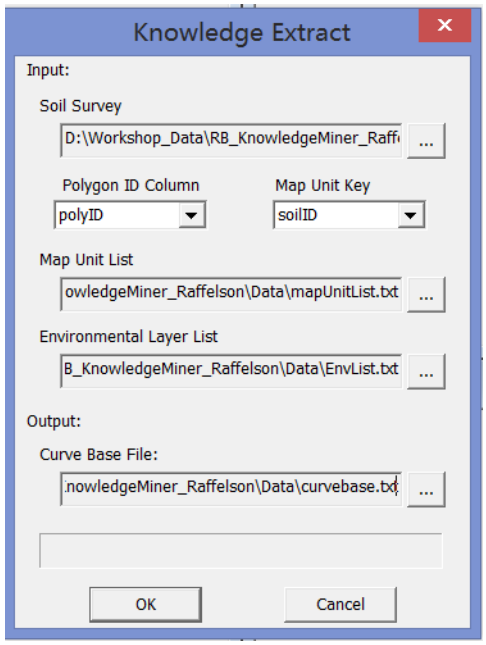


Start the knowledge extraction interface by hitting the button  or go to "Knowledge --> Extract…" on the main menu.



In the pop-up window, specify the path to the Soil Survey File (\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data\nrcs_soil.shp). Choose the field (polyID) for polygon ID and the field (SoilID) for Map Unit Key in the following dropdown lists. Specify the path to the Map Unit List File (\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data\mapUnitList.txt) and Environmental Layer List File (\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data\EnvList.txt). Set the name and the saving path of the output curve base file (\Workshop_Data\RB_KnowledgeMiner_Raffelson\Data\curvebase.txt), then click "OK".

Caution: the curve base file, the map unit list file and the environmental layer list file must be in the same folder.

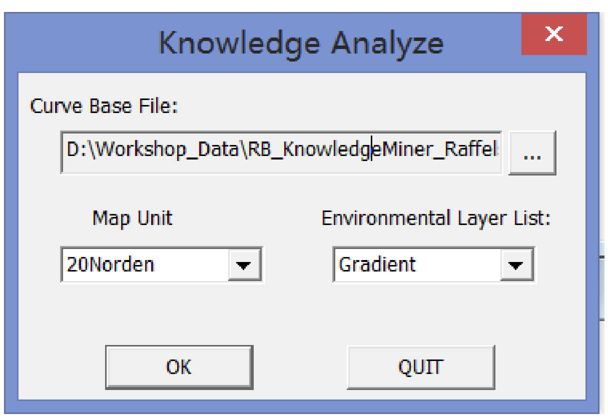


Step 2: Knowledge analysis

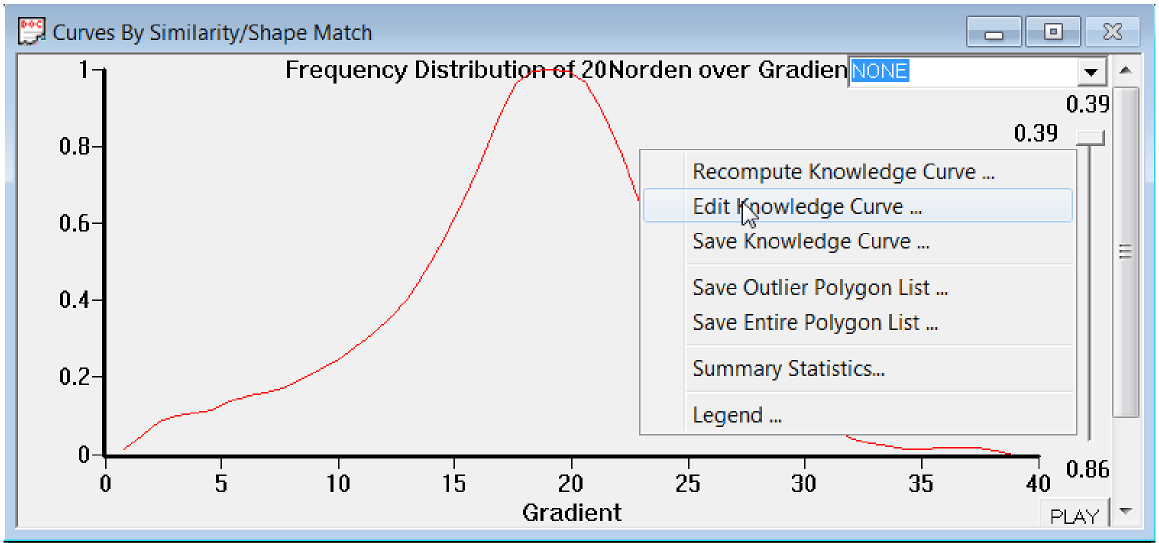
Knowledge analysis is normally performed for every combination of map unit and environmental data layer.

Go to “Knowledge --> Analyze …” or hit the button  on the tool bar to start the knowledge analysis interface.

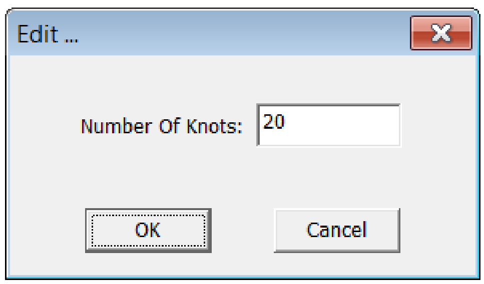
In the pop-up window “Knowledge Analyze” window, specify the path to the curve base file generated in “Extracting Knowledge” (step 1). Select a map unit and an environmental layer to be analyzed.



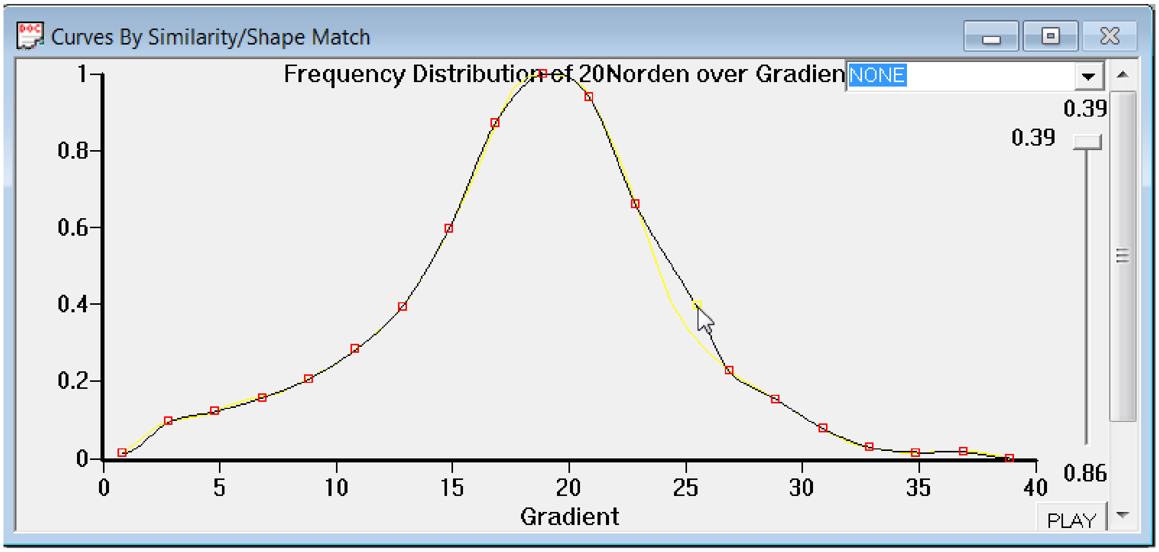
Click on “OK” and you will see the frequency distribution of the soil type you choose over the specified environmental variable displayed in the top-left panel. This curve can be used as the membership curve to the soil type over the specified environmental variable. This curve is editable. Right click on the curve and choose "Edit the Knowledge Curve..."



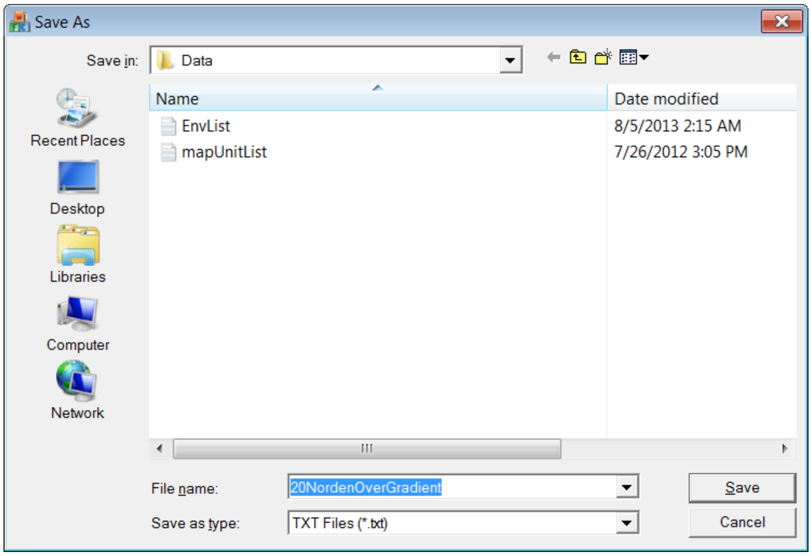
You may set the number of knots used to define the shape of the curve：



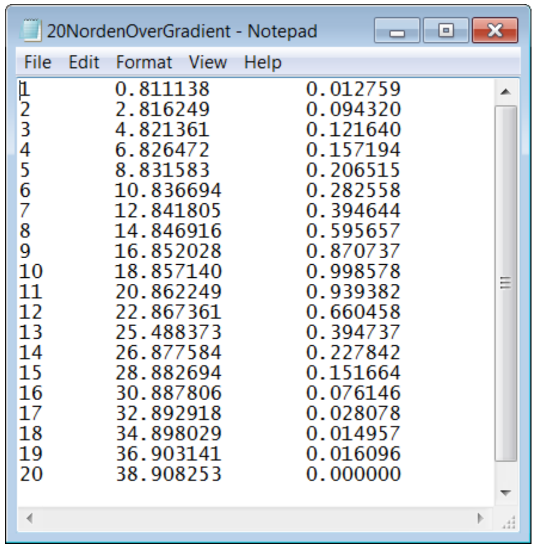
Click on OK and you will see the knots showing up along the curve. You can drag the knot(s) to adjust the shape of the curve.



When you finish editing, you can save your edits. Right click on the curve and choose "Save Knowledge Curve", the curve will be saved in a .txt file as following:



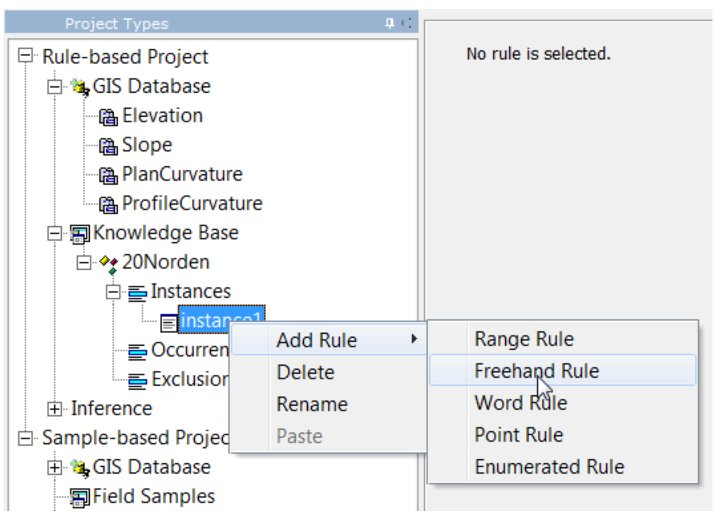
An example of membership curve is shown below. The first column is the IDs of the knots used to define the shape of the curve, the second column is the values of environmental variable at those knots and the third column is the corresponding membership to the soil type.



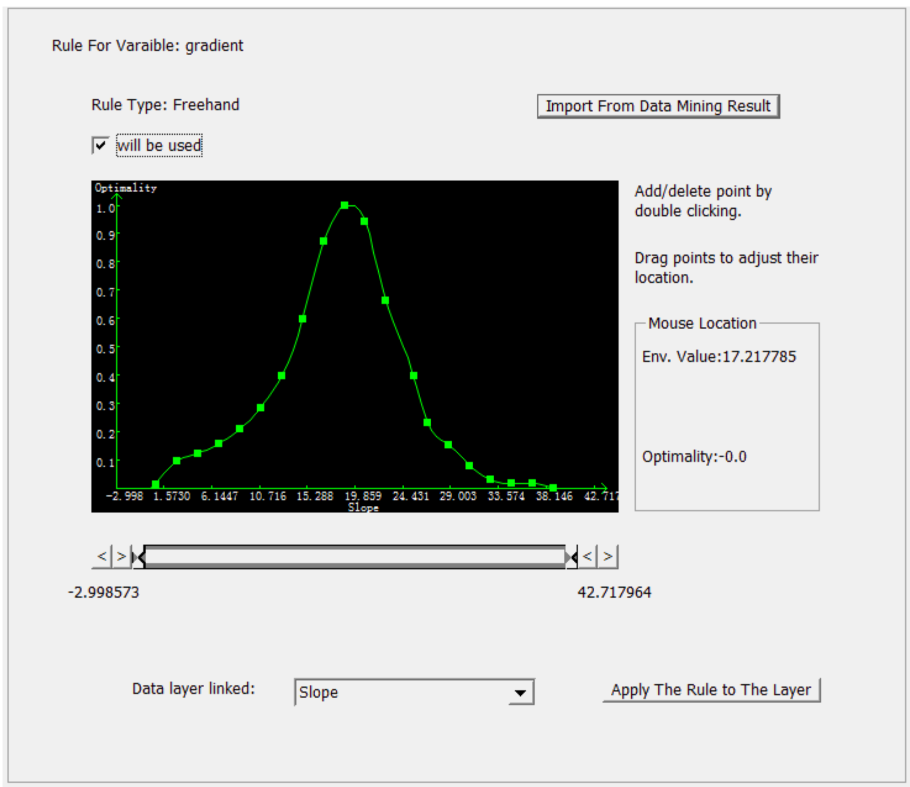
1	0.811138	0.012759
2	2.816249	0.094320
3	4.821361	0.121640
4	6.826472	0.157194
5	8.831583	0.206515
6	10.836694	0.282558
7	12.841805	0.394644
8	14.846916	0.595657
9	16.852028	0.870737
10	18.857140	0.998578
11	20.862249	0.939382
12	22.867361	0.660458
13	25.488373	0.394737
14	26.877584	0.227842
15	28.882694	0.151664
16	30.887806	0.076146
17	32.892918	0.028078
18	34.898029	0.014957
19	36.903141	0.016096
20	38.908253	0.000000

Step 3: Import knowledge curves into SoLIM Solutions for soil mapping

The generated curves can be imported into SoLIM Solutions for soil mapping. In step 3 of the training case described in 2.2.1 (Soil Mapping Based on Knowledge from Expert), the type of the new rules is "Range Rule". In order to use the knowledge curve, choose "Freehand Rule" as the rule type instead.



Click "Import From Data Mining Result" and specify the knowledge curve file (.txt file). The specified curve will be imported.



Step 4: The remaining steps

Other steps are the same as using range rules. After adding rules you can inference soil type. The method is same as the step 5 of the training case described in 2.2.1 (Soil Mapping Based on Knowledge from Expert).

2.2.3 Soil Mapping Based on Knowledge from Purposive Sampling

- **Case Study:**

  Heshan, Nenjiang County, Heilongjiang Prov., China

- **Data:**

  Environmental data (Slope gradient, Planform curvature, Profile curvature, Topographic wetness index)

- **Learning Goals:**

  Fuzzy clustering to generate sample design table

  Prepare field sample table

  Extract knowledge base from field sample table for soil mapping (soil type and property)

- **References:**

Zhu, A.X., L. Yang; B. Li, C. Qin, E. English, J. E. Burt, C.H. Zhou, 2008. "Purposefully sampling for digital soil mapping". In: A.E. Hartemink, A.B. McBratney and M.L. Mendonca Santos (eds.) Digital Soil Mapping with Limited Data, Springer-Verlag, New York, pp. 233-245.

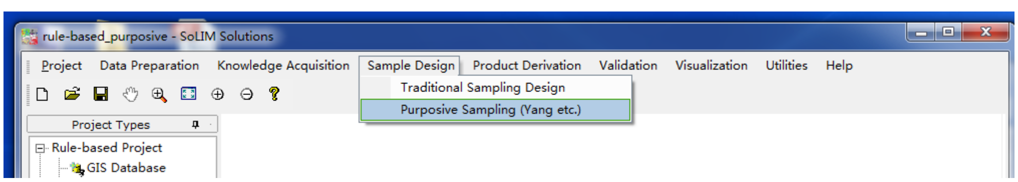
Zhu, A.X., L. Yang, B.L. Li, C.Z. Qin, T. Pei, B.Y. Liu, 2010. "Construction of membership functions for predictive soil mapping under fuzzy logic", *Geoderma*. Vol. 155, No. 3-4, pp. 166-174.

Yang, L., A.X. Zhu, F. Qi, C. Qin, B. Li, T. Pei, 2012. "An integrative hierarchical stepwise sampling strategy for spatial sampling and its application in digital soil mapping", *International Journal of Geographical Information Science*, pp.1-23.

- **Operation Procedure:**

Step 1: Generate suggested sample locations

In SoLIM Solutions, choose "Sample Design ->Purposive Sampling (Yang etc.)"



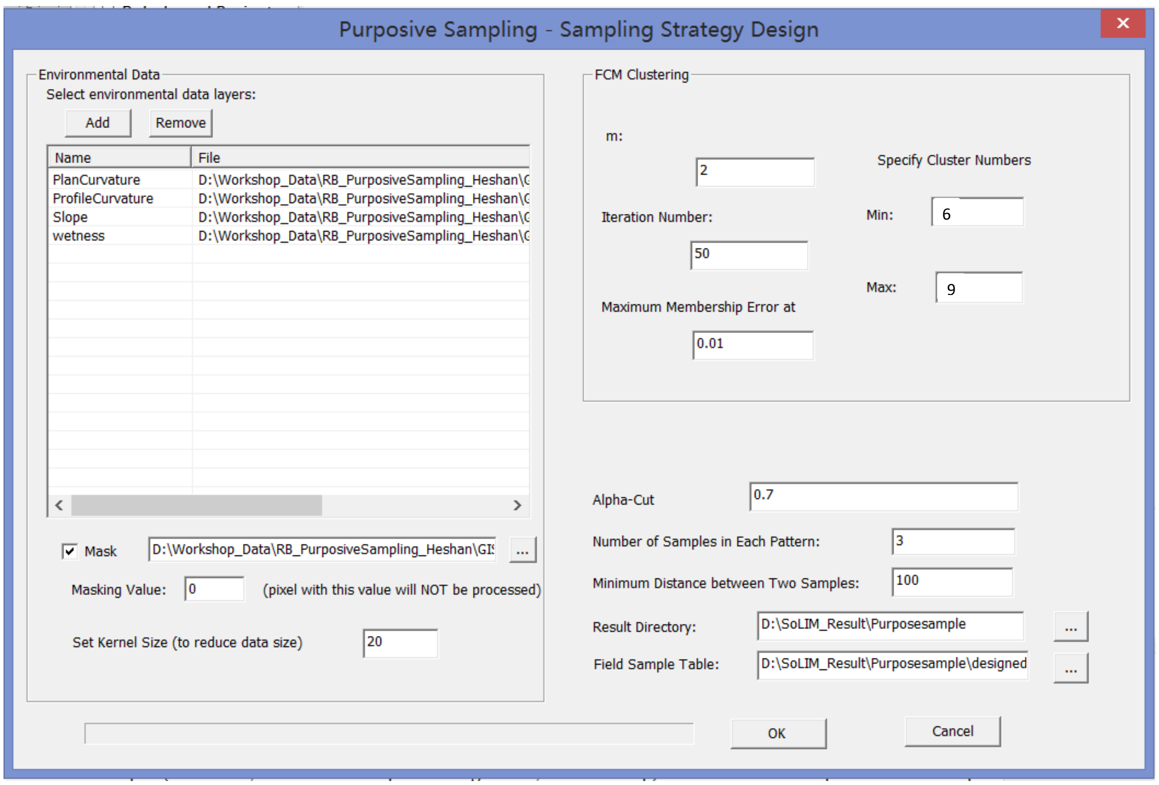
Add the four provided environmental layers by clicking "Add" button (\Workshop_Data\RB_PurposiveSampling_Heshan\GISData). Specify the mask file (\Workshop_Data\RB_PurposiveSampling_Heshan\GISData\mask02.3dr) and masking value. Pixels with the masking value will not be processed.

As FCM clustering is computational intensive in extensive area, you can also set the kernel size larger than 1 to resample the data so that the computation takes less time. In this exercise, we set the kernel size to 100.

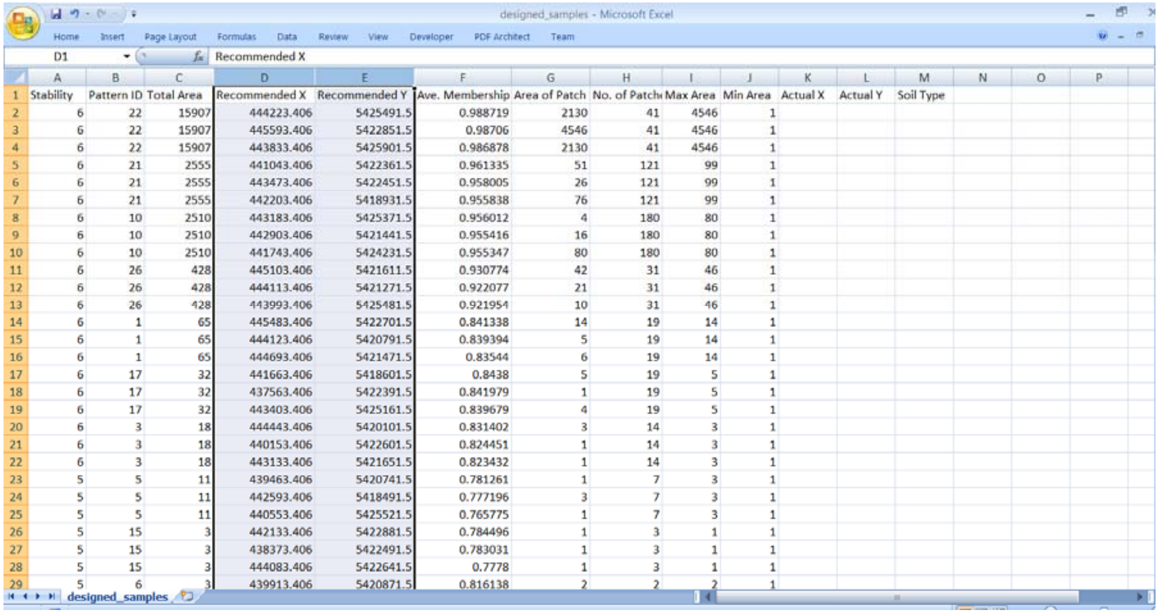
Set the basic parameters for FCM clustering. For cluster numbers, set the minimum cluster number to 6 and maximum cluster number to 8 for this case study. For other parameters, keep the default setting.

You also need specify the alpha-cut value above which a fuzzy membership value can be regarded as high fuzzy membership. You can also determine how many samples you want to get for each detected pattern and the minimum distance between two samples. Keep the default setting for this case study.

Finally, specify the result directory which holds all intermediate results and the path to the field sample table which records the designed samples.



Click "OK". The computation may take some time. When the execution is finished successfully, a table containing the suggested sample locations can be found in the path you specified. In this table, the recommended x and y coordinates of each suggested sample are listed. You can also find the stability of each sample (how many times this sample has high fuzzy membership) and the ID of the pattern each sample belongs to.



Stability	Pattern ID	Total Area	Recommended X	Recommended Y	Ave. Membership	Area of Patch	No. of Patch	Max Area	Min Area	Actual X	Actual Y	Soil Type
6	22	15907	444223.406	5425491.5	0.988719	2130	41	4546	1			
6	22	15907	445593.406	5422851.5	0.98706	4546	41	4546	1			
6	22	15907	443833.406	5425901.5	0.986878	2130	41	4546	1			
6	21	2555	441043.406	5422361.5	0.961335	51	121	99	1			
6	21	2555	443473.406	5422451.5	0.958005	26	121	99	1			
6	21	2555	442203.406	5418931.5	0.955838	76	121	99	1			
6	10	2510	443183.406	5425371.5	0.956012	4	180	80	1			
6	10	2510	442903.406	5421441.5	0.955416	16	180	80	1			
6	10	2510	441743.406	5424231.5	0.955347	80	180	80	1			
6	26	428	445103.406	5421611.5	0.930774	42	31	46	1			
6	26	428	444113.406	5421271.5	0.922077	21	31	46	1			
6	26	428	443993.406	5425481.5	0.921954	10	31	46	1			
6	1	65	445483.406	5422701.5	0.841338	14	19	14	1			
6	1	65	444123.406	5420791.5	0.839394	5	19	14	1			
6	1	65	444693.406	5421471.5	0.83544	6	19	14	1			
6	17	32	441663.406	5418601.5	0.8438	5	19	5	1			
6	17	32	437563.406	5422391.5	0.841979	1	19	5	1			
6	17	32	443403.406	5425161.5	0.839679	4	19	5	1			
6	3	18	444443.406	5420101.5	0.831402	3	14	3	1			
6	3	18	440153.406	5422601.5	0.824451	1	14	3	1			
6	3	18	443133.406	5421651.5	0.823432	1	14	3	1			
5	5	11	439463.406	5420741.5	0.781261	1	7	3	1			
5	5	11	442593.406	5418491.5	0.777196	3	7	3	1			
5	5	11	440553.406	5425521.5	0.765775	1	7	3	1			
5	15	3	442133.406	5422881.5	0.784496	1	3	1	1			
5	15	3	438373.406	5422491.5	0.783031	1	3	1	1			
5	15	3	444083.406	5422641.5	0.7778	1	3	1	1			
5	6	3	439913.406	5420871.5	0.816138	2	2	2	1			

Step 2: Determine the actual sampling locations and conduct field sampling

You can then use this table to guide your field sampling. You may determine the actual sampling locations with the considerations on accessibility and other field conditions. After you finish filed sampling, fill in ActualX, ActualY and Soil Type column.

Step 3: Extract knowledge base from field sample table

A field sample table should be in .csv format and contain at least three columns: ActualX, ActualY and Soil Type. A field sample table has been provided to you for this exercise (\Workshop_Data\RB_PurposiveSampling_Heshan \Actual Samples.csv).

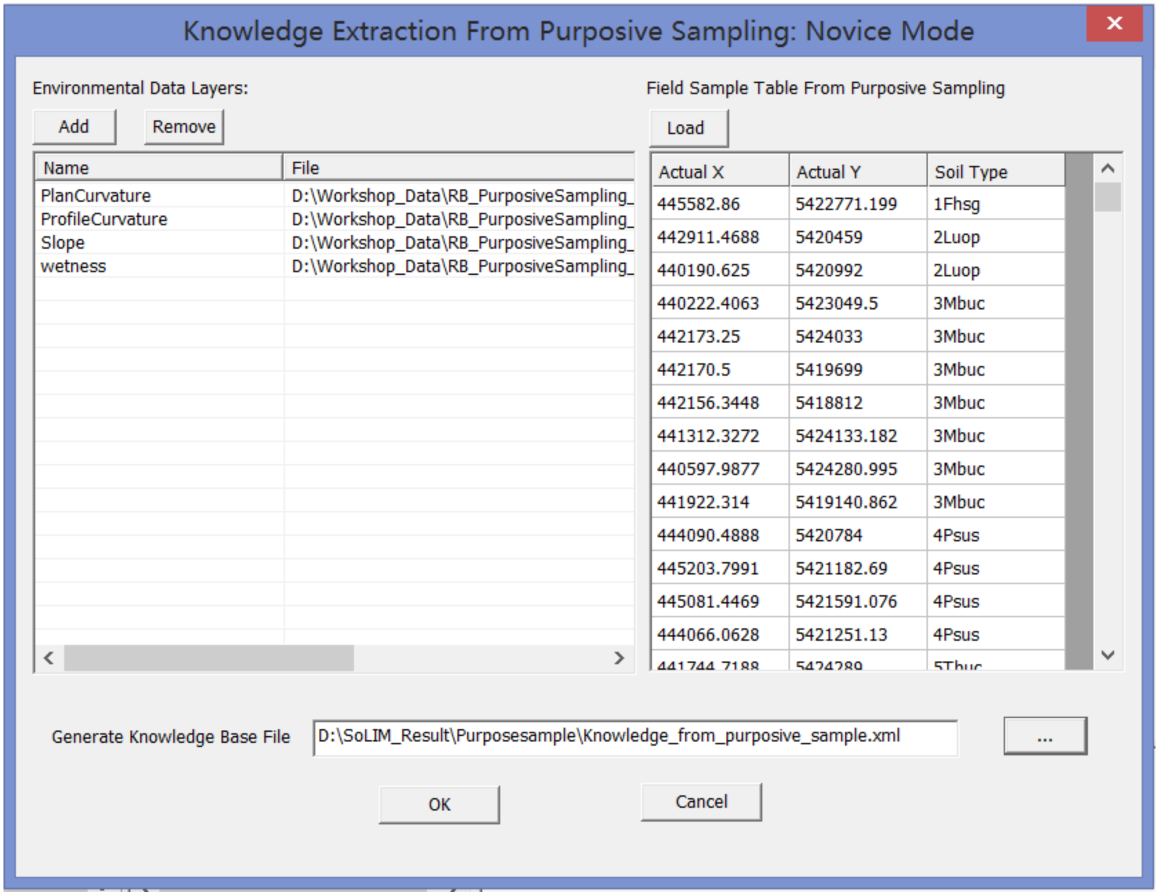
A	B	C
Actual X	Actual Y	Soil Type
445582.9	5422771	1Fhsg
442911.5	5420459	2Luop
440190.6	5420992	2Luop
440222.4	5423050	3Mbuc
442173.3	5424033	3Mbuc
442170.5	5419699	3Mbuc
442156.3	5418812	3Mbuc
441312.3	5424133	3Mbuc
440598	5424281	3Mbuc
441922.3	5419141	3Mbuc
444090.5	5420784	4Psus
445203.8	5421183	4Psus
445081.4	5421591	4Psus
444066.1	5421251	4Psus
441744.7	5424289	5Tbuc
441389	5419952	5Tbuc
441304.8	5420240	5Tbuc
439655.4	5420429	5Tbuc
440876.8	5426079	5Tbuc
439653	5420111	5Tbuc
441804.2	5423861	5Tbuc
442172.7	5419961	5Tbuc

With this table, you can start to extract knowledge.

Choose “Knowledge Acquisition-> From Purposive Sampling -> Novice Mode”. In the pop-up dialog, Click “Add” to load environmental layers from (\Workshop_Data\RB_PurposiveSampling_Heshan\GISData) directory. Those layers are the same as the layers you used in sampling design.

Load field sample table by clicking “Load” button (\Workshop_Data\RB_PurposiveSampling_Heshan\Actual Samples.csv).

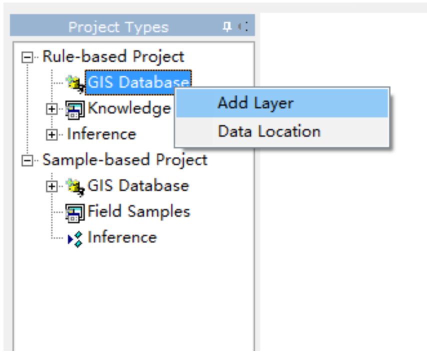
Specify the output knowledge base file: (\SoLIM_Result\Purposesample \Knowledge_from_purposive_sample.xml).

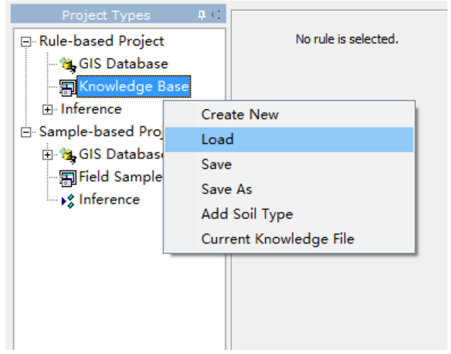


Click "OK". A knowledge base file will be created.

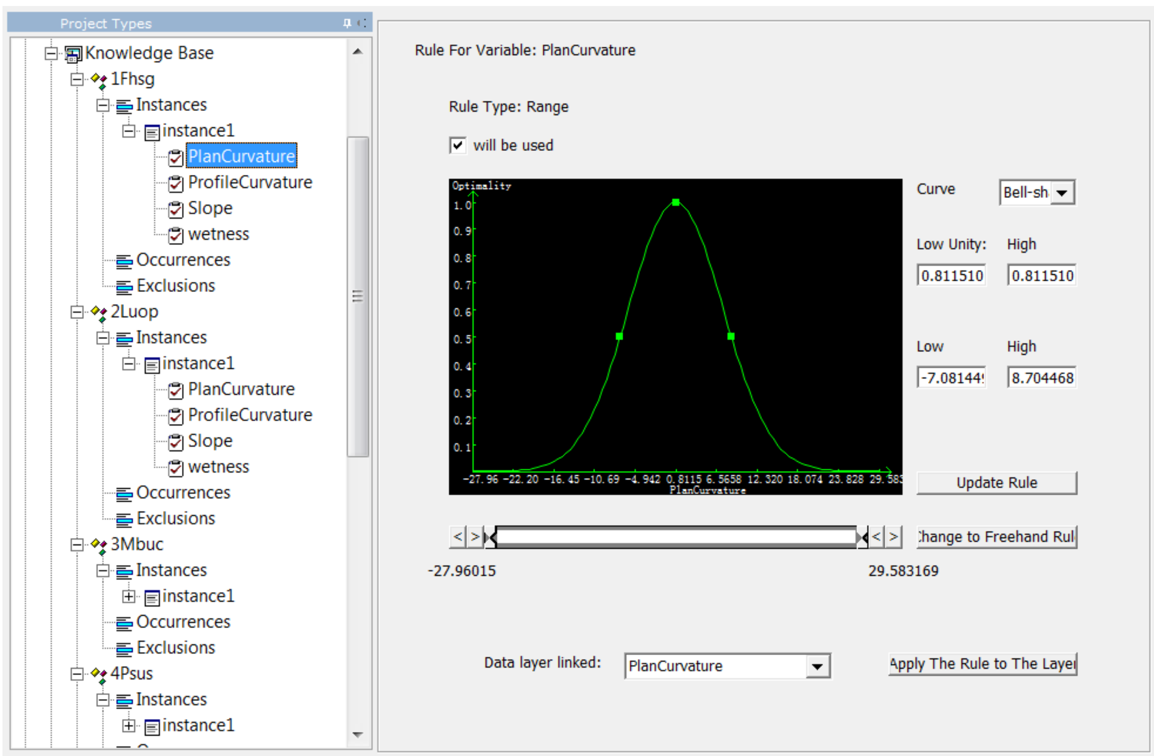
Step 4: Use knowledge base file in SoLIM Solutions:

After you obtain a knowledge base file from purposive sampling, you can use it for soil mapping. The steps are: 1) setup a project (rule-based); 2) load environmental data; 3) load the knowledge base created from the purposive sampling; 4) conduct soil inference. Details on the first two steps were described in the training case described in 1.2.1 (Soil Mapping Based on Knowledge from Expert). ***Due to a program issues you must display every single environmental data layers on the screen to avoid program crashes in the later steps.*** The details on the step 3 (load the knowledge base) are described below. First, right click on "Knowledge Base" node in the left panel. Then, choose "Load" in the menu and load the knowledge base file.





If it is loaded successfully, you will see the rule definition for each soil type that appears in your field sample table.



The soil inference step is the same as the above case training.

2.3 Sample-based Soil Mapping

- **Case Study:**

Pleasant Valley, Wisconsin, USA

- **Data:**

Field samples (split into training and validation set)

Environmental data (Parent material, Elevation, Slope gradient, Profile curvature, Topographic wetness index)

- **Learning Goal:**

Soil mapping based on ad-hoc samples

Property validation

- **References:**

Liu, J., A.X. Zhu, S. Zhang, C. Qin, 2013. "Large-scaled Soil Attribute Mapping Method Based on Individual Representativeness of Sample Sites", *ACTA PEDOLOGICA SINICA*(in Chinese), 50(1): 12-20.

Zhang, S., A.X. Zhu, J. Liu, L. Yang, 2012. "Sample-based Digital Soil Mapping Methods and Related Sampling Schemes". Soils (in Chinese). 12:881-889.

- **Operation Procedure:**

Let's assume that we want to map soil sand content over an area. In this exercise, we do not have direct knowledge on the relationship between soil sand content and environmental conditions. Instead, we have some field samples where soil sand content of a given horizon has been measured. However, these samples were not collected based on a well-established sampling design (e.g. random sampling, regular sampling, etc.). SoLIM Solutions 2015 provides a way (referred as Sample-based Inference) to map the spatial variation of soil by just using such samples. Certainly you can use this approach with samples collected according to a well-established sampling design.

Suppose that the following table contains the information of the field samples you have for an area. The "X" and "Y" columns contain the x and y coordinates of each sample location. These coordinates should be in the same coordinate system as that of the environmental data layers. Column "Sand" contains the soil sand content of a given horizon for each sample location.

	A	B	C	D
1	SampleID	X	Y	Sand
2	6	712053.4	490942.2	85
3	10	711597.4	491446.6	47
4	13	712003.4	490352.2	3.2
5	14	711889	490101.7	18.8
6	24	707233.4	494502.2	47
7	27	708145.8	494860.6	13.4
8	33	711553.4	489622.2	7.8
9	36	711239.6	489710.5	27.3
10	37	710697.6	489697.9	22.8
11	40	711260.4	493236.2	56.2
12	42	711493.4	492922.2	19.3
13	45	711028.4	493811.6	76.2
14	46	711224.4	494142.7	7.7
15	47	711274.5	494443.6	11.5
16	52	708055.4	493405.6	21.8
17	54	707941.4	492985.6	12
18	53	707923.4	493192.2	35.3

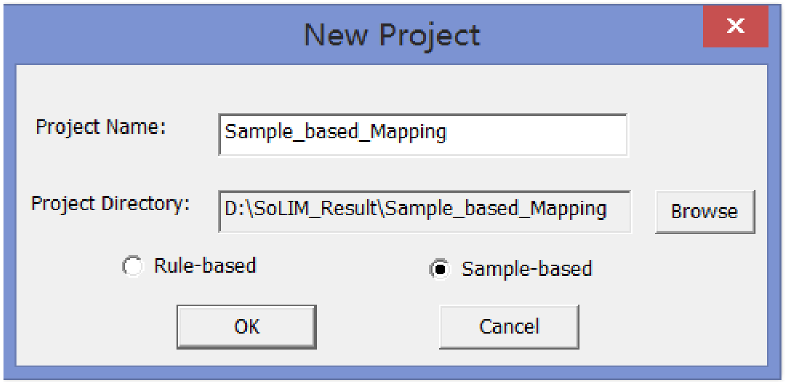
We will create a sample-based project to utilize these field samples to produce a soil sand map for the area.

Step 0: Prepare the sample file (done in this exercise)

We need to first create a .csv file to store the field samples. The easiest way to do this is to enter the field sample data into a spreadsheet and save it as a .csv file. In this exercise, this step has been done for you. A file named as "field_samples_training.csv" can be found in the directory "\Workshop_Data\SB_AdhocSample_PleasantValley\AdhocSamples"

Step 1: Create a Blank Sample-based Project:

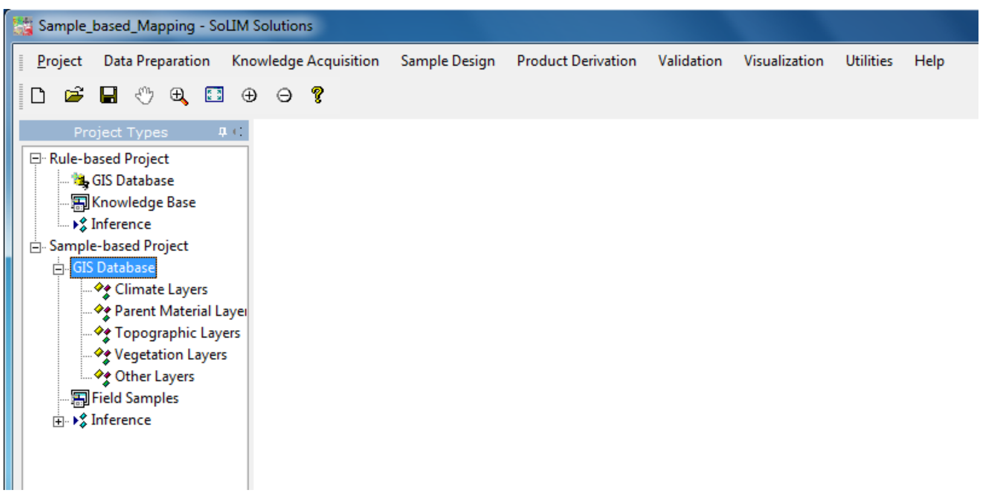
On the main menu of SoLIM Solutions 2015, select "Project->New" to create a new project. Specify Project Name as "Sample_based_Mapping" and create a new folder "Sample_based_Mapping" in the directory "D:\SoLIM_Result\". Use this folder as the project directory.



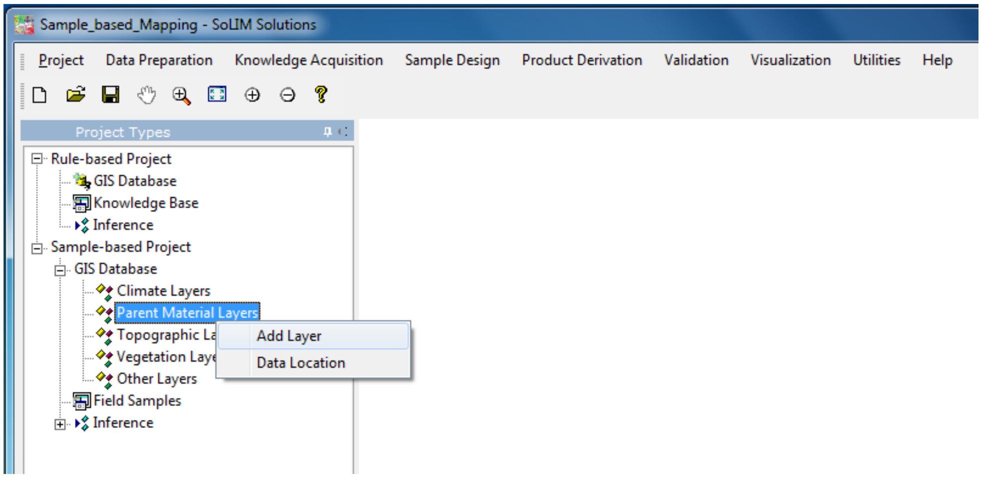
Select "Sample-based" and click on "OK", a new blank sample-based project will be created. You may notice that the "Sample-based Project" node in the left project panel has been activated.

Step 2: Add GIS Data Layers

In the left project panel, you will see five sub-nodes under the "GIS Database" node: "Climate Layers", "Parent Material Layers", "Topographic Layers", "Vegetation Layers" and "Other Layers". We will add the environmental data layers to different sub-nodes.

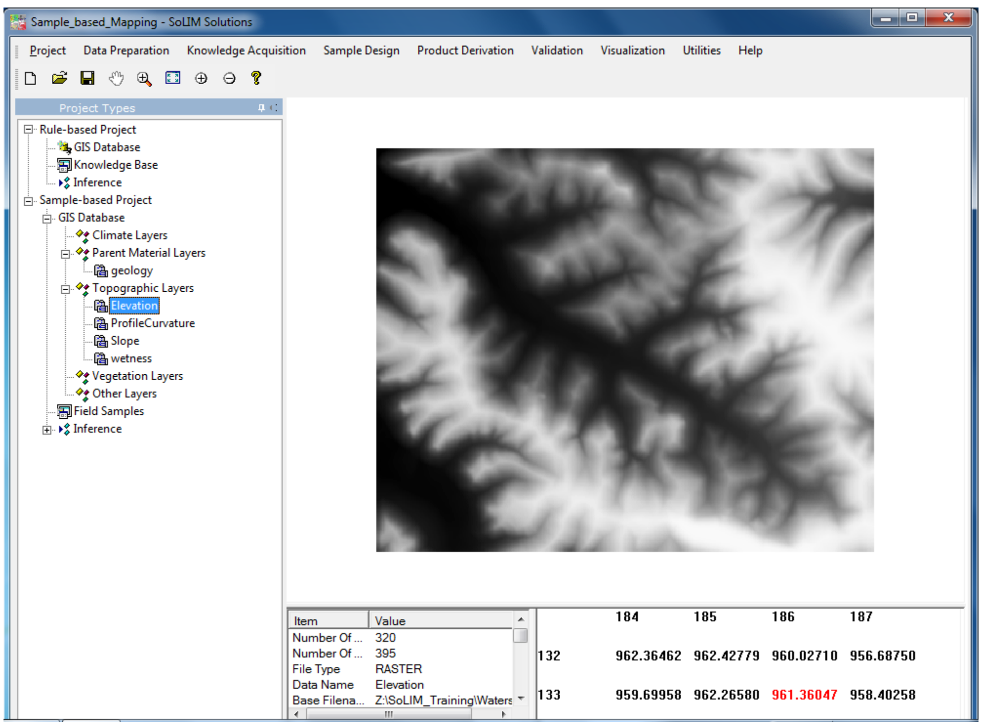


In "\Workshop_Data\SB_AdhocSample_PleasantValley\GISData" directory, you can find several environmental data layers. One of them ("geology.3dr") is for parent material. Other data layers are topographic layers. Right click on "Parent Material Layers" node, select "Add Layer" and browse to "geology.3dr".



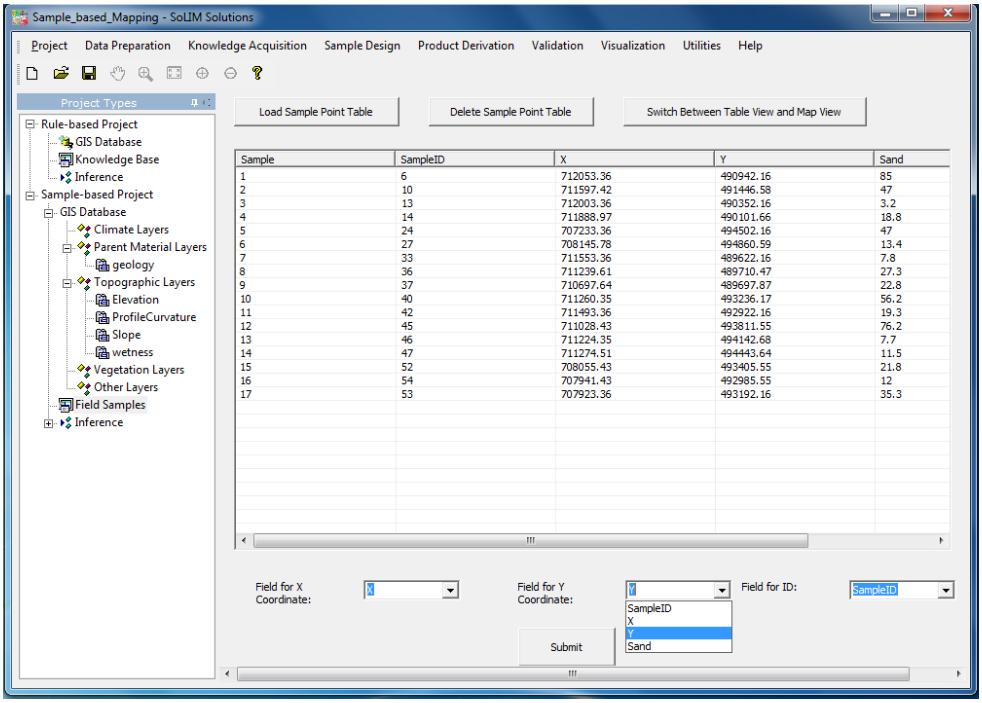
Similarly, add other layers to the "Topographic Layers" node.

After loading environmental data layers, you can view each layer by clicking on the corresponding layer name.



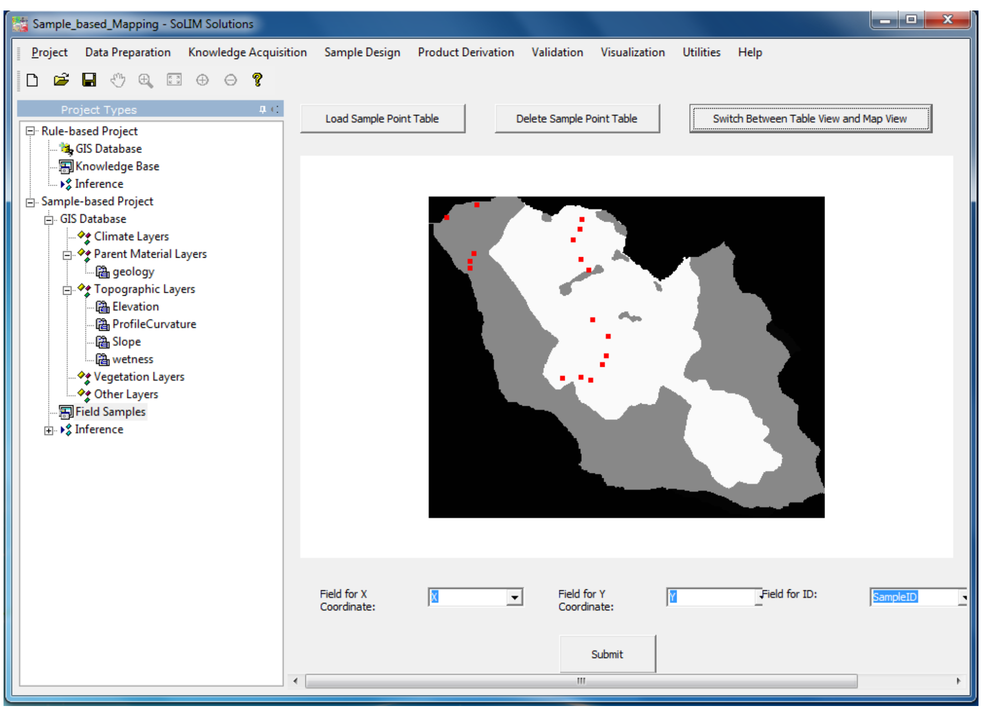
Step 3: Add the samples

Left click on the "Field Samples" node. You may find that the panel on the right side will switch to a blank table correspondingly. Press the "Load Sample Point Table" button on the top. As mentioned in Step 0, you may use the sample file "field_samples_training.csv" in the directory "\Workshop_Data\SB_AdhocSample_PleasantValley\AdhocSamples\".



Specify the columns that record x coordinate and y coordinate by selecting column names from dropdown list at the bottom. Here “X” is for x coordinate and “Y” is for y coordinate.

Click on “Submit”, you will see the spatial distribution of the samples in the map view. The red points represent the field samples in the study area.

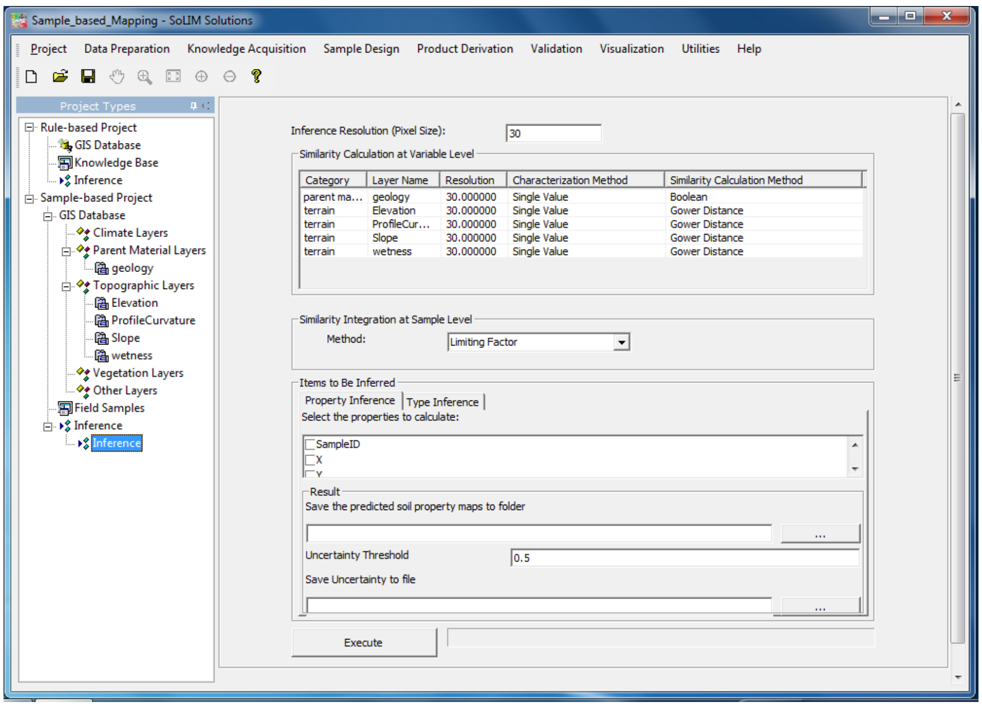


You can switch back to the table view by clicking on “Switch between table view and map view”.

Step 4: Run Inference

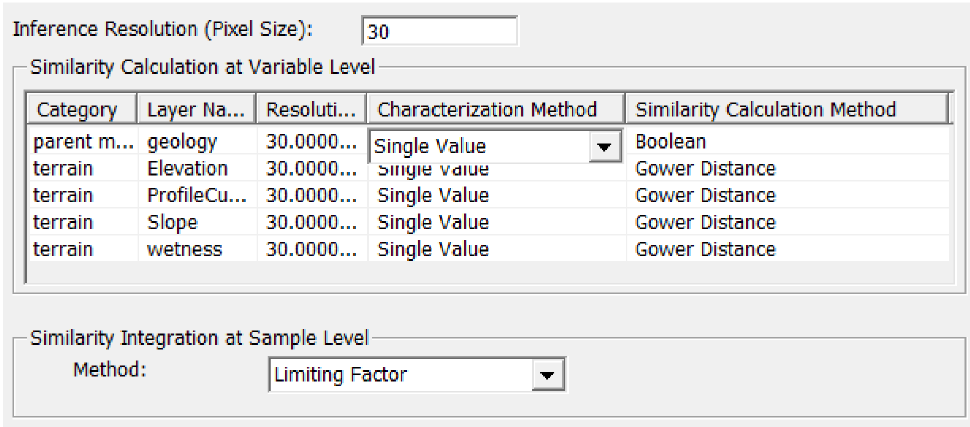
SoLIM Solutions can generate multiple soil property maps using GIS database and field sample points. In this exercise, we only infer soil sand content.

Click on “Inference” node to unfold it. Under that node, click “Inference”. The panel on the right side will switch to the inference interface.

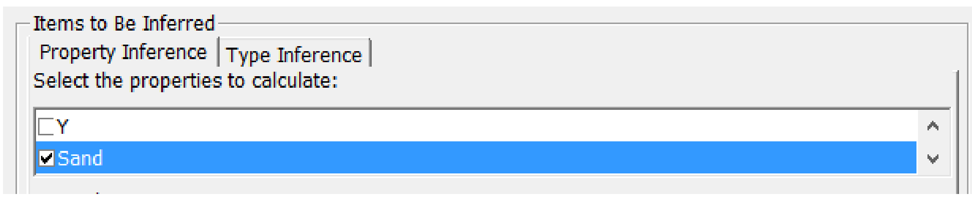


The basic idea of sample-based project is to infer soil property/type according to the environmental similarity between the position to be inferred and the existing field samples. In the inference interface, we need to specify the methods used to characterize environmental conditions and the methods used for similarity calculation.

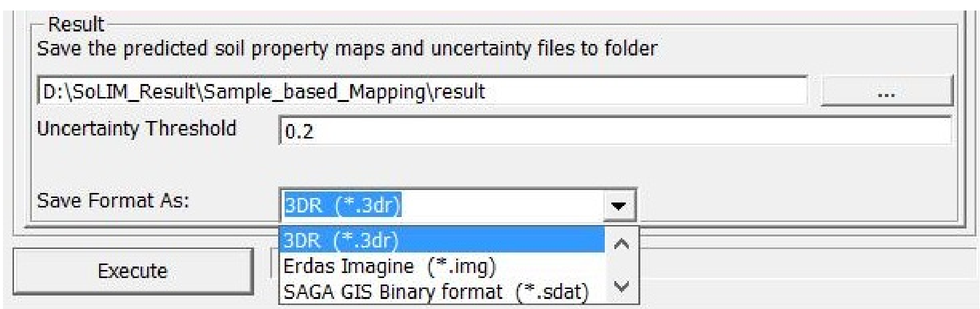
In this exercise, we use “Single Value” as the characterization method for all environmental layers. “Single Value” is the mean value of a given environmental variable over the inference resolution (e.g. 30m in this exercise). Similarity calculation is conducted at two levels: variable level and sample level. In this exercise, we choose “Boolean” as the similarity calculation method for parent material variable and “Gower Distance” for all terrain variables; we choose “Limiting Factor” as the method to integrate variable-level similarity into sample-level similarity.



Then we need to specify the soil property to be inferred. Make sure you are in Property Inference tab. Check the box before "Sand".



Click on button and specify the output directory. In the pop-up dialog, you may create a new directory named as "result" in the "Sample_based_Mapping" directory to host the output soil sand content map and uncertainty map.



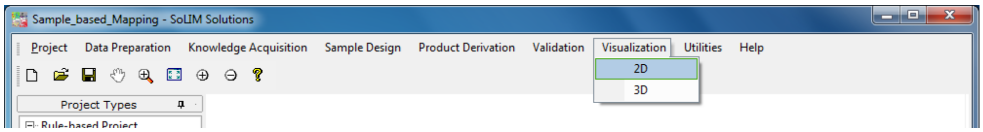
SoLIM Solutions can also provide uncertainty measurement associated with prediction at each location. If the prediction uncertainty at a position is higher than an uncertainty threshold, SoLIM Solutions will assign NoData to that position in the predicted soil property map. In this exercise, we change the default uncertainty threshold to a stricter one: 0.2.Specify the save format in the dropdown list (e.g. ".3dr").

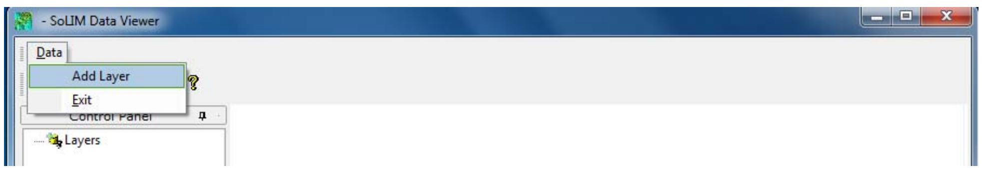
Click on "Execute" to run the inference. If the inference has been executed without error, a dialog will pop up to inform the completion of the inference. Otherwise, an error message will show up to tell what was wrong.

Step 5: Viewing Results

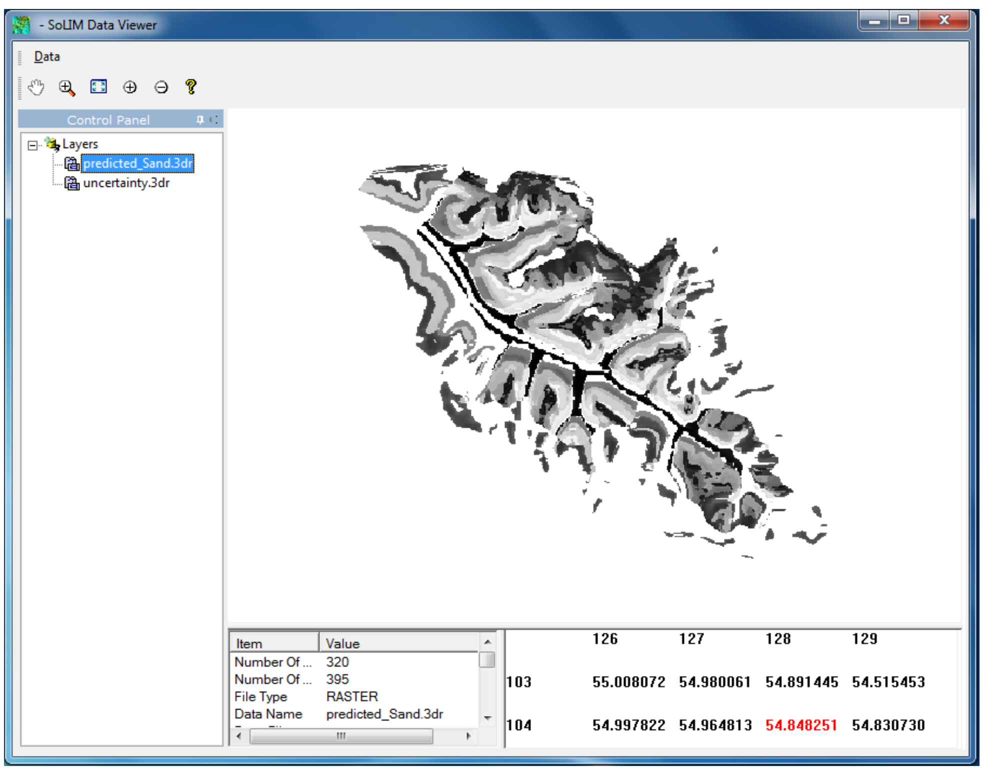
You should be able to find “predicted_Sand.3dr” file and “uncertainty.3dr” file in “\SoLIM_Result\Sample_based_Mapping\result” directory if the inference has been executed successfully. You may view the results by using the Data Viewer in SoLIM Solutions 2015.

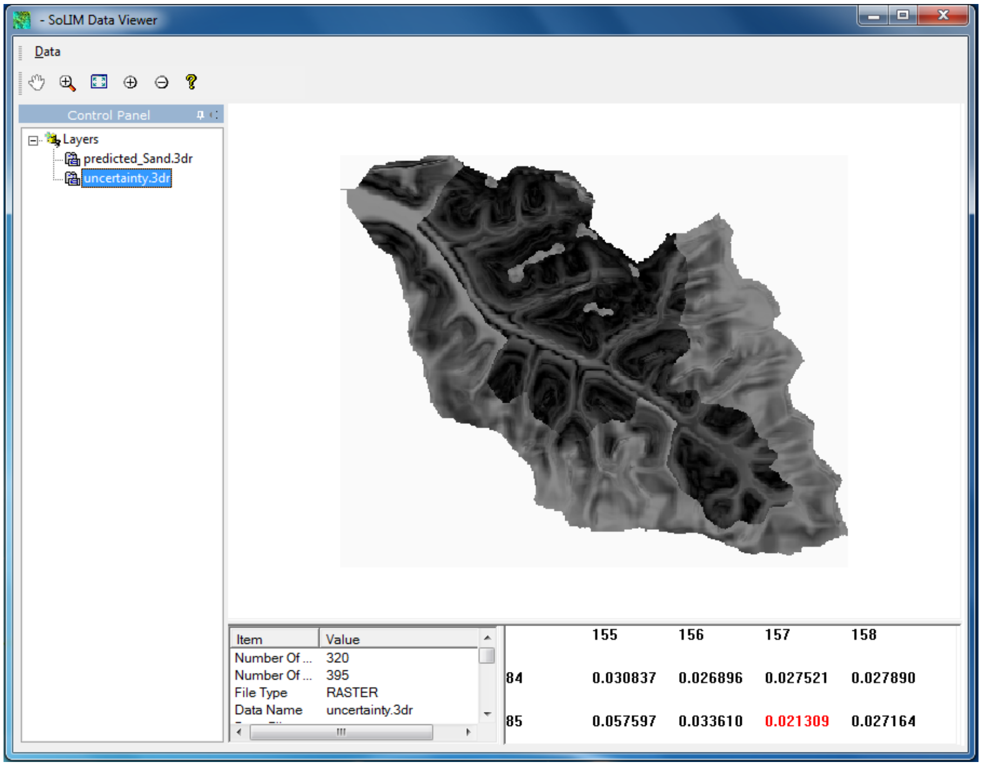
Choose “Visualization->2D” on the main menu of SoLIM Solutions 2015 and open the Data Viewer. In SoLIM Data Viewer, choose “Data->Add Layer”





Select “predicted_Sand.3dr” and “uncertainty.3dr” in your output directory and add them to the layer list panel on the left side. You can choose one layer to view each time.



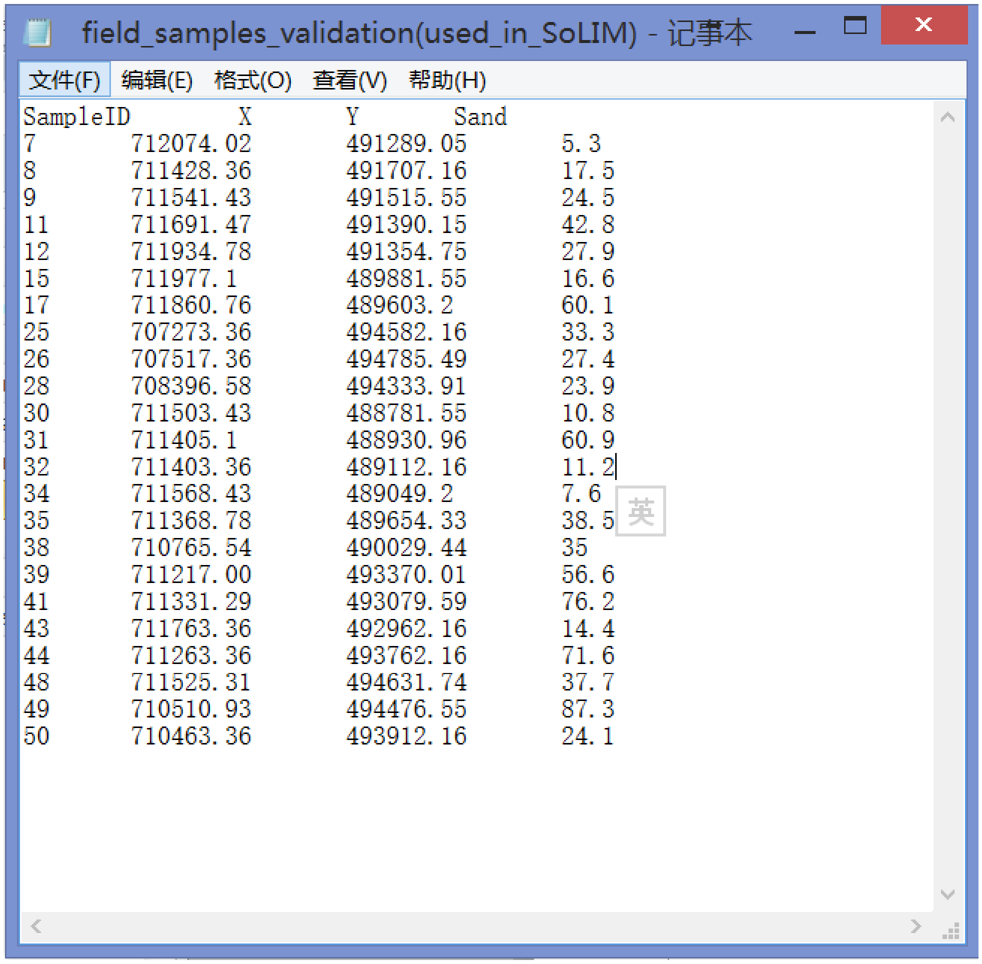


For the soil sand content layer, bright grey represents high sand content and dark grey represents low sand content. Note that some pixels have a negative value -0.015 which means NoData. This is due to the corresponding uncertainty values at those locations are higher than the uncertainty threshold (0.2 in this exercise).

For the uncertainty layer, bright grey represents high uncertainty and dark grey represents low uncertainty.

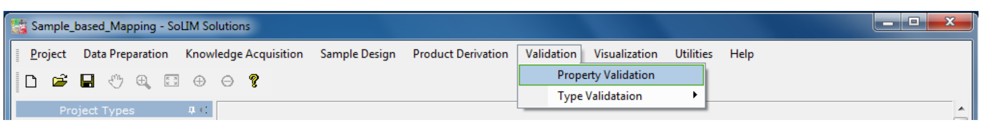
Step 6: Validation

In this exercise, a set of independent validation samples has been prepared for validation. Those samples are stored in a text file named as "field_samples_validation (used_in_SoLIM).txt".



SampleID	X	Y	Sand
7	712074.02	491289.05	5.3
8	711428.36	491707.16	17.5
9	711541.43	491515.55	24.5
11	711691.47	491390.15	42.8
12	711934.78	491354.75	27.9
15	711977.1	489881.55	16.6
17	711860.76	489603.2	60.1
25	707273.36	494582.16	33.3
26	707517.36	494785.49	27.4
28	708396.58	494333.91	23.9
30	711503.43	488781.55	10.8
31	711405.1	488930.96	60.9
32	711403.36	489112.16	11.2
34	711568.43	489049.2	7.6
35	711368.78	489654.33	38.5
38	710765.54	490029.44	35
39	711217.00	493370.01	56.6
41	711331.29	493079.59	76.2
43	711763.36	492962.16	14.4
44	711263.36	493762.16	71.6
48	711525.31	494631.74	37.7
49	710510.93	494476.55	87.3
50	710463.36	493912.16	24.1

On the main menu of SoLIM Solutions, choose "Validation" -> "Property validation"

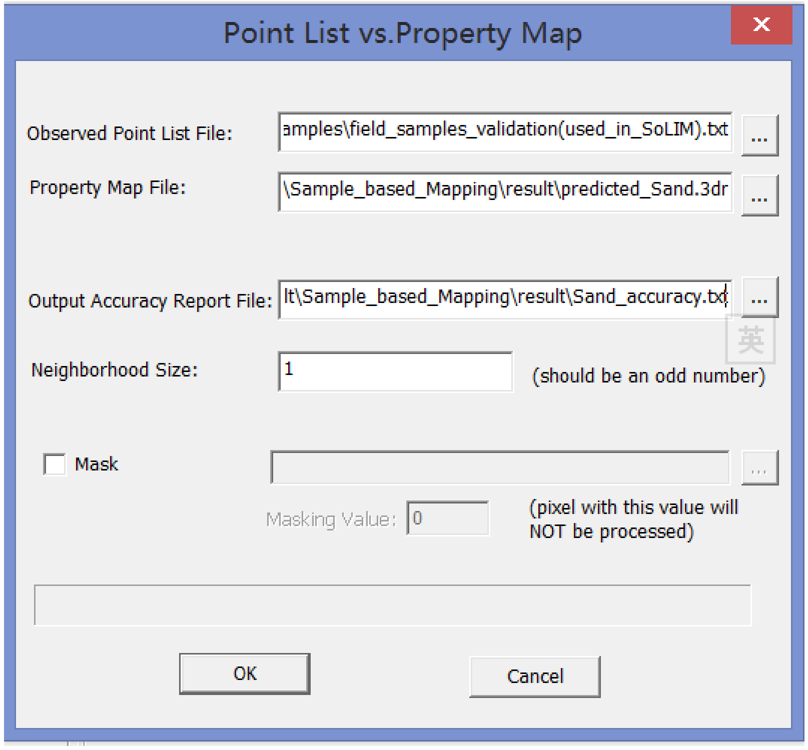


In the pop-up dialog window, specify the "Observed Point List File" by browsing to the directory "\Workshop_Data\SB_AdhocSample_PleasantValley\AdhocSamples\field_samples_validation(used_in_SoLIM).txt"

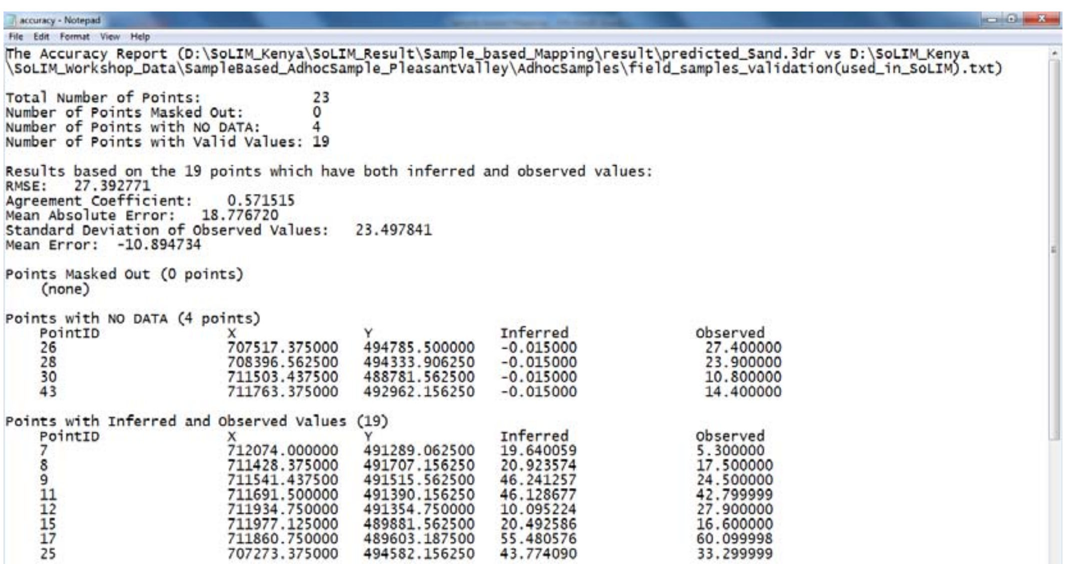
For "Property Map File", browse to the predicted soil sand content map ("predicted_Sand.3dr"). Specify the output location for the Accuracy Report File as "\SoLIM_Result\Sample_based_Mapping\result\Sand_accuracy.txt"

Leave the value of "Neighborhood Size" as default. And we do not need "Mask" in this exercise.

Click on "OK" to generate the accuracy report.



A summary of the validation can be found in the report. It should be noted that not all the validation samples are used – some of them are located at the positions where the prediction uncertainty is higher than the uncertainty threshold, therefore no prediction has been made at those positions. Multiple indices such as RMSE, AC, MAE, etc. are used to evaluate the prediction accuracy. The inferred value and the observed value of each validation sample are also listed.



```
The Accuracy Report (D:\SoLIM_Kenya\SoLIM_Result\Sample_based_Mapping\result\predicted_Sand.3dr vs D:\SoLIM_Kenya
\SoLIM_Workshop_Data\SampleBased_AdhocSample_PleasantValley\AdhocSamples\field_samples_validation(used_in_SoLIM).txt)

Total Number of Points:           23
Number of Points Masked Out:      0
Number of Points with NO DATA:    4
Number of Points with Valid Values: 19

Results based on the 19 points which have both inferred and observed values:
RMSE:   27.392771
Agreement Coefficient:    0.571515
Mean Absolute Error:   18.776720
Standard Deviation of Observed Values:   23.497841
Mean Error:  -10.894734

Points Masked Out (0 points)
    (none)

Points with NO DATA (4 points)
    PointID              X              Y              Inferred           Observed
    26                   707517.375000  494785.500000  -0.015000           27.400000
    28                   708396.562500  494333.906250  -0.015000           23.900000
    30                   711503.437500  488781.562500  -0.015000           10.800000
    43                   711763.375000  492962.156250  -0.015000           14.400000

Points with Inferred and Observed Values (19)
    PointID              X              Y              Inferred           Observed
    7                    712074.000000  491289.062500  19.640059          5.300000
    8                    711428.375000  491707.156250  20.923574          17.500000
    9                    711541.437500  491515.562500  46.241257          24.500000
    11                   711691.500000  491390.156250  46.128677          42.799999
    12                   711934.750000  491354.750000  10.095224          27.900000
    15                   711977.125000  489881.562500  20.492586          16.600000
    17                   711860.750000  489603.187500  55.480576          60.099998
    25                   707273.375000  494582.156250  43.774090          33.299999
```