SoLIM Solutions Help Manual

For SoLIM Solution 2016

SoLIM Group

2016

State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences

School of Geography, Nanjing Normal University

Department of Geography, University of Wisconsin-Maison

Table of Contents

Welcome to SoLIM Solutions

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 the Third Law of Geography, recent developments in Geographic Information Sciences (GISc), Artificial Intelligence (AI), and information representation theory. SoLIM was designed to overcome th
e limitations of existing soil survey methods and to improve the efficiency and accuracy of soil survey. Although it was designed for soil mapping, it can be also used for spatial prediction such as landslide susceptibility mapping, wildlife habitat mapping, even epidemic disease prediction and crime mapping.

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.

The SoLIM Solutions software was developed through a joint effort among the colleagues at the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Nanjing Normal University and The University of Wisconsin-Madison, led by Dr. A-Xing Zhu, Dr. Cheng-zhi Qin and Dr. James E. Burt. The other key developers involved include (in alphabetical order of last name): Fei Du, Peng Liang, Rongxun Wang, Lei Zhang, Fanghe Zhao.

SoLIM Solutions is currently distributed for free given that its use is non-commercial. Materials included with SoLIM Solution should not be copied without explicit citation of this manual. For updates and new release, please visit the SoLIM website: http://solim.geography.wisc.edu/index.htm

Where to start: This help manual is organized in four sections: Soil mapping using SoLIM, Quick start tutorial, Reference manual, and Appendices. The first section "Soil mapping using SoLIM" provides a brief overview about how to use SoLIM for soil mapping. The second section provides a step by step tutorial in using the SoLIM Solutions software through a hypothetical example. Section 3 provides a complete description of the functionality available through the SoLIM Solutions software. Section 4 contains a list of known problems and their solutions. References, tutorial data and other related materials are provided in Appendices.

If you are new to SoLIM, please start by going through Quick Start Tutorial section.

To learn how to do specific tasks, please consult Reference Manual. All sections in the reference manual have extensive links to the other relevant sections so there is no preferred starting point.

SoLIM Group[1,2,3]

1 Stat Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences
11A Datun Road, Chaoyang District, Beijing 100101, PR China

2 School of Geography, Nanjing Normal University
No.1, Wenyuan Road, Xianlin University District, Nanjing 210023, China

3 Department of Geography, University of Wisconsin-Maison
550 N. Park St., Madison, WI 53706, USA

Contact Information:

Email: dsm_solim@yahoo.com

Twitter: https://twitter.com/DsmSolim

Website: http://solim.geography.wisc.edu/index.htm

1. Soil Mapping Using SoLIM

Overview

This section provides a brief introduction to soil mapping using the SoLIM approach. A list of key references on SoLIM is provided in Appendix D. SoLIM Solutions supports two types of soil mapping: **rule-based mapping** and **sample-based mapping.** Both of them are based on the idea that soil types and/or soil properties can be inferred from soil-related environmental conditions (environment covariates). Thus, both require a set of environmental layers (or covariates) (stored in a **GIS Database**) which depict the environmental conditions indicative of soil conditions. The difference between rule-based and sample-based is that in rule-base project (which is based on fuzzy logic and expert system), users need to define a set of rules (**Knowledge Base**) describing the soil-environment relationship explicitly, while in sample-base project (which is based on the Third Law of Geography), users need only to provide field sample points (**Field Samples**). The sample points can be those collected either based on a well-designed sampling strategy or based on ad-hoc sampling activities (meaning that these samples are not collected based any specific sampling design at all).

Rule-Based Soil Mapping

There are three key ideas underlining the rule-based soil mapping. **The first** is that SoLIM maps soil type (taxonomic class or user-defined soil concepts), not soil mapping units, under the assumption that soil properties are fairly homogeneous over a small spatial extent (such as 10 meters by 10 meters area). Thus, it takes a raster-based approach which means that it divides the area to be mapped into small pixels and determines the soil type for each pixel. **The second** idea is that the soil at each pixel is expressed in terms of its similarity to a set of prescribed soil types (user defined categories, also referred to as prototypes in the literature). This idea is often referred to as "fuzzy soil mapping". The object of such mapping is to avoid assigning a single soil type to a given location, but instead to assign similarity values (fuzzy membership values) expressing the similarity of the local soil to each of the prescribed soil types (categories or prototypes). For this reason, the user must know the types of soil existing in the area to be mapped. **The third** idea is that it predicts the similarity value of a local soil to each prescribed soil type (category) by assessing the environmental conditions at the location according to knowledge on how these conditions are related to the development of each soil type. In other words, SoLIM takes a knowledge-based approach to predict the similarity values. The two key inputs to SoLIM are: data on the selected environmental variables (covariates) related to soil conditions in the area (stored in GIS database), and knowledge that describes the relationships between soils and the environmental variables (referred to as the soil-landscape model and stored in knowledge base).

SoLIM uses an inference engine to link the GIS database with the knowledge base to calculate the similarity values (See Appendix D). For example, a piece of soil-environment relationship knowledge could be "If the elevation is 1000 feet and slope is 12%, then the soil there is a **typical soil type A**". In this case, the inference engine will use the GIS database to identify all the locations where elevation values are 1000 feet and slope values are 12%, then assign full membership (similarity) to those locations as the soils at these locations are typical cases of soil type A. While in this example we used two environment variables for calculating the membership values, SoLIM can calculate membership values based as many variables as the user wants to use.

Not all locations in the area meet the conditions perfectly. For example, "soil type A occurs in areas with elevation from 500 ft to 1500 ft and slope from 6% to 20%". This does not mean that all places with this range of values will have the same soil (A in this case). Instead, SoLIM acknowledges that places within that range will be more or less similar to another soil type (B for example) depending on the particular values of the environmental variables. In addition, soils in areas just a bit outside of the range may still bear some similarity to soil type A. They will not be perfect examples of typical soil type A, but they will not be totally dissimilar to that type either. For these locations (actually they constitute majority of the landscape), SoLIM will assign partial membership values based on how similar the environmental conditions at other locations to the conditions stated above. SoLIM does this by adopting a rule which is expressed as a function that defines how changes in an environmental variable affect the optimality of that environment variable for a specific soil type. If a range of values of an environmental variable is less optimal for a soil type, i.e., the environment condition is "sub-optimal", then the optimality value would be low for that range of values. If the optimality value is high, on the other hand, we would expect the local soil under that environmental condition to have higher similarity to the typical soil of that type. In other words, an optimality function describes how the similarity to the typical case of a given soil type changes as the local environmental conditions deviate from the ideal conditions.

The inference engine then looks at the optimality values of all environment variables related to the soil and determines an overall measure indicating the similarity of the local soil to a named soil type. This procedure is repeated for all defined soil types, yielding a vector of similarity values for each pixel.

Soil-landscape model (knowledge on soil-environment relationships) can be obtained using different methods. Soil scientists may provide their knowledge in the form of optimality functions directly, or they may use some words to express the knowledge on the relationship between soil type and environmental conditions, or they may identify locations as places where the soil is typical for the given soil type using a topographic map, DEM, or orthophoto. Alternatively, soil-landscape model can be obtained through spatial data mining or purposive sampling techniques (Appendix D).

The basic output of SoLIM is a set of fuzzy membership maps, one for each soil type. If one wants to tag a pixel with a single soil for producing a soil type map, it would be natural to do so by selecting the class with the highest membership. This process is called **hardening**. It produces a single "best guess" soil type for each pixel. Hardening can be done through the **production derivation** menu of SoLIM Solutions. If one wants to estimate the value of a given property for each pixel for producing a soil property map, one can use the fuzzy membership values as weights through a weighted average approach. This can be achieved through the **Property Map** function of the production derivation menu.

Key Terminology and Concepts in Rule-based Soil Mapping

- It is a **raster-based approach** that assigns soils type to all pixels in a study area.
- SoLIM uses a **fuzzy representation** for soils. That is, SoLIM allows a pixel to have membership in multiple soil types. Each membership takes on a value between zero (no membership) and unity (full membership). We call these the **fuzzy membership values** or **similarity value** of a pixel. The membership values need NOT sum to unity.
- Some environmental settings are more conducive to a particular soil type than others. We define the **optimality** as an indicator of the degree to which a soil is favored by a particular value of an environmental variable. For example, if a soil is less likely to be developed on a slope of 30%, the optimality would be low for that slope value. Like fuzzy membership values, the optimality is a number between zero and unity.
- Knowledge on soil-environment relationships is the key to successful soil mapping. SoLIM can accept different kinds of knowledge and express them as **rules**. Each rule corresponds to one environmental variable and characterizes the relationship between the optimality of one soil type and that environmental variable. Knowledge are divided into two types: global knowledge and local knowledge.
- **Global knowledge** refers to the knowledge that is effective **in the whole mapping area**. In SoLIM, global knowledge is expressed as **instances**. Each instance is a representation of the soil scientist's knowledge on the relationship between a soil type and its environmental conditions. An instances contain one or more rules. One rule defines the relationship between one environmental variable and the soil type.

- **Local knowledge** refers to the knowledge that takes effect **within limited area**. In SoLIM, local knowledge is expressed as **occurrences and exclusions**. An occurrence is a positive exception, which means a particular soil type will occur in places where the global knowledge does not cover. An exclusion is a negative exception, which means a particular soil type will be very unlikely to occur in some places where the global knowledge does cover. An occurrence/exclusion contains one or more rules plus **spatial setting** which defines the area under influenced.
- The final membership value for each pixel of the mapping area is obtained by **linking the knowledge base (global knowledge and local knowledge) with the GIS database**.

Sample-Based (Point-Based) Soil Mapping

Different from rule-based soil mapping, sample-based soil mapping, also referred to as point-based soil mapping, relies on the field samples instead of explicit knowledge (rules). The key ideas underlying sample-base mapping is that each field sample reflects an underlying relation between soil and its relative environmental conditions, and this relation would recur over the space. It is assumed that locations with similar environmental conditions will have similar soil type/property. Therefore, each sample can be considered representative over locations with similar environmental conditions. That is, each sample has an **individual representativeness**. Moreover, the representativeness level of an individual sample for an unsampled location can be approximated by the environmental similarity between the sample location and the unsampled location. Based on this concept, the soil property value or soil type at unsampled locations can be predicted by referring to environmentally similar samples. Besides, the uncertainty introduced by the samples' representativeness can be quantified by analyzing the nature of environmental similarity values.

SoLIM uses inference engine to link the GIS database with the field samples to estimate soil properties or soil types. Soil information at unsampled locations can be predicted by referring to environmentally similar samples. Through comparing environmental condition of existing samples and that of unsampled locations, environmental similarities between them can be estimated. Soil information at unsampled locations then can be predicted through integrating environmental similarities and the attributes of the corresponding samples. The uncertainty of prediction at each location due to the limitation of samples' representativeness can be quantified through analyzing the environmental similarities.

When inferring soil property values for an unsampled location, the soil property of the unsampled location is determined by weighting the soil property values of all the environmentally similar samples and the similarity values to these samples. With soil type inference, for each soil type the field samples with that soil type are selected, the environmental similarities between an unsampled location and the locations of the selected samples are measured and then the maximum similarity among the computed similarities is assigned to the unsampled location for the soil type. Therefore, the final results are a set of similarity files, one for each soil type, which is the same as rule-based soil mapping.

Key Terminology and Concepts in Sample-based Soil Mapping

- It is a **raster-based approach** that assigns soil properties or soil types to all pixels in a study area.
- Soil property and type at unsampled locations are predicted through integrating **environmental similarities** and the attribute values of the corresponding samples.
- Each field sample has its own **individual representativeness**. The field samples do not need to be based on any sampling designs. You can use samples of any number with whatever spatial distribution as long as the coordinates of the sample location are determined accurately and the observations at that location are made correctly. The coordinates of sample locations should be converted into the same coordinate systems as the GIS data layers which should be in one coordinate system.
- The final output (soil property map or soil type similarity files) of the mapping area is obtained by **linking the field samples with the GIS database.** The basic idea is to infer soil property/type based on **similarity** between an unknown position and existing field samples in terms of their environmental conditions. Therefore, how environmental conditions are characterized and how similarity is calculated are most critical in the inference.
- There are two basic characterization methods in this implementation: **single value** and **probability density:** Single value means that the mean value of a given environmental variable over the inference resolution is used in calculation of similarity when the inference resolution (the pixel size) is larger than that of the environmental data layer. If the resolutions of the environmental data layers are the same and the inference resolution is also that of the environmental data layers, the single value is simply the original value of the pixel. Probability density means that a probability density function is derived from the values of the pixels over the area of inference pixel when the inference pixel size is greater than the pixel size of the environmental data layer. Probability density method is much more computationally expensive than the single value approach.
- Similarity calculation is conducted at two levels: **variable** level and **sample** level: At the variable level, the similarities between an unknown position and sample points are calculated using each environmental layer in GIS database. At the sample level, the similarities derived at the variable level are integrated to yield the final similarity between each unknown position and each sample point.

- For sample-based soil property inference, soil property of the unsampled location is inferred by integrating the soil property values of the **environmentally similar samples.**
- For sample-based soil type inference, SoLIM uses a **fuzzy representation** for soils types. That is, SoLIM allows a pixel to have membership in multiple soil types. Each membership takes on a value between zero (no membership) and unity (full membership). We call these the **fuzzy membership values** or **similarity value** of a pixel. The membership values need NOT sum to unity.
- For sample-based soil type inference, the basic output of SoLIM is a set of fuzzy membership maps, one for each soil type. If one wants to tag a pixel with a single soil type, it would be natural to do so by selecting the class with the highest membership. This process is called **hardening**. It produces a single "best guess" soil type for each pixel. One can use the functions in the Product Derivation menu to produce soil type map and/or property map based on the derived fuzzy membership maps.

Implementation

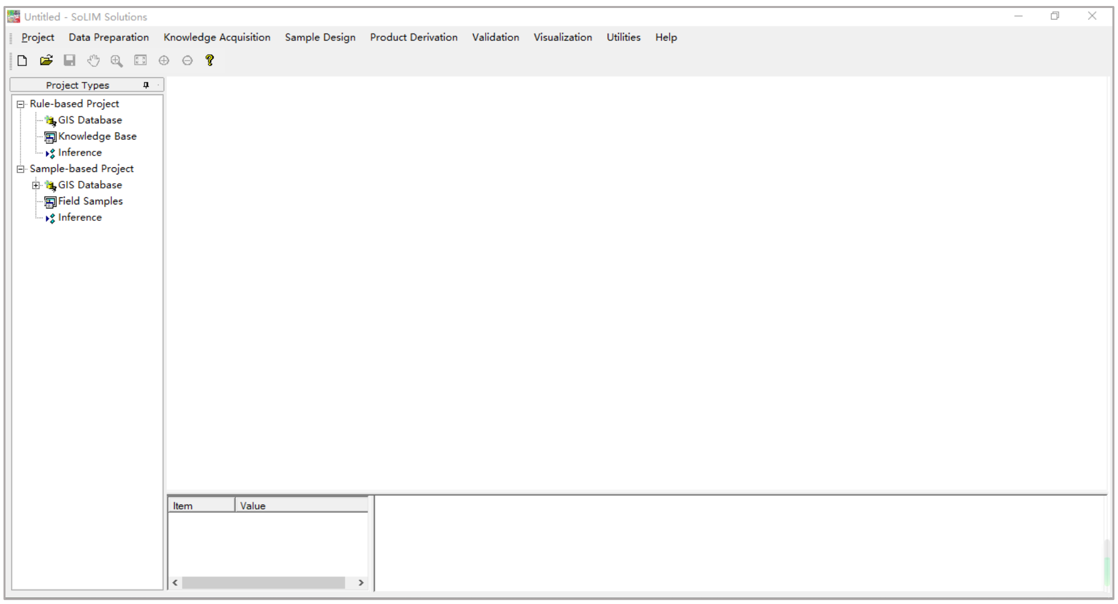
SoLIM Solution provides a graphic user interface to construct GIS database, to define the knowledge on soil-environment relationships and import field samples, and to utilize the SoLIM inference engines described above. In addition, it includes modules for users to prepare the environmental data layers (GIS Layers). The functions are described in detail in the [Reference Manual](#).

2. Quick Start Tutorial

2.1 Overview of SoLIM Solutions

The directory in which SoLIM Solutions 2015 is installed is referred to <SoLIM Solutions installed directory>. The data sets needed for the tutorial should have been installed in a directory named "Tutorial_Data" under <SoLIM Solutions installed directory>.

After starting SoLIM Solutions 2015 you will see the following user interface:



The Menu:

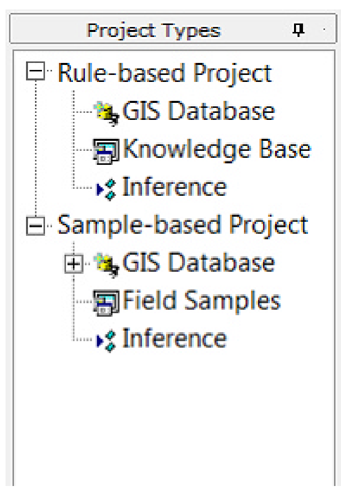
- The "**Project**" menu provides the interface for managing the projects. A project maintains all necessary information needed for performing a SoLIM inference (soil mapping). Project menu also supports importing existing projects created with the old versions of SoLIM software.
- The "**Data Preparation**" menu provides terrain analysis tools which implement the most recently developed ideas in digital terrain analysis and it also provides tools for extracting soil-related surface dynamic feedback from remote sensing data (Zhu et al., 2010).
- The "**Knowledge Acquisition**" menu provides tools to extract knowledge from experts or through purposive sampling or through Map.
- The "**Sample Design**" menu provides traditional sample design tools as well as purposive sampling design tool (Yang et al., 2012).

- The "**Production Derivation**" menu provides the interface to produce soil property maps derived using fuzzy soil membership maps (Property Map) and soil type maps (Hardened Map).
- The "**Validation**" menu provides tools to evaluate the results (type maps and property maps) using field point data including property validation and type validation.
- The "**Visualization**" menu provides tools to visualize environmental data layer and inference results in 2D or 3D form.
- The "**Utilities**" menu provides file conversion and other data analysis utilities.
- The **"Help"** menu provides SoLIM Solutions 2015 user manual for users to query and the software version information.

The Toolbar:

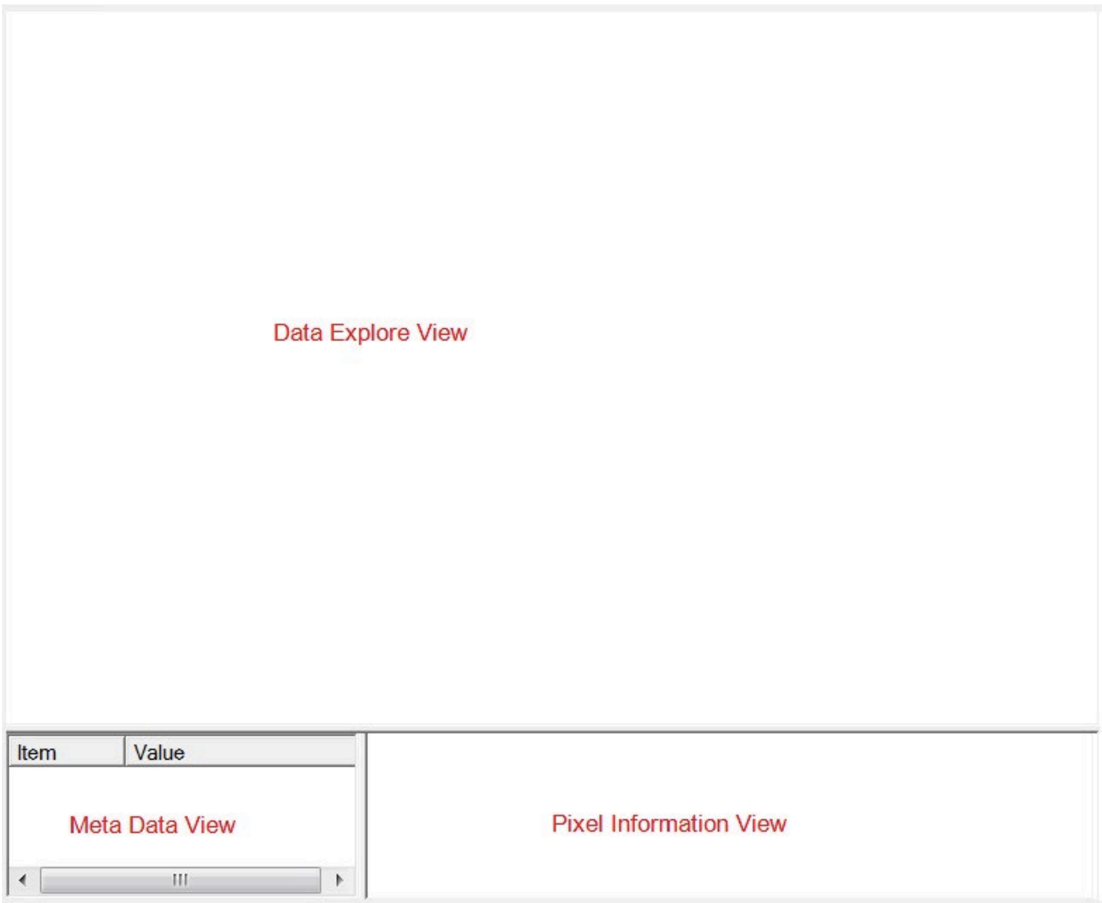
The toolbar provides some frequently used functions, including creating project, saving project and exploring environmental data layers.

The Project Panel:



The project panel contains a tree-style list. It is used to control SoLIM projects. As you can see from this panel, SoLIM Solutions 2015 supports two types of projects: rule-based and sample-based. Both rule-based project and sample-based project are based on the idea that soil types and/or soil properties can be inferred from soil-related environmental conditions, so they both require a set of GIS layers (GIS Database) which depict the soil-related environmental conditions. The difference between them is that in rule-base project, we need to define a set of rules (Knowledge Base) to express the soil-environment relationship explicitly, while in sample-base project, we just need to provide field sample points (Field Samples). As stated before the sample-based project does not need the samples based on any spatial sampling design. You will have a better understanding of the difference after you go through this tutorial. Currently, the two projects cannot be mixed.

Views:



The view area is for displaying current information based on the node selected in the project panel. The figure above shows the view for GIS database management.

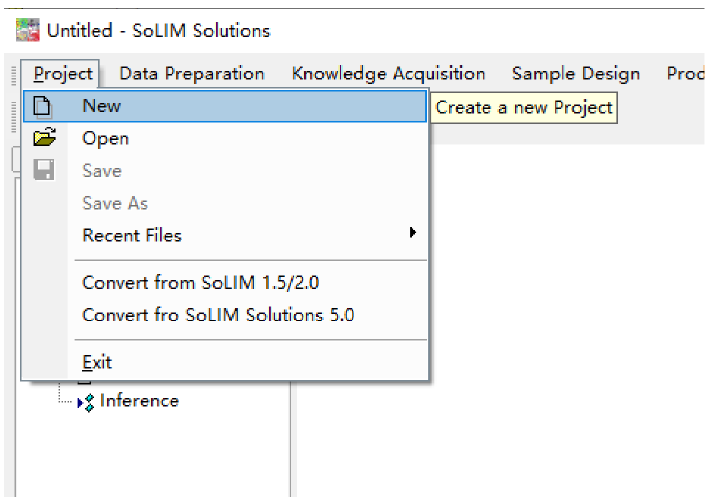
2.2 Rule-based Project

Clearly, with a rule-based project, one is assuming that one has the soil-environment relationships explicitly defined and they can be expressed as rules.

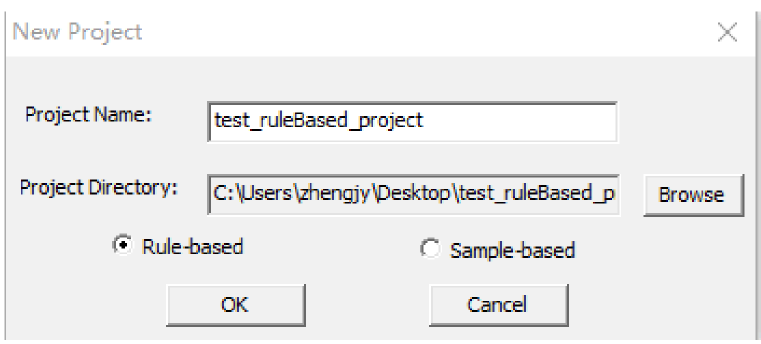
For this tutorial, let us assume that we want to map a soil type (say soil type 1) which occurs under the following conditions: **"South facing slopes with linear profile curvature, sandstone geology (indicated by the number 15 in the geology data layer) and slope gradient greater than 12%"**. This description of environmental conditions is the knowledge on soil-environment relationship.

Create a Blank Rule-based Project

Select "Project->New" to create a new project.



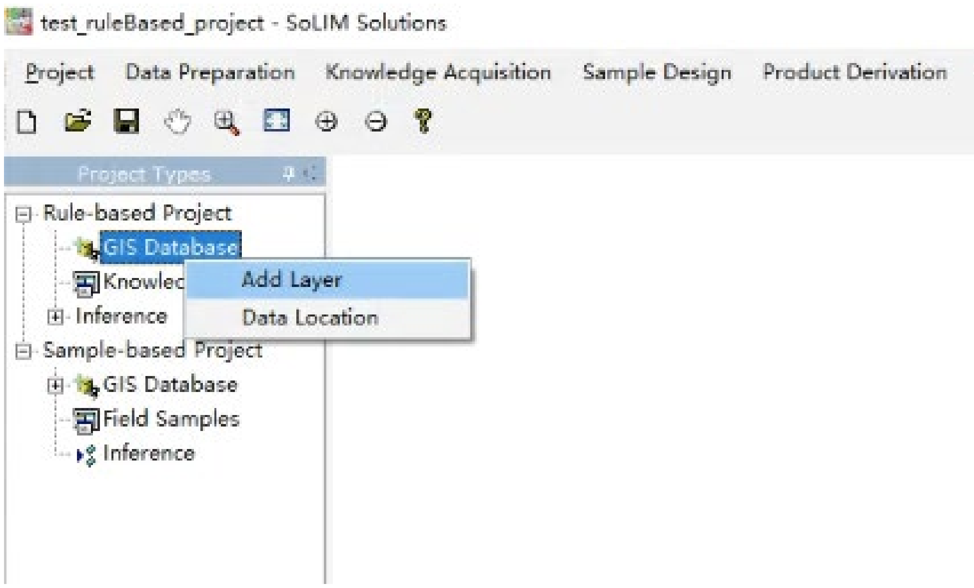
Specify Project Name and Project Directory. Here we enter "test_ruleBased_project" as the project name.



The default project type is rule-based, so we don't need to make any change. Click on "OK", a new blank rule-based project will be created for you. The Rule-based Project node in the left project panel is activated. In the project directory you just specified, the program generated test_ruleBased_project.sip (the project file) and a directory named "GISData" (directory for storing GIS data layers on environment variables) which is currently empty.

Add GIS Data Layers

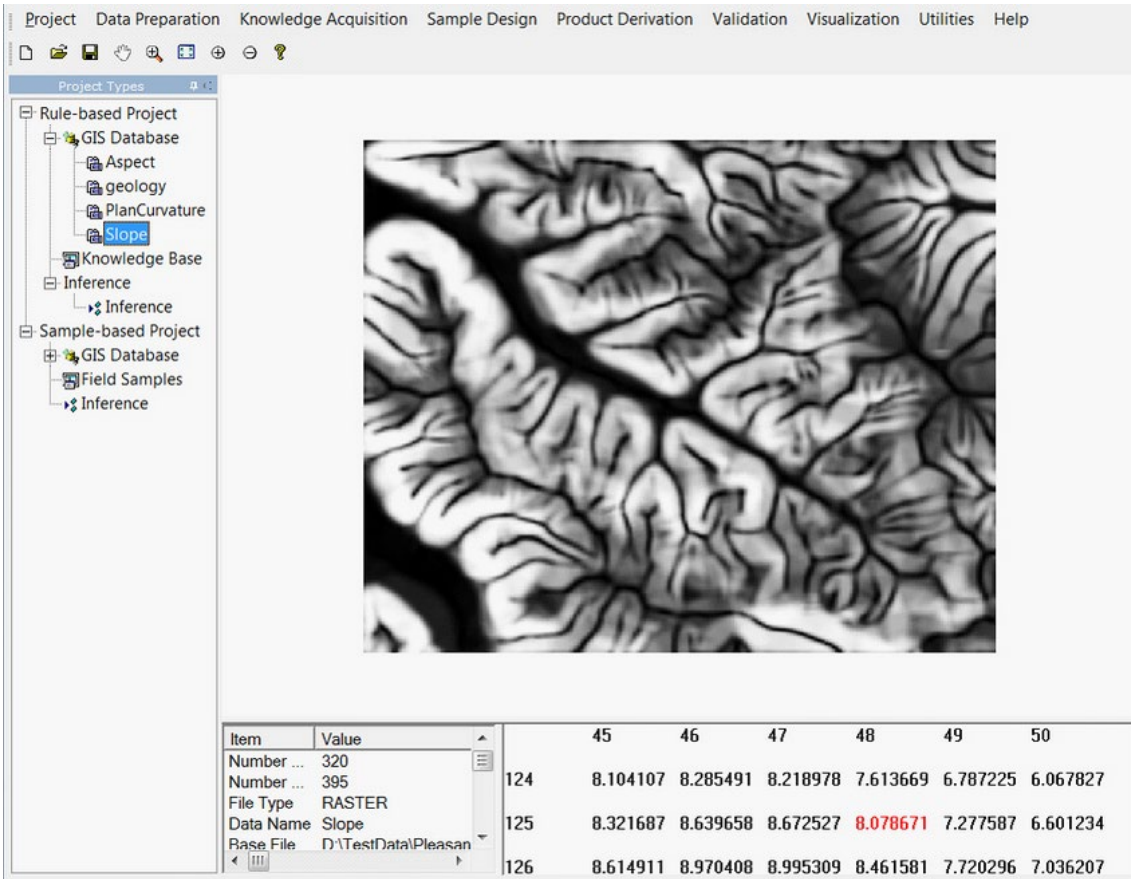
In the left project panel, right click on "GIS Database" under "Rule-based Project" node. In the pop-up menu, select "Add Layer" to add new environmental data layer(s) to the project.



For this tutorial we will add the following four data layers (in 3dr format) from "[SoLIM Solutions installed directory]\Tutorial_Data\Rule_Based" directory: Geology, ProfileCurvature, and Slope based on the needed environmental conditions stated in the knowledge stated earlier. You can also include other environmental data layers at this stage which may be needed for mapping other soil types. The native file format for SoLIM Solutions is 3dr format. However, SoLIM Solutions can convert other commonly-used raster file format (e.g. TIFF, ERDAS IMAGE) to .3dr format on the fly. SoLIM Solutions can also rasterize commonly-used vector file format (e.g Shapefile) into 3dr format. See File Converters in the Utilities menu for how.

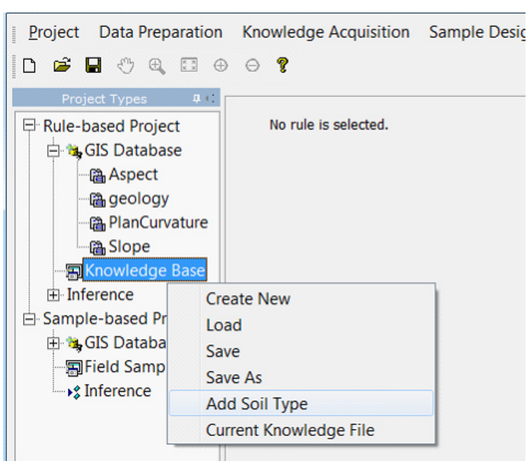
After selecting the files, click on "Open". The left project panel should now list the data layers you just selected. These data layers are also copied to the "GISData" directory in the project directory you specified earlier. In that way, they become part of the project and can be easily moved around with the project directory.

After the data layers are loaded, you can click the name of each layer in the left project panel to view the layer.

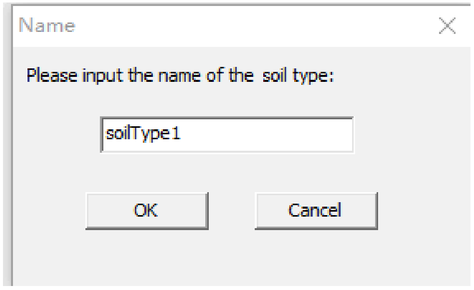


Add Soil Types

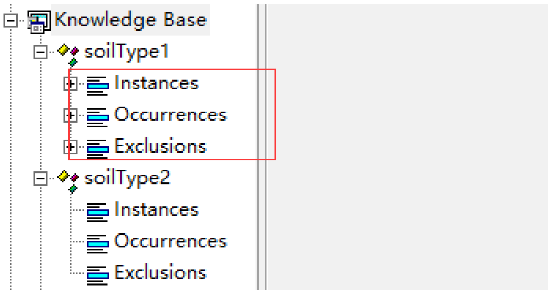
In the left project panel, right click on "Knowledge Base" node and select "Add Soil Type" from the pop-up menu.



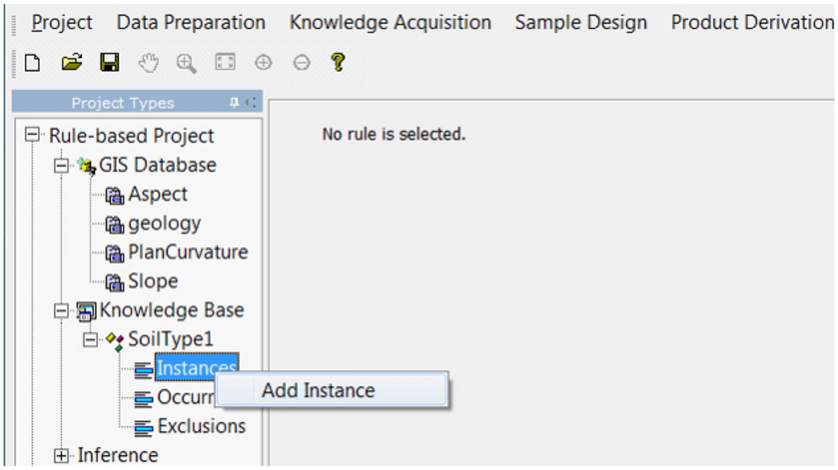
This displays a dialog box for you to specify the name of the soil type. Enter "SoilType1" and click on "OK".



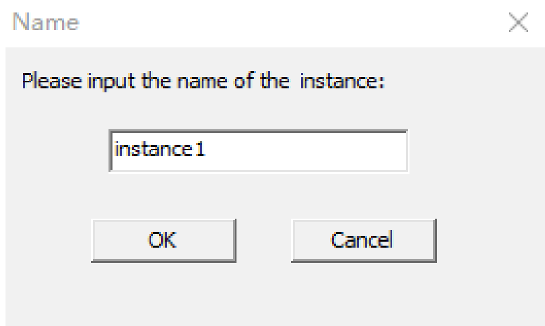
A new soil type is added into the knowledge base. Unfold the soil type node, you will see three sub-nodes are create: Instances, Occurrences, Exclusions. They are used to hold different kinds of knowledge (see Soil Mapping Using SoLIM for discussion on the different kinds of knowledge).



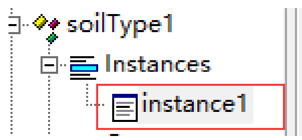
Based on the soil-environment knowledge for soil type 1 **("South facing slopes with linear planform curvature, sandstone geology (indicated by the number 15 in the geology data layer) and slope gradient greater than 12%".)** there is only one environmental configuration under which soil type 1 occurs. The environmental configuration takes effect in the whole mapping area, so only one instance is needed to represent the knowledge (global knowledge) in the knowledge base. Right click on the "Instances" node under the "SoilType1" node and select "Add Instance" from the pop-up menu.



This will display a dialog to let you input the instance name. Enter "Instance1" and click on "OK", a new blank instance will be created.



Under the "Instances" node, you can see the newly-created instance.



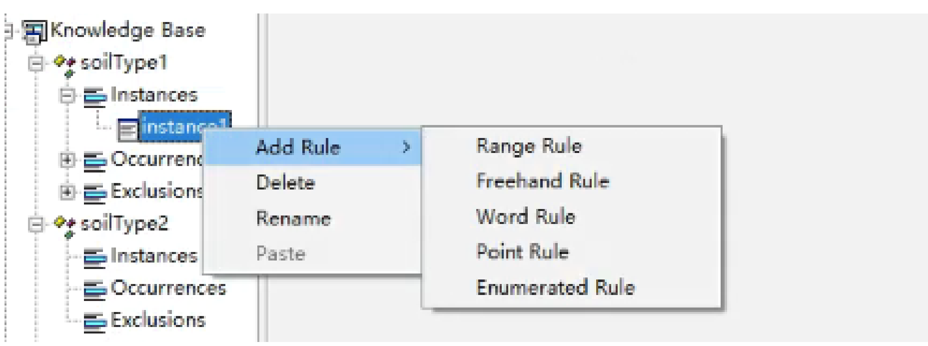
Create Rules

Four environmental variables are used in the soil-environment knowledge for this soil type. Thus, the next task is to create rules for each environmental variable.

Rule for Slope

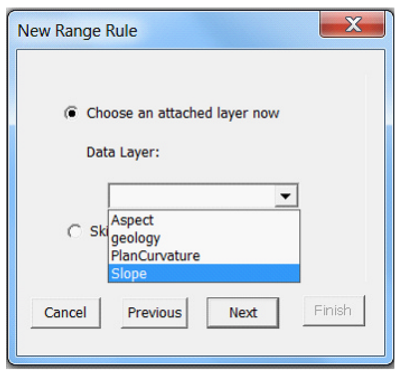
We can use **range rule** to express the knowledge on slope.

1. Right click on the "Instance1" node. In the pop-up menu, select "Add Rule" and then select "Range Rule".

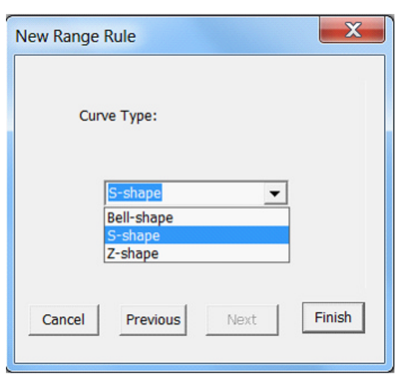


2. Enter the name of the variable for the rule. Here we type in "Slope" then click on "Next".

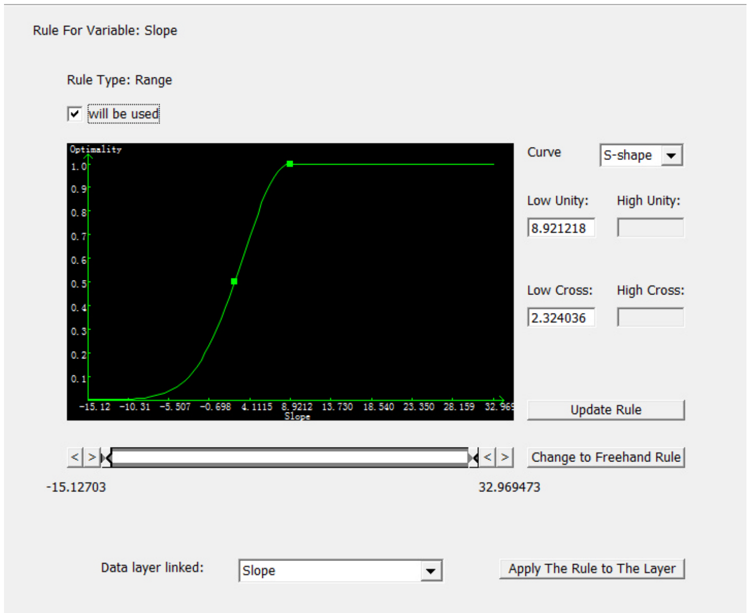
3. Select "Choose an attached layer now", and then select "Slope" from the "Data Layers" drop down list, and then click on "Next". This will allow the inference engine to link the rule defined here with the GIS data layer "Slope" which was defined in the GIS database earlier.



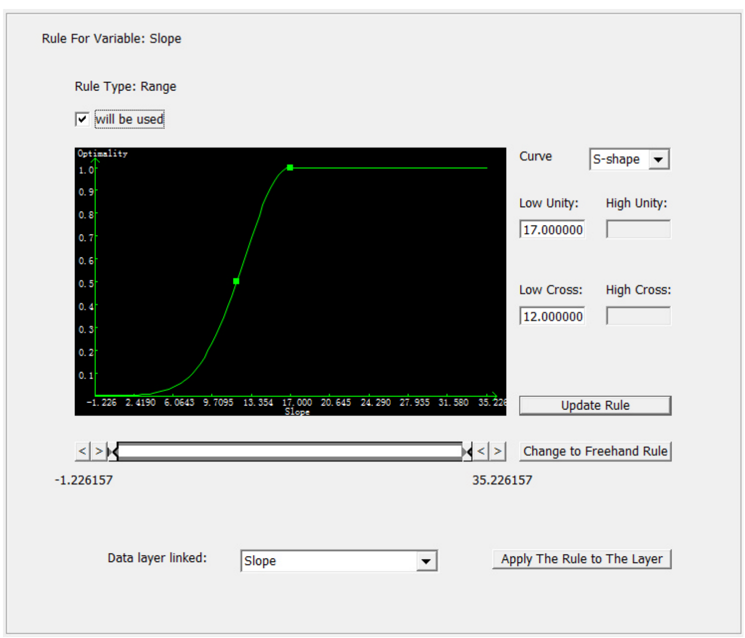
Based on the soil environmental conditions, Soil Type 1 occurs on slopes steeper than 12%. This calls for an S-shaped curve which says that with the increase of slope the membership value for being this soil type also increases (For discussion on curve types see the internet article "Overview of SoLIM" in the **Overview of SoLIM** section and the **Zhu et al. (2013)** article in the **Representation Scheme** section in Appendix D). Click on the "Curve Type" drop down list and select "S-shaped", then click on "Next".



4. Click on "Finish", a default S-shape curve is created for you. The view will switch to rule editing interface.



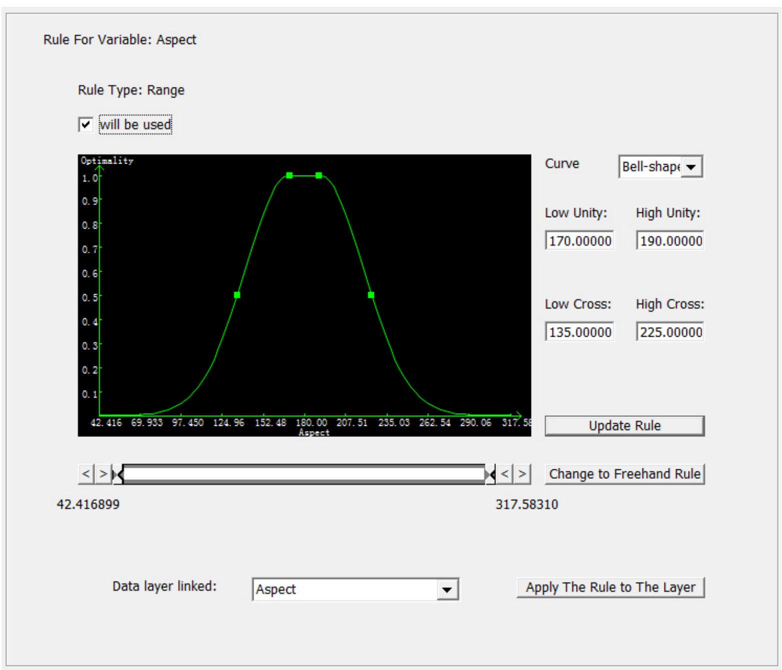
5. We can customize the newly created rule to better represent the knowledge on slope. The current rule as shown above represents that the soil occurs in areas with slope gradient greater than 2.324 %. To make it represent greater than 12%, we need to determine the low unity and low cross. Low unity in this case represents the ideal slope value for the soil to be Soil Type 1. This is another piece of knowledge soil scientists need to provide. Let assume in this case low unity to be 17%. Low cross (the environmental condition at which the membership for this soil type is 0.5) is set to be 12%. When low unity and low cross are all entered, click on "Update Rule" to submit. This brings up the interface shown below



You can also click on and drag the green handles (dots) on the curve to adjust the rule. Please try to drag the handles respectively and see their individual impacts on curve and the values in text boxes. You can also click the "Apply The Rule to The Layer" button to see the effect of the rule.

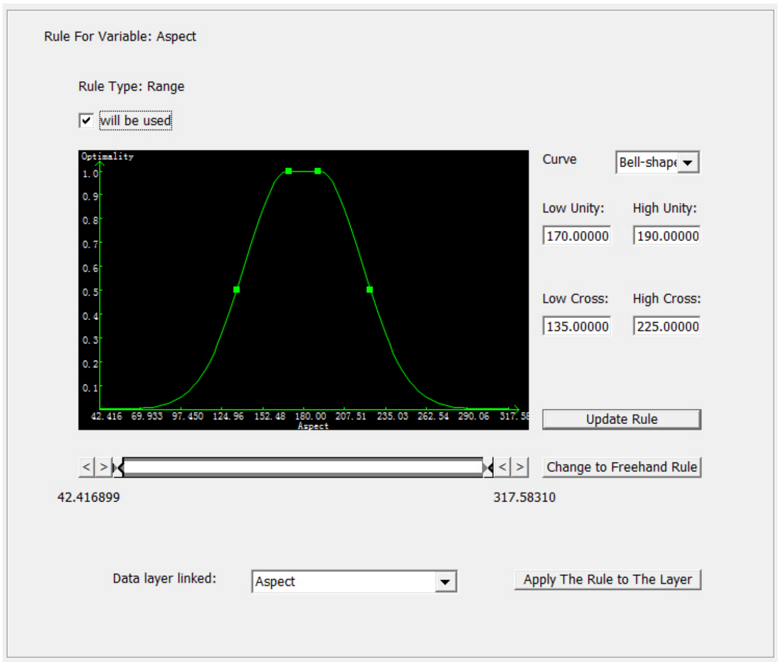
Rule for Aspect

The way to create rule for aspect is much the same as that for slope, except using Bell-shape rather than S-shape curve in the second step. The reason for using the Bell-shape rule here is that the knowledge states that the soil type is on "south facing slope". Thus, the membership value for being "south" should increase first when the slope aspect value increases from 0 and then should decrease when the slope aspect value is over a certain value. When the default rule is created, we should adjust the rule to capture the essence of "South". Absolute "South" means aspect is 180°. We will assume that a range of values around 180° is perfectly suitable for the soil. Thus we will take "South" to mean 170° (low unity) to 190° (high unity). A slightly larger range, from 135° (low cross) to 225° (high cross) will cover less than optimal aspect values. Try to make your aspect rule look like the one shown below. The specific values for low unity, low cross, high unity and high cross are case specific and much of this information is provided by the local experts.



Rules for Planform Curvature

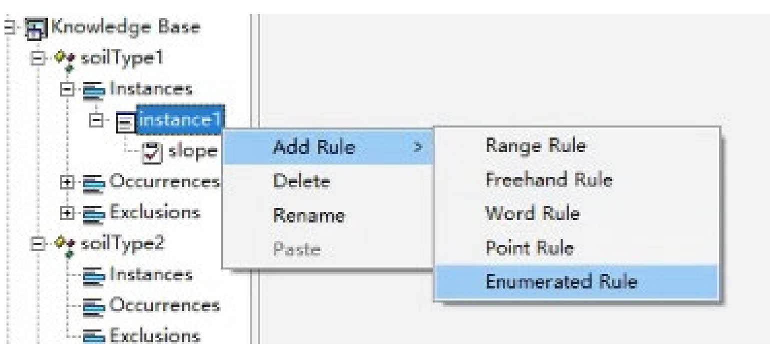
The way to create rule for planform curvature is the same as that for Aspect because linear planform curvature means that the membership value for being linear increases when the planform curvature value approaches zero from the negative and decreases when the planform curvature values deviate from zero into the positive. When adjusting the rule, we set low unity and high unity to 0 to reflect "linearity" (zero curvature) and set low cross and high cross to -0.015, 0.015 to cover less than optimal planform curvature values.



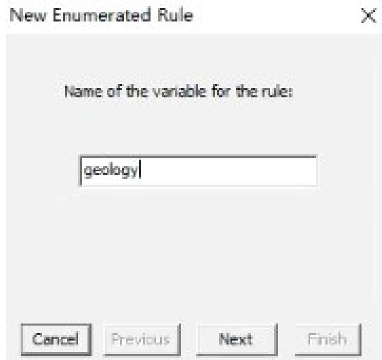
Rule for Geology

Geology is different from other environmental variables because it is a categorical variable. So you should create an enumerated rule for it. The way to create enumerated rule is very similar to that of creating a range rule.

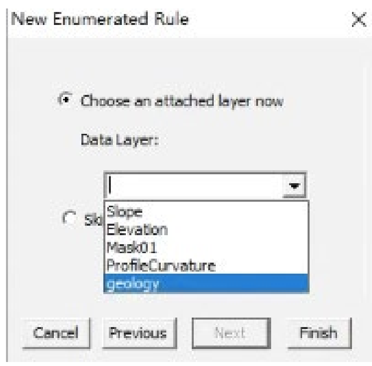
1. Right click on the "Instance1" node. In the pop-up menu, select "Add Rule" and then select "Enumerated Rule".



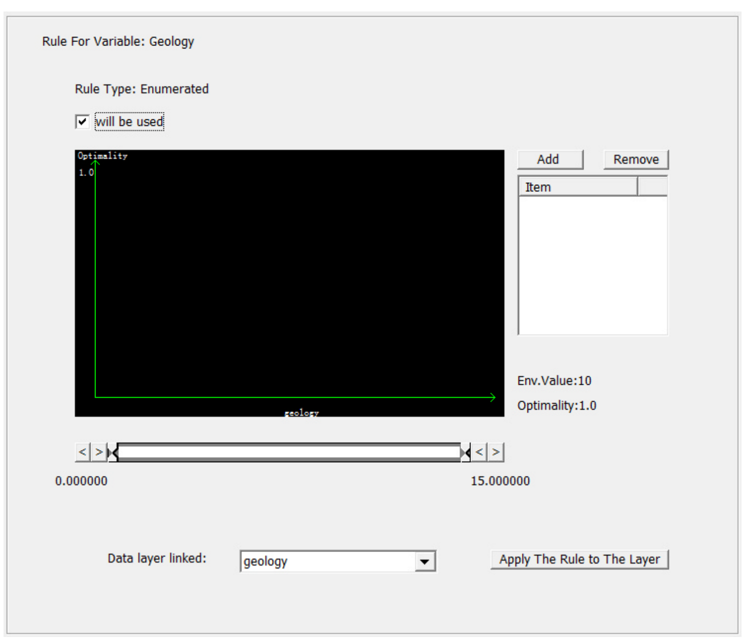
2. Enter the name of the rule.



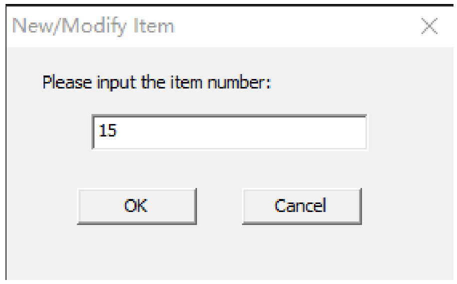
3. Check "Choose an attached layer now", and then select "geology" in the drop down list.



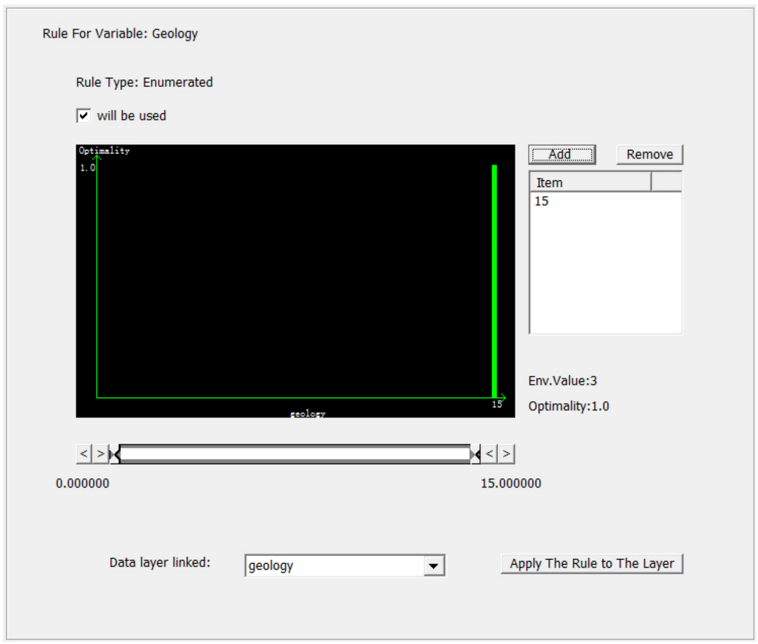
4. Click on "Finish", a blank enumerated rule will be created for you. The view will switch to rule editing interface.



5. In the editing interface, click on "Add". In the pop-up dialog, enter 15.



Then, you will see this:



Now we have encoded the knowledge on soil environmental conditions as rules. You can repeat the process for the other soil types.

Save the Project

Let's save the project through "Project->Save".

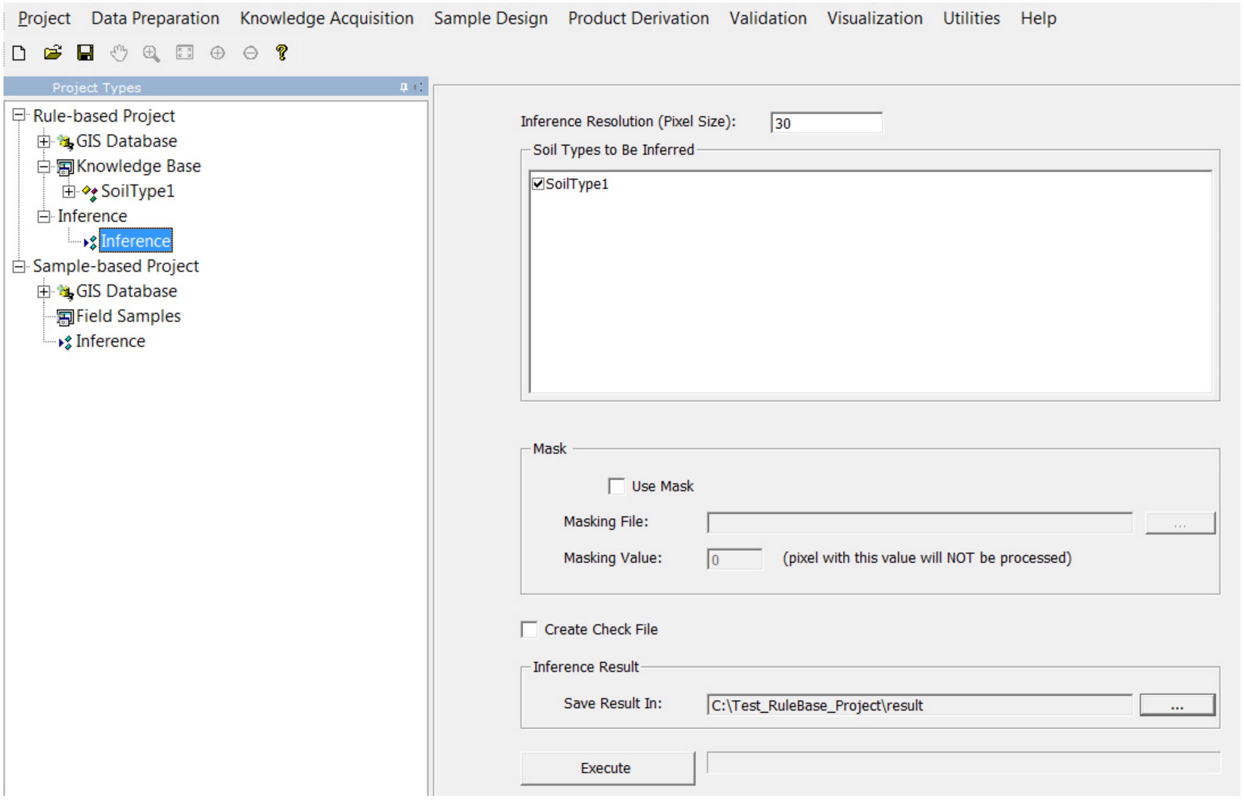
The project configuration and knowledge base will be saved into "test_ruleBased_project.sip" in the project directory.

At this point your project has been created and saved. You can use the control panel to navigate to the rules that have been created and make changes to them. The next step is to run an inference using the encoded knowledge to produce a soil fuzzy membership map.

Run Inference

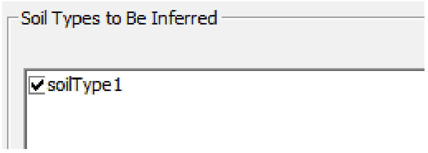
SoLIM Solutions will create a set of soil fuzzy membership maps, one for each soil type which has soil-landscape knowledge defined in the current knowledge base. In this tutorial, we only have the knowledge for one soil type, so there will be only one soil membership map.

Click "Inference" node to unfold it. Under that node, click "Inference", the view will shift to Inference interface.

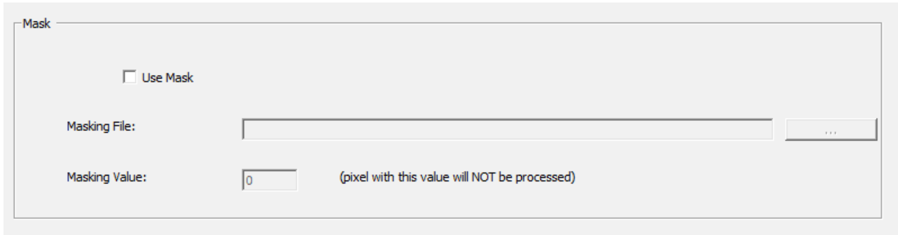


In this view, you can specify the resolution of the inference result . The default resolution is the same as the resolution of the first GIS data layer. We use the default resolution here.

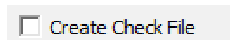
In the "Soil Types to Be Inferred" list, check "SoilType1" to specify that a soil membership map will be created for this soil type.

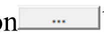


Uncheck the box next to "Use Mask" so that no mask will be used in this tutorial. The mask is used to limit the area of inference. Inference is only performed over areas with values of 1 in the mask.



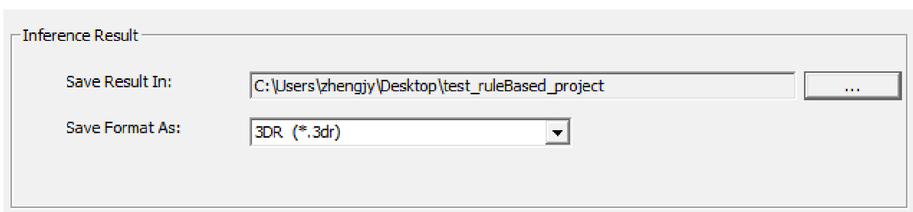
Uncheck the box next to "Create Check File", so no check file will be created. The check file is used to log the information on inference at every location where inference is performed.



Click on  button within the region of "Inference Result", a dialog will appear for you to create or choose output directory.

Create a results directory "result" in the "test_ruleBased_project" directory to host the output membership maps using the "Make New Folder" button.

Click on "OK", the directory will appear on the "Save Result in" text box.



Click on "Execute", in less than a minute a dialog should pop up to inform the completion of the inference. Otherwise, an error message should show up to tell you what was wrong.

Viewing Results

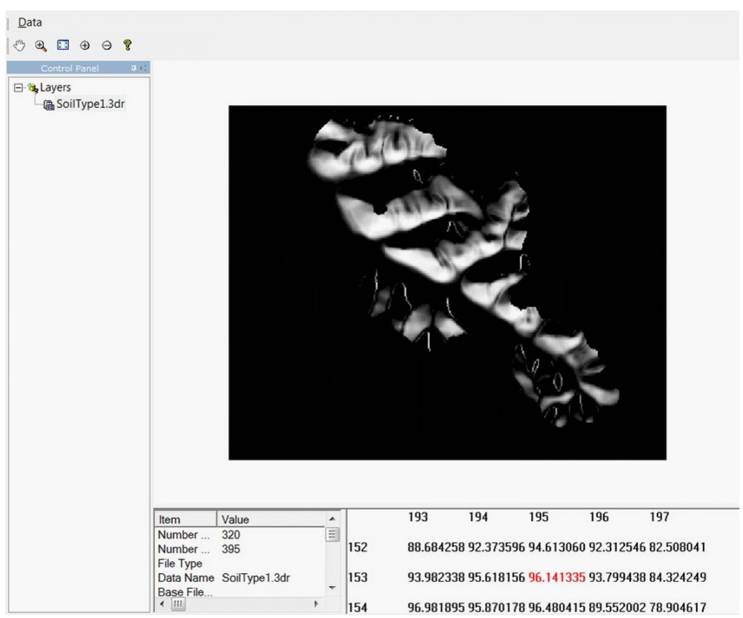
After the inference has completed, you may want to view the results in graphic form. Inference output consists of one fuzzy membership map in a raster form for each named soil. Each raster is stored in a separate file with name NameID.3dr where ID is the soil identification number (unique to each soil type) and Name is the soil type name. After the tutorial inference is complete, the following file will exist:
SoilType1.3dr

Please examine the relationship between the filename of the output fuzzy membership map and the name used in the knowledge base for the soil type. Be sure that each soil type has a unique ID number.

There are three ways to view the results. One is to convert the 3dr file into GRID ASCII format and view the results in other GIS software, such as ArcGIS (**see Utilities->File Converter for details on exporting 3dr file into GRID ASCII format**). The second is to use the 3dMapper (available at http://solim.geography.wisc.edu/software). The third way is to view the results using the Data Viewer provided by SoLIM Solutions 2015. In this tutorial, we will use the SoLIM Data Viewer to visualize the results.

Choose Data->Add Layer.

Navigate to the output directory and select SoilType1.3dr to open it. You should now see this new raster data layer is added to the left layer list panel. At the same time, the layer (the fuzzy membership map for soil type 1) is displayed in the graphic area.



You have now learned the basics of soil mapping using rule-base project in SoLIM Solutions 2015. Feel free to play around with the membership functions and examine the impact on the output. As an exercise, you could create another soil with different rules and re-inference. By bringing the new soil's membership map into Data Viewer, you could compare its distribution with SoilType1.

2.2 Sample-based Project

For this tutorial, let us assume that we want to map soil silt content over an area, but we don't have direct knowledge on the relationship between soil silt content and environmental conditions. However, **we have some field sample points** at each of which soil silt content for a given horizon was measured but these samples were not collected based on any rigid sampling designs. SoLIM Solutions now provides a way (referred to as sample-based inference) to map the spatial variation of soil just using field samples which does not need to be collected based on any rigid sample designs. You can certainly use this approach with samples collected based on a well-designed sampling scheme.

Suppose that the following table is what you have in terms of information about 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 "silt" contains the soil silt content of a given horizon for each sample location.

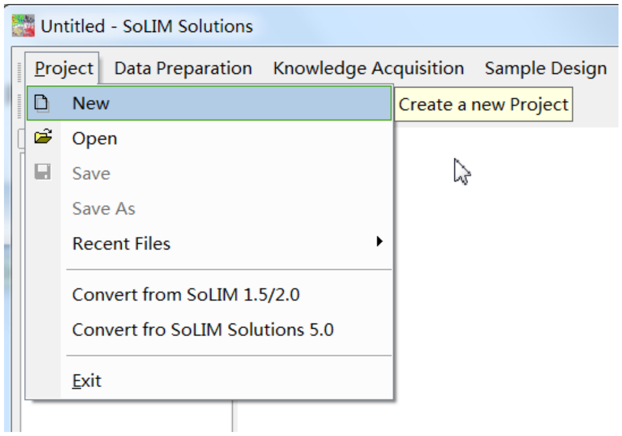
PointID	***X***	***Y***	***sand***	***silt***	***clay***
47	*663487*	*4873463*	*0.070*	*0.712*	*0.218*
48	*663503*	*4873467*	*0.276*	*0.558*	*0.166*
49	*663509*	*4873476*	*0.257*	*0.590*	*0.153*
50	*663516*	*4873473*	*0.232*	*0.623*	*0.145*
51	*663522*	*4873486*	*0.495*	*0.392*	*0.133*
52	*663528*	*4873494*	*0.544*	*0.318*	*0.138*
...	...	...	...	...	...

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

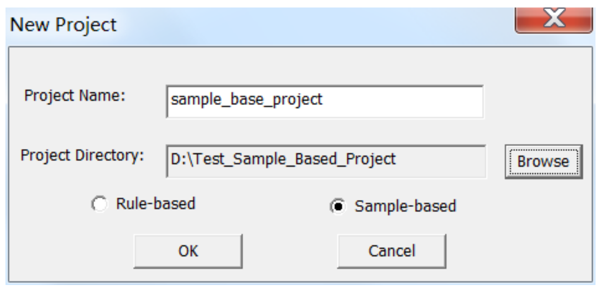
We need to create a .csv file to store these field samples first, which is completed outside SoLIM Solutions. The easiest way to do this is to enter the field sample data into a spreadsheet and save it as a .csv file. You can see a "field_samples.csv" file as an example for this tutorial in "<SoLIM Solutions installed directory>\Tutorial_Data\Sample_Based" directory.

Create a Blank Sample-based Project:

Select "Project->New" to create new project.



Specify Project Name and Project Directory. Here we enter "Test_SampleBased_Project" as the project name and create a new folder "D:\Test_SampleBased_Project" as the project directory.

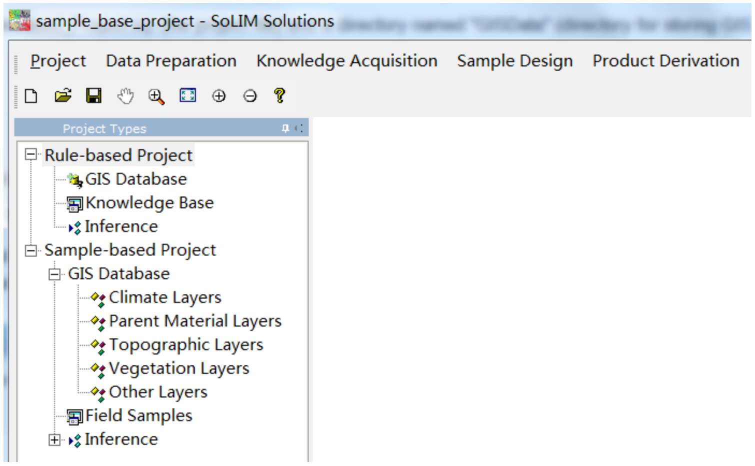


Select "Sample-based" and then click on "OK", a new blank sample-based project will be created for you. The "Sample-based Project" node in the left project panel is activated. In the project directory you just specified, the program generated Test_SampleBased_Project.sip (the project file) and a directory named "GISData" (directory for storing GIS data layers on environment variables) which is currently empty.

Add GIS Data Layers:

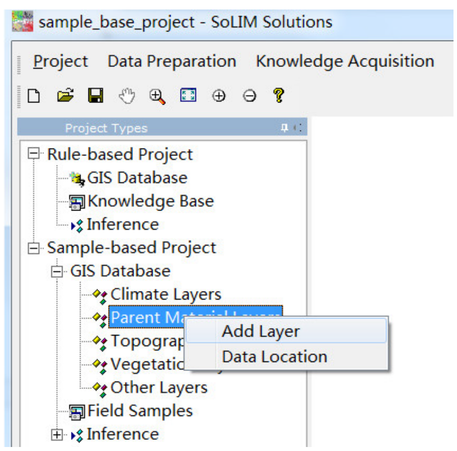
In the left project panel, you will notice that under "GIS Database" node, there are five sub-nodes: Climate Layers, Parent Material Layers, Topographic Layers, Vegetation Layers and Other Layers. We will add our environmental data layers to different sub-nodes.

The native file format for SoLIM Solutions is 3dr format. However, SoLIM Solutions can convert other commonly used raster file format (e.g. TIFF, ERDAS IMAGE) to .3dr format on the fly. SoLIM Solutions can also rasterize commonly used vector file format (e.g Shapefile) into 3dr format. See File Converters in the Utilities menu for how.



There are a few data layers in "<SoLIM Solutions installed directory>\Tutorial_Data\Sample_Based" directory. One of them (Geology.3dr) is for parent material. Other data layers are all topographic layers.

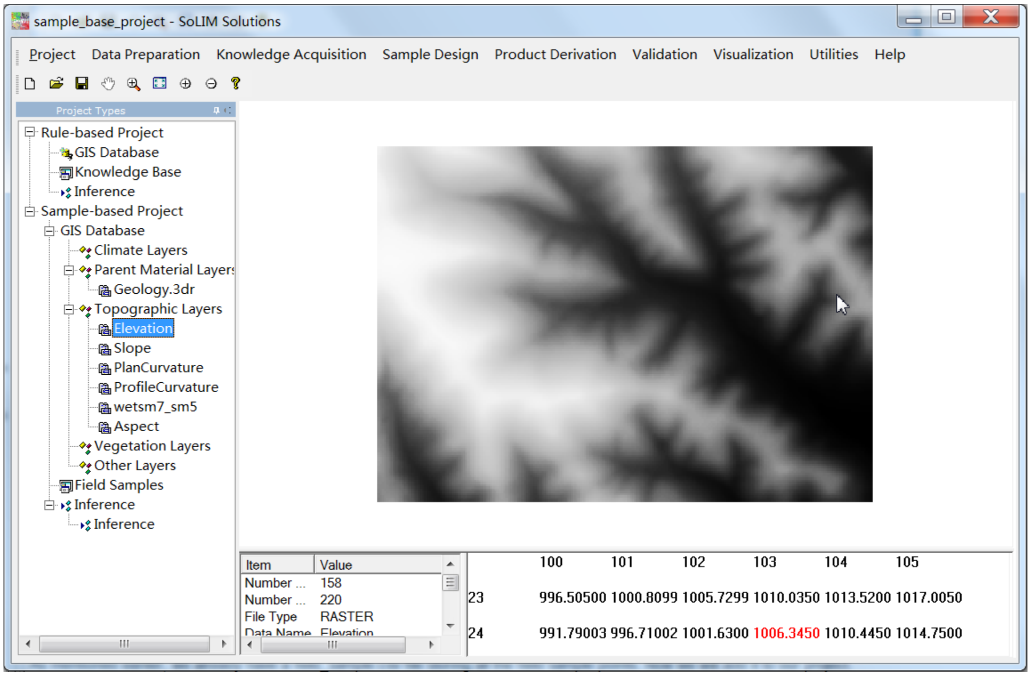
Right click on "Parent Material Layers" node. In the pop-up menu, select "Add Layer" to add Geology.3dr to the project.



Similarly, add all the other environmental data layers to "Topographic Layers" node. Note that you can add multiple layers at one time.

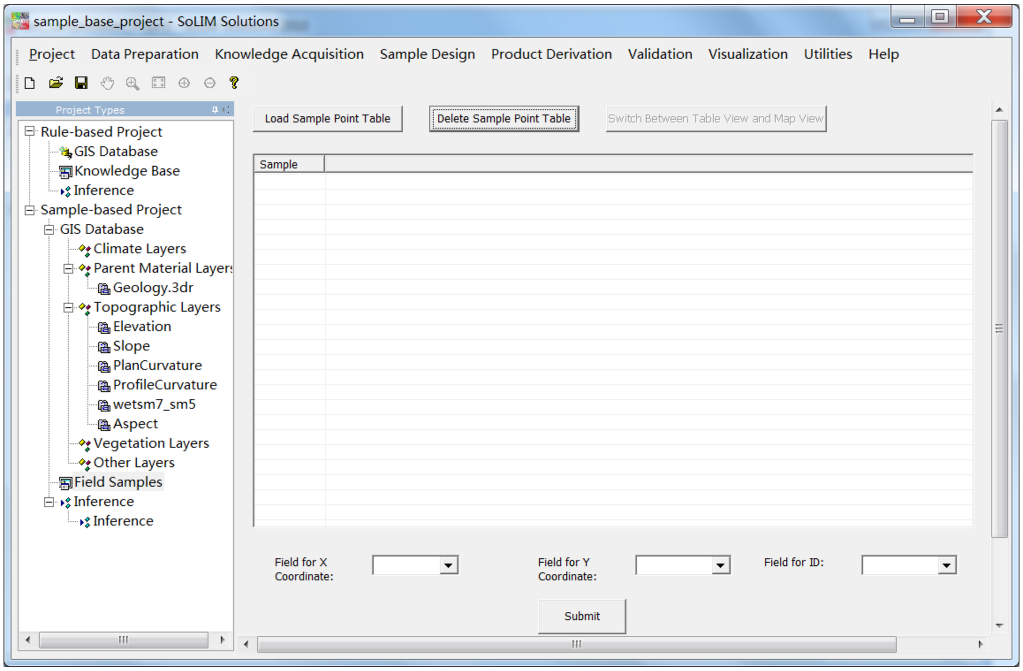
The left project panel should now list the data layers you just added. These data layers are also copied to the "GISData" directory in the project directory you specified earlier (here is "D:\Test_SampleBased_Project"). In that way, they become part of the project and can be easily moved around with the project directory.

After the data layers are loaded, you can click each layer name in the left project panel to view the layer.

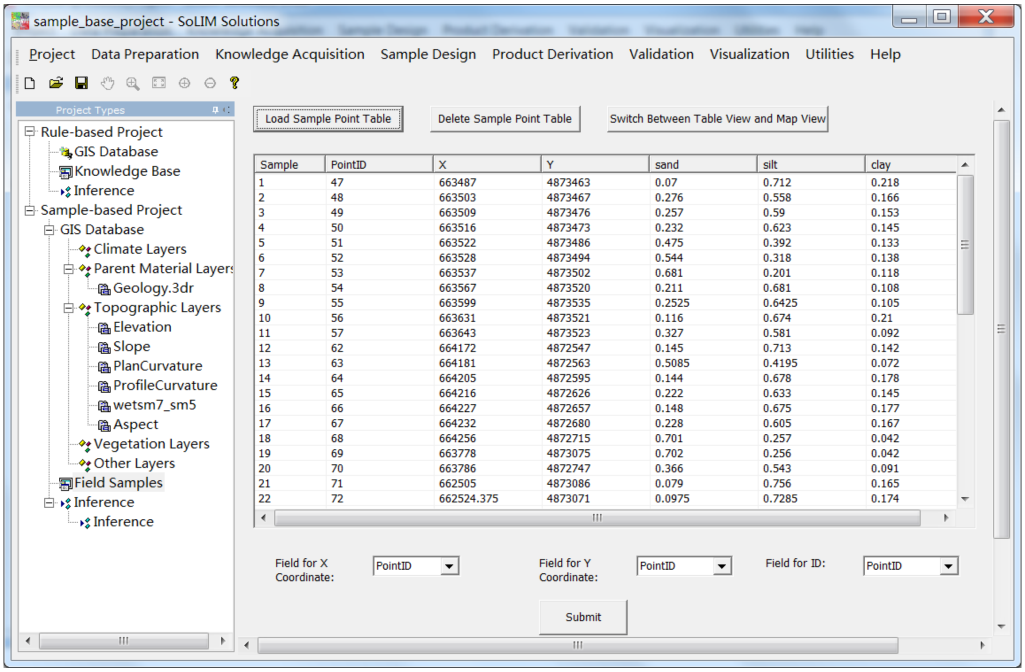


As mentioned earlier, we already have a field_sample.csv file storing all the field sample points. Now we will add it to our project.

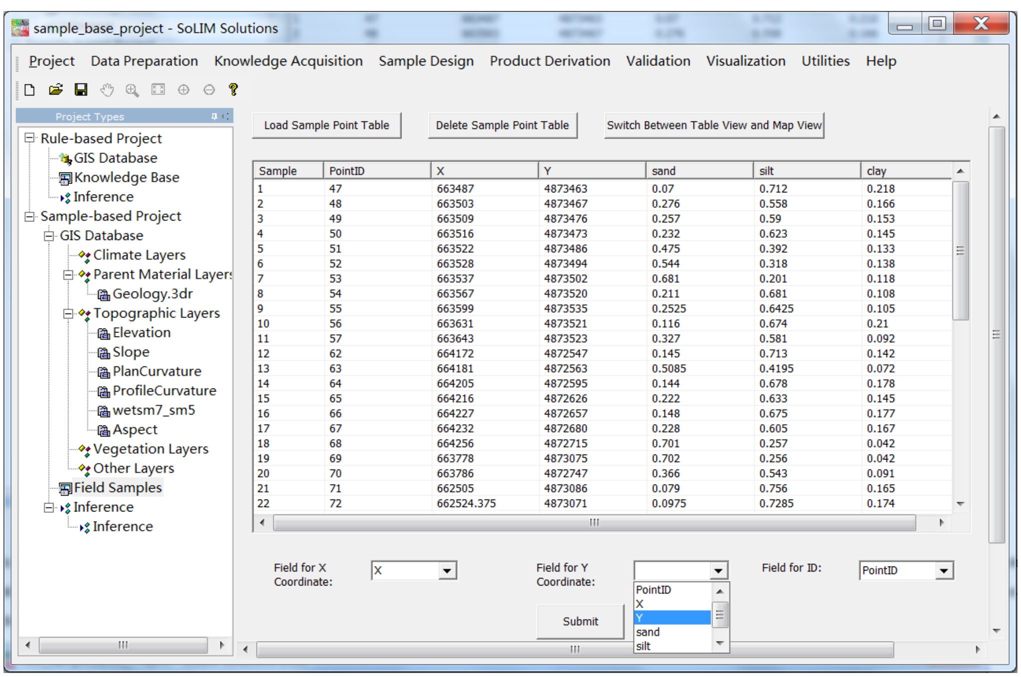
Left-click "Field Samples" node in the left project panel.



Click "Load Sample Point Table" to add the .csv file to our project.

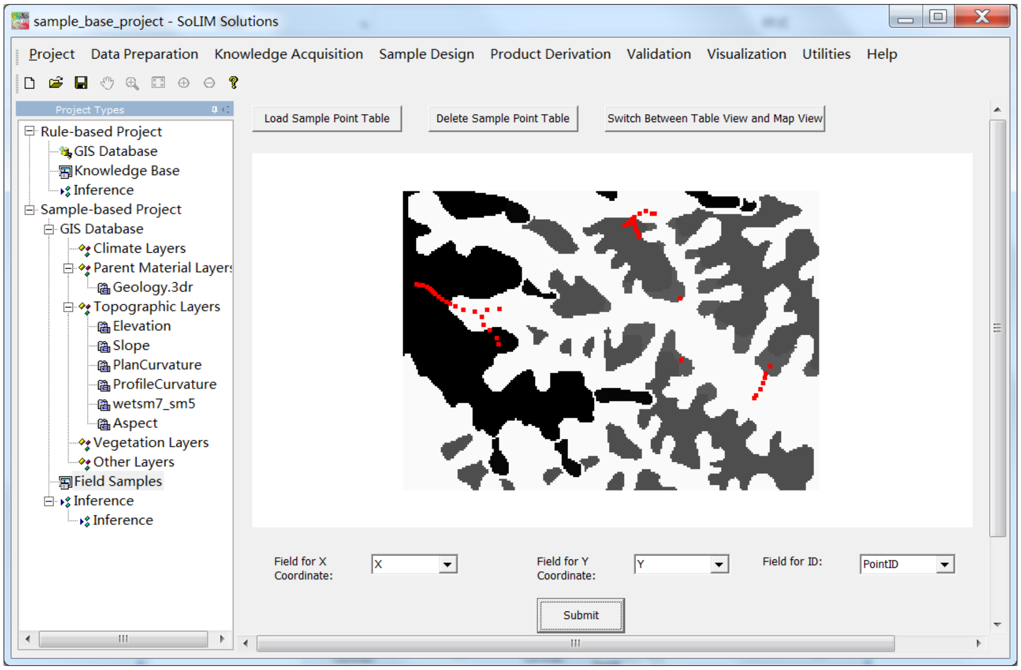


Specify the columns that record X coordinate and Y coordinate by selecting column names from drop down list. Here "X" is for X coordinate and "Y" is for Y coordinate.



Click "Submit", you will see the distribution of the sample points in map view. The red points represent the field sample point in the study area. It should be noted that **at least one layer** needs to be in the GIS

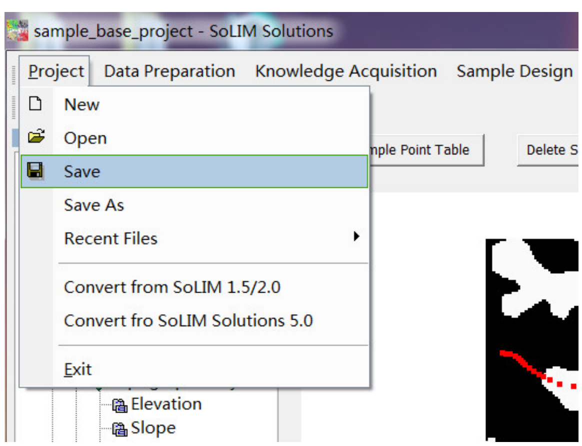
Database. Otherwise, you cannot view the distribution in map view. Also, if the coordinates of your sample points are not in the same coordinate system as that of the environmental data layer in the GIS database, you will not be able to see the sample points in the map view.



You can always switch to table view by clicking "Switch between table view and map view".

Save the Project:

The project can be saved through "Project->Save".



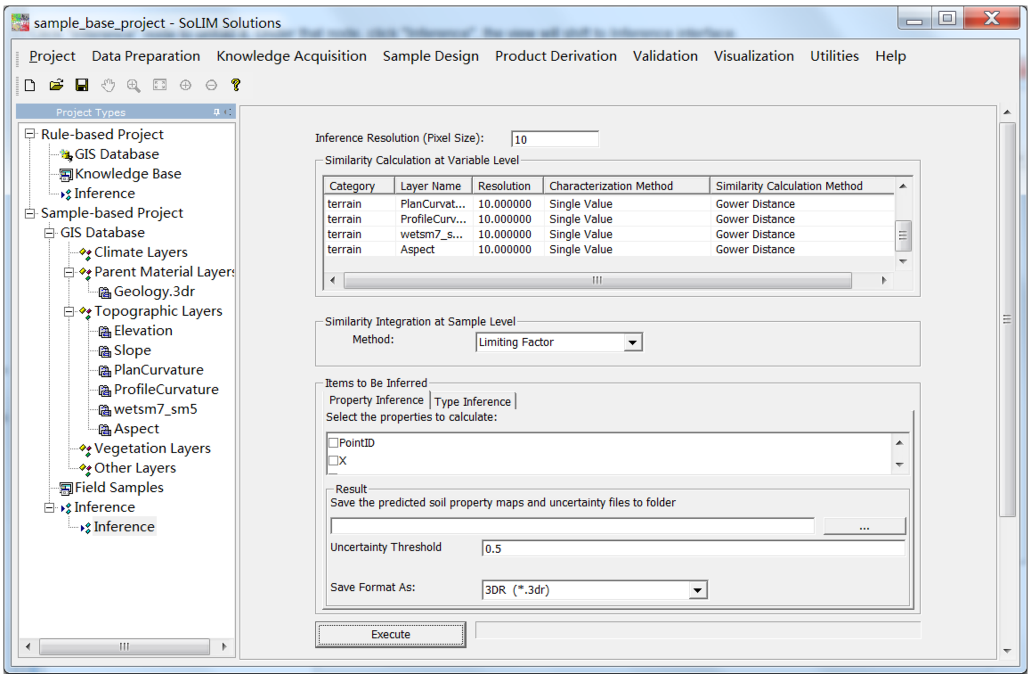
The project configuration and field samples will be saved into "Test_SampleBased_Project.sip" in the project directory.

At this point your project has been created and saved. The next step is to run an inference using the GIS database and field sample points to produce soil silt content map.

Run Inference:

SoLIM Solutions can create multiple soil property maps and soil type map using GIS database and field sample points. In this tutorial, we only infer soil silt content.

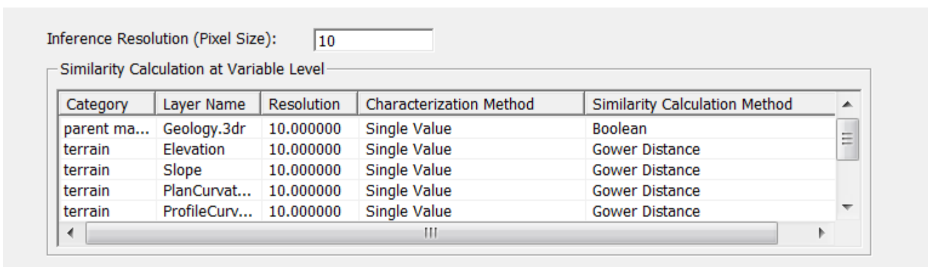
Click "Inference" node to unfold it. Under that node, click "Inference", the view will shift to Inference interface.



Before we make any change to the inference setting, we need understand a few concepts behind sample-based project. The basic idea of sample-based project is to infer soil property/type based on **similarity** between an unknown position and existing field samples based on the environmental conditions. Therefore, how environmental conditions are characterized and how similarity is calculated are most critical in the inference.

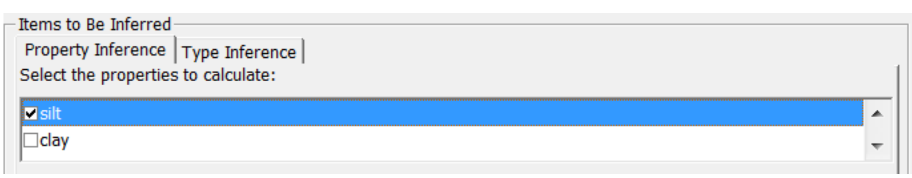
Similarity calculation is conducted at two levels: **variable** level and **sample** level. At the **variable** level, the similarities between an unknown position and sample points are calculated using each environmental layer in GIS database. At the **sample** level, the similarities derived at the variable level are integrated to yield the final similarity between each unknown position and each sample point. There are only one basic characterization methods in this implementation at present: single value. Single value means that the mean value of a given environmental variable over the inference resolution (The default resolution is the same as the largest pixel size of all the GIS data layers. It can also be customized by the user.) is used in calculation of similarity when the inference resolution (the pixel size) is large than that of the environmental data layer. If the resolutions of the environmental data layers are the same and the inference resolution is also that of the environmental data layers, the single value is simply the original value of the pixel.

While you can change the similarity calculation method at both variable level and sample level. In this tutorial, we choose "Boolean" as Similarity Calculation Method for Geology layer and "Gower Distance" for all the terrain layers. You can refer to the user manual for the differences of those methods.



Once that is done we need to specify the soil property or soil type we want to infer and then run the inference. In this tutorial, we only infer soil silt content. During the inference process, SoLIM Solutions does not only predict soil property value for every location but also provides the uncertainty associated with the prediction at each location. If the uncertainty is too high (higher than the uncertainty threshold), SoLIM Solutions will assign NoData to that position in the final soil property map.

Make sure you are in Property Inference tab. Check the box before "silt".

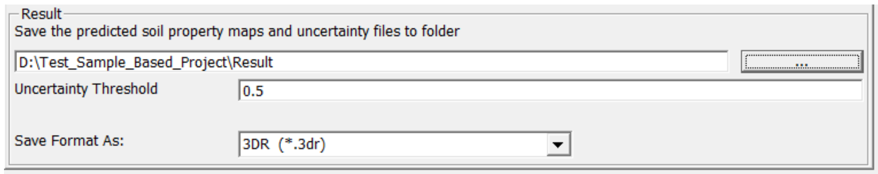


Click on ... button within the region of "Result", a dialog will appear for you to create or choose output directory.

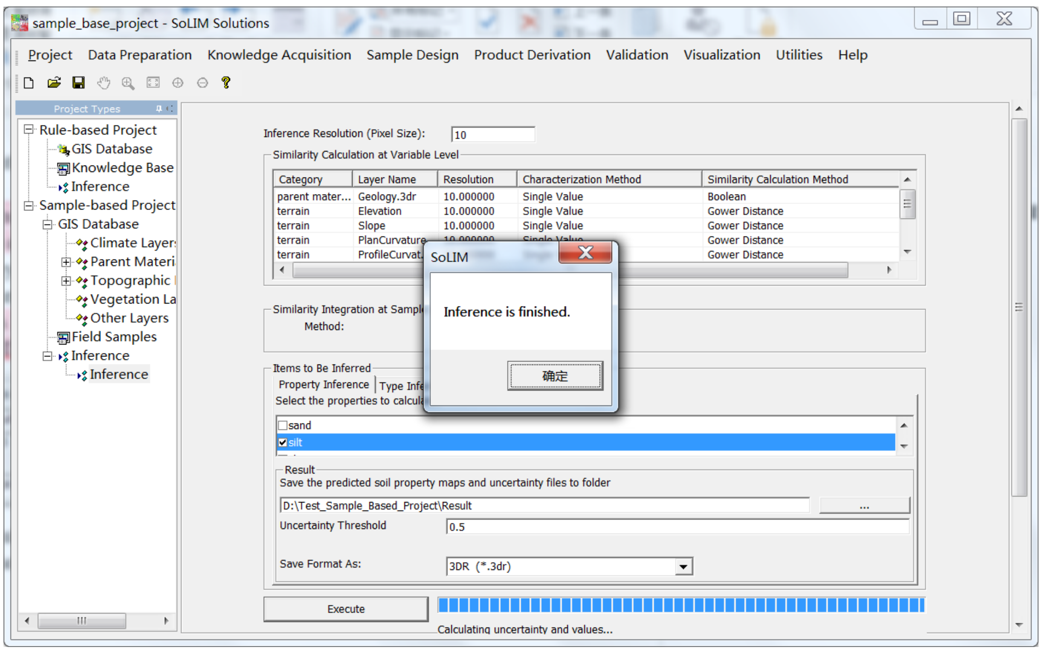
Create a results directory "result" in the "Test_Sample_Based_Project" directory to host the output soil silt content map and uncertainty file using the "Make New Folder" button.

Click on "OK", the directory will appear on the "Save Result in" text box.

We will use the default uncertainty threshold 0.5. And we have multiple choices for the format of the output data, such as 3DR format (.3dr), Erdas Imagine format (.img), SAGA format (.sdat), TIFF/BigTIFF/ GeoTIFF format (.tif). For example, we can use default format (.3dr).



Click on "Execute", if the inference is finished without an error, a dialog should pop up to inform the completion of the inference. Otherwise, an error message should show up to tell you what was wrong.

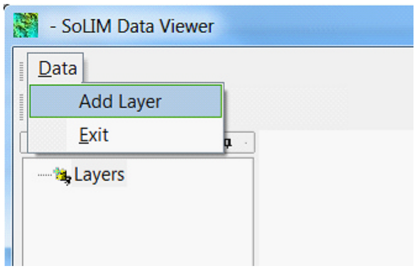


Viewing Results:

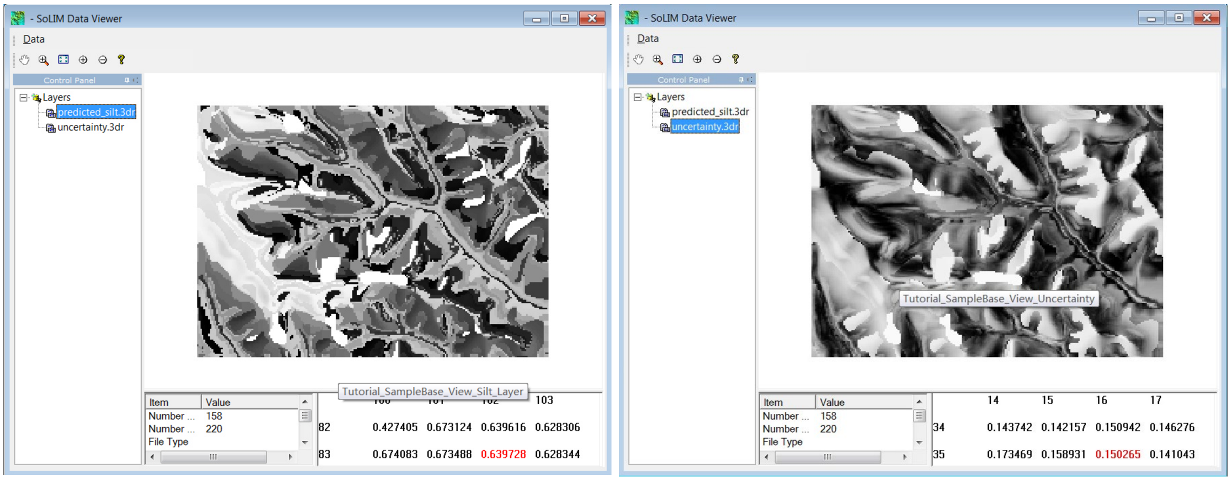
After the inference has completed, you may want to view the results in graphic form. If you have been following this tutorial so far, you should see "predicted_silt.3dr" file in "D:\Test_SampleBased_Project\ result" directory. This file is the final prediction result. You can also find "uncertainty.3dr" file in the project folder.

There are three ways to view the results. The first one is to convert the 3dr file into GRID ASCII format and view the results in other GIS software, such as ArcGIS (see Utilities->File Converter for details on exporting 3dr file into GRID ASCII format). The second one is to use the 3dMapper (available at http://solim.geography.wisc.edu/software). The third way is to view the results using the Data Viewer provided by SoLIM Solutions 2015. In this tutorial, we will use the 2D Data Viewer to visualize the results.

Choose "Visualization->2D" and open the Data Viewer. Then choose "Data->Add Layer"



Navigate to the output directory and select "predicted_silt.3dr" and "uncertainty.3dr" to open them. You should now see these two data layers are added to the left layer list panel. You can choose one layer to view it the graphic area by clicking the corresponding node in the layer list panel.



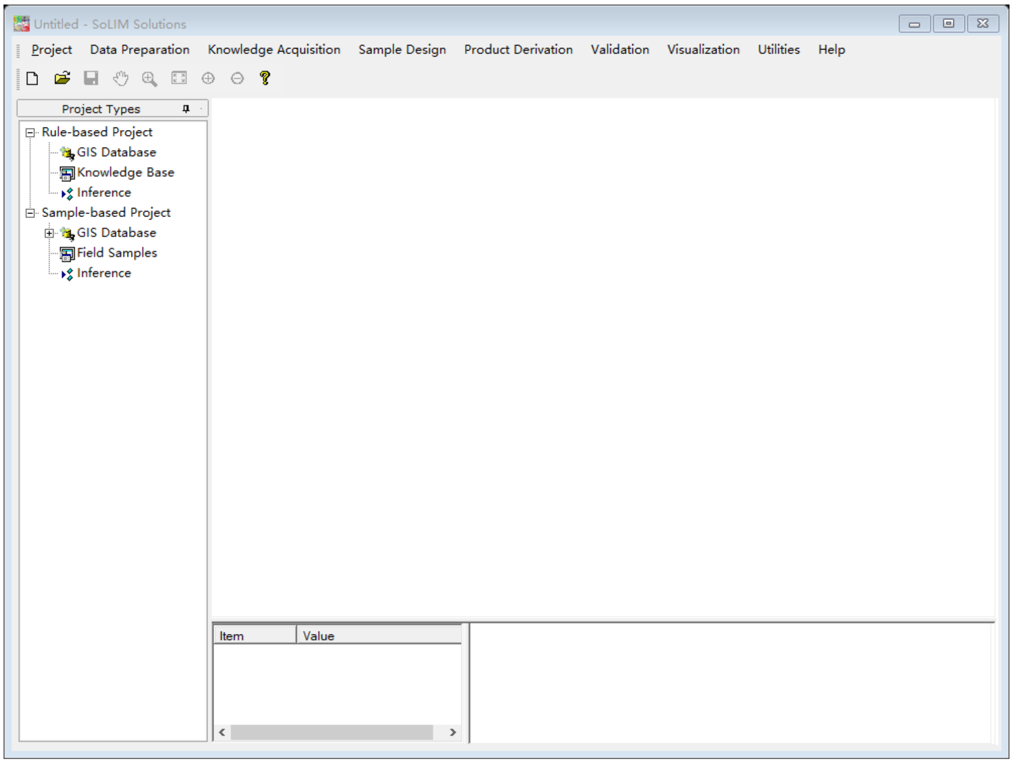
For the soil silt content layer, bright grey represents high silt content and dark grey scale represents low silt content. Note that some pixels have value -9999 which means *NoData*. This is due to the corresponding uncertainty values at those locations are higher than the uncertainty threshold (0.5 in this tutorial).

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

3. Reference Manual

3.1 Interface of SoLIM Solutions

The directory in which SoLIM Solutions is installed is referred to as <SoLIM Solutions installed directory>. Start SoLIM Solutions which will present you with the following user interface:



The Menu:

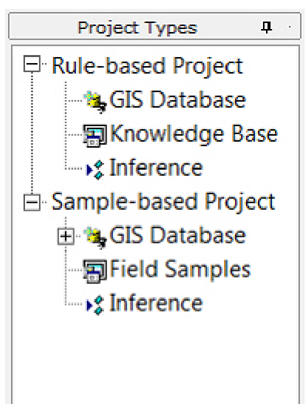
- The "**Project**" menu provides the interface for managing the projects. A project maintains all necessary information needed for performing a SoLIM inference (soil mapping). Project menu also supports importing existing projects created with the old versions of SoLIM software.
- The "**Data Preparation**" menu provides terrain analysis tools which implement the most recently developed ideas in digital terrain analysis and it also provides tools for extracting soil-related surface dynamic feedback from remote sensing data (Zhu et al., 2010).
- The "**Knowledge Acquisition**" menu provides tools to extract knowledge from experts or through purposive sampling or through Map.
- The "**Sample Design**" menu provides traditional sample design tools as well as purposive sampling design tool (Yang et al., 2012).

- The "**Production Derivation**" menu provides the interface to produce soil property maps derived using fuzzy soil membership maps (Property Map) and soil type maps (Hardened Map).
- The "**Validation**" menu provides tools to evaluate the results (type maps and property maps) using field point data including property validation and type validation.
- The "**Visualization**" menu provides tools to visualize environmental data layer and inference results in 2D or 3D form.
- The "**Utilities**" menu provides file conversion and other data analysis utilities.
- The **"Help"** menu provides SoLIM Solutions 2015 user manual for users to query and the software version information.

The Toolbar:

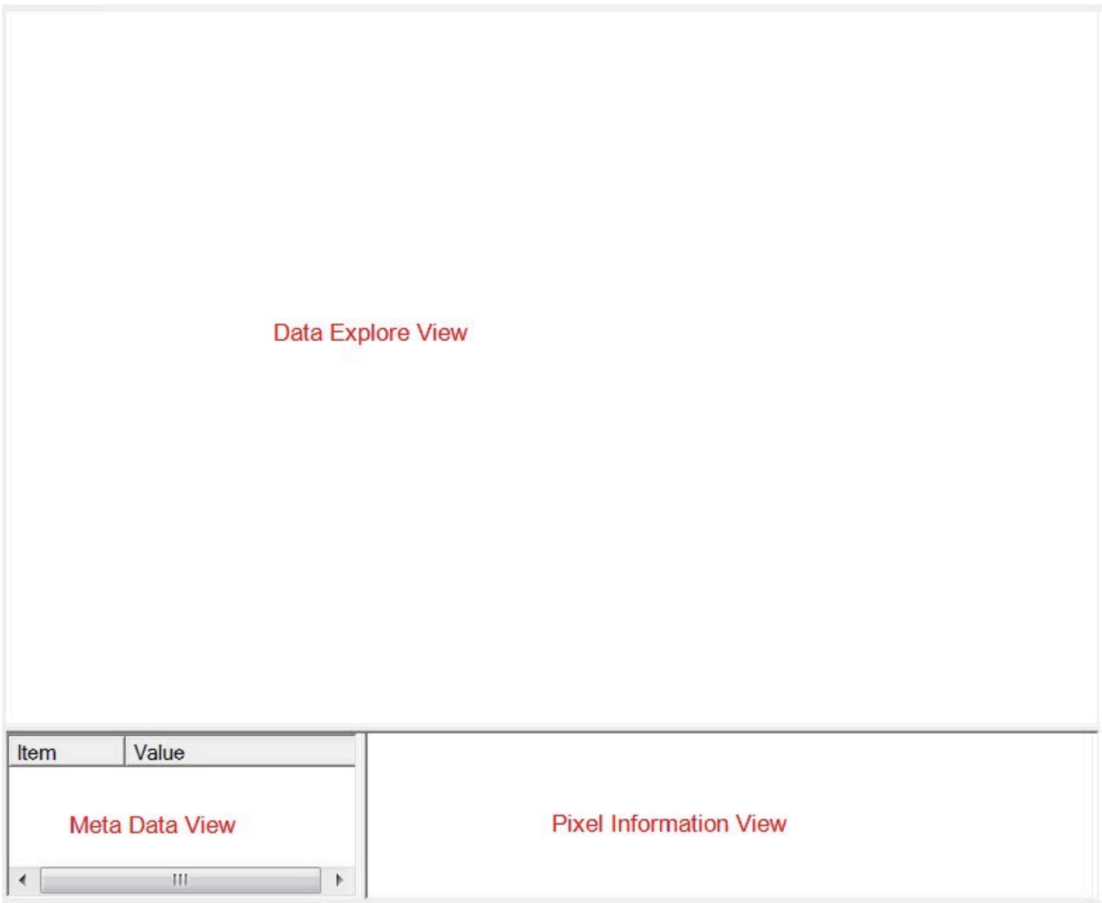
The toolbar provides some frequently used functions, including creating project, saving project and exploring environmental data layers.

The Project Panel:



The project panel contains a tree-style list. It is used to control SoLIM projects. As you can see from this panel, SoLIM Solutions 2015 supports two types of projects: rule-based and sample-based. Both rule-based project and sample-based project are based on the idea that soil types and/or soil properties can be inferred from soil-related environmental conditions, so they both require a set of GIS layers (GIS Database) which depict the soil-related environmental conditions. The difference between them is that in rule-base project, we need to define a set of rules (Knowledge Base) to express the soil-environment relationship explicitly, while in sample-base project, we just need to provide field sample points (Field Samples). As stated before the sample-based project does not need the samples based on any spatial sampling design. You will have a better understanding of the difference after you go through this tutorial. Currently, the two projects cannot be mixed.

Views:



The view section displays current information based on the node selected in the project panel. The figure above shows the view for GIS database management.

3.2 Project

SoLIM Solutions supports two types of soil mapping projects: **rule-based project** and **sample-based project.** Both of them are based on the idea that soil types and/or soil properties can be inferred from soil-related environmental conditions, so they both require a set of environmental layers (or covariates) (stored in a **GIS Database**) which depict the environmental conditions which are indicative of soil. The difference between rule-based and sample-based is that in rule-base project, users need to define a set of rules (**Knowledge Base**) to express the soil-environment relationship explicitly, while in sample-base project, users need only to provide field sample points (**Field Samples**) to express the soil-environment relationship implicitly. The sample points can be those collected from a well-designed sampling strategy or from ad-hoc activities (meaning not design at all).

3.2.1 Rule-based Project

Project Management Overview

A rule-based project is the unit on which SoLIM Solutions performs database management, knowledge definition and soil inference. A complete definition for a project consists of three components:

1. **Project file(.sip):** This file contains project configurations and knowledge base used in a project.
2. **Environmental data layers (GIS database) containing raster data layers needed for inference:** The files must be in 3dMapper raster format (.3dr) in a subdirectory named "GISData". This subdirectory must be in the same directory as the project file.
3. **Knowledge base file (.xml):** This file records the rules (knowledge).

A project is stored in a project directory. Soil inference requires a GIS database and a knowledge base.

Optionally:

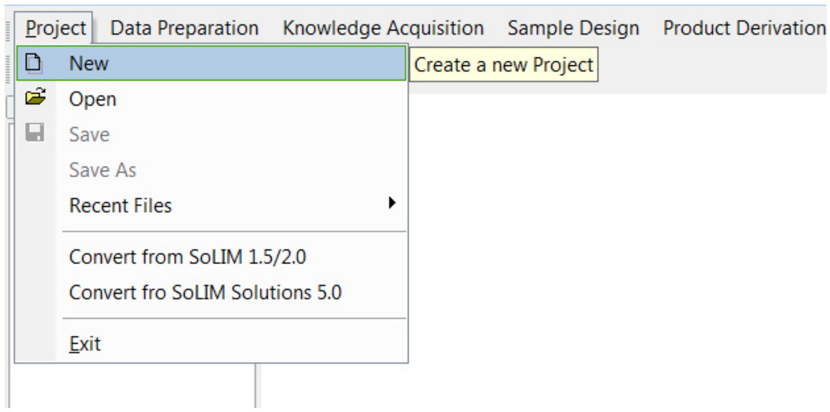
The word library file should be in the project directory if word rule(s) are used in the knowledge base.

In this section, you will find information about how to

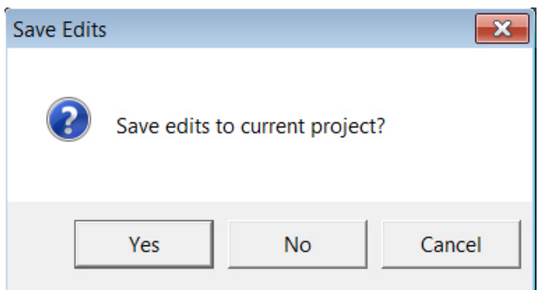
- Create a New Project
- Open an Existing Project
- Edit a Project
- Save a Project

Create a New Project

Select "Project->New" to create a new project.

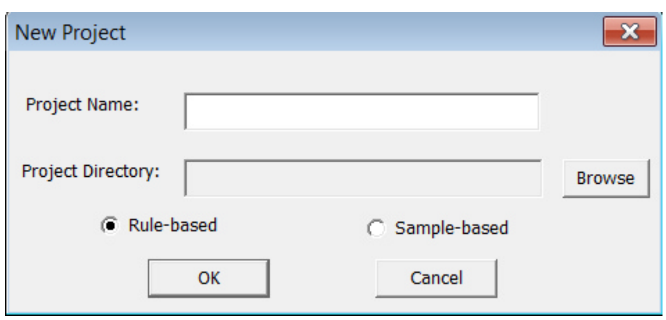


If you are editing another project, the program will remind you to save the current project.



Click on "Yes" to save it, "No" to skip or "Cancel" to return.

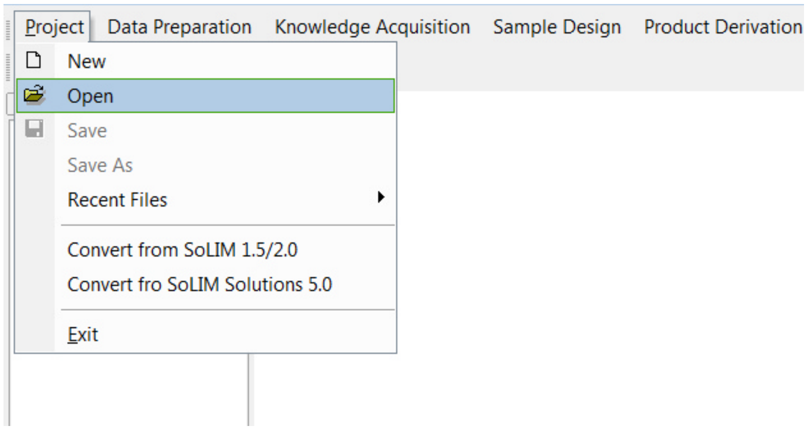
If you choose "Yes" or "No", the dialog below will appear. Specify project name and project directory.



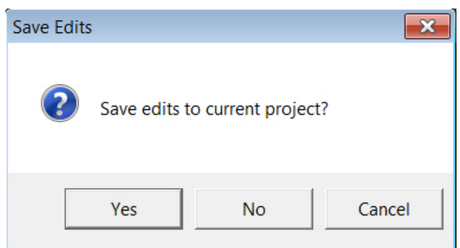
Open an Existing Project

If a SoLIM Solutions project has already been saved, you can open it and edit.

Select "Project->Open".



If you are editing another project ,the program will remind you to save the current project.



Click on "Yes" to save it, "No" to skip or "Cancel" to return.

If you choose "Yes" or "No", the original project is closed. A dialog will appear to let the user specify the project file (.sip) to be opened. After selecting a file, click on "Open". If the specified project file is valid, it will be opened. Otherwise, the program will pop up an error dialog.

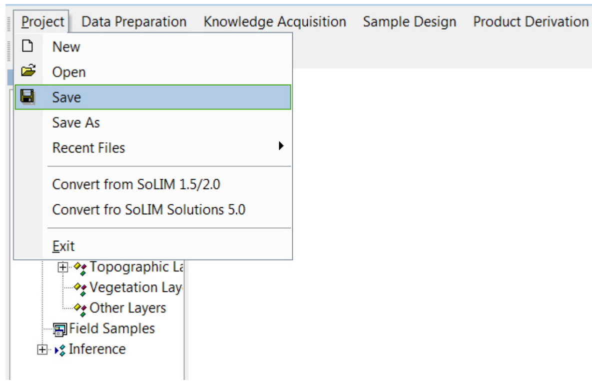
Edit a Project

Please refer to GIS Database Management and Knowledge Base Management.

Save a Project

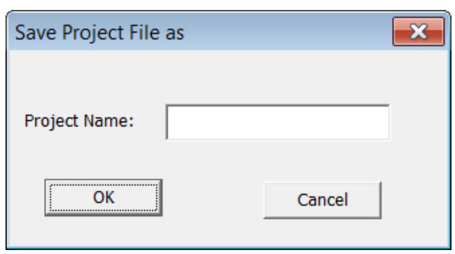
You can save a project at any time.

Select "Project->Save".



You can save a project to a file with another name by clicking "Project->Save As". Save As is useful when you need to save different versions of a project with different modification. In fact, they (these versions) can be treated as different projects with the GIS data layers in the same directory (GISData). This can save disk space if the majority of the GIS data layers are common to multiple projects.

In the dialog below, specify the name you want to use.



click on "OK" and the program will save the project file(.sip) with the name you specified in the project directory.

GIS Database Management

Before any inference can be performed, the needed environmental data layers must be ready and properly referenced in the project file. The collection of these references to environmental data layers constitutes the GIS database. The GIS database:

- provides data on the environmental variables used in the soil inference; and
- defines the spatial extent of the resulting soil maps.

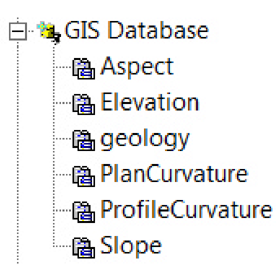
If the data layer for an environmental variable is not referenced or does not exist, the variable will not be used in the inference. When loading environmental data layers, the layers must be in the same coordinate system.

The native file format for SoLIM Solutions is 3dr format. However, SoLIM Solutions can convert other commonly used raster file format (e.g. TIFF, ERDAS IMAGE) to .3dr format on the fly. SoLIM Solutions can also rasterize commonly used vector file format (e.g. Shapefile) into 3dr format. See File Converters in the Utilities menu for how.

SoLIM Solutions automatically checks the spatial extents and resolutions of the data layers when they are loaded. If the spatial extents of the data layers do not match, SoLIM Solutions only perform soil inference over the common area covered by all data layers. Now data layers with different resolutions can be used together in SoLIM Solutions.

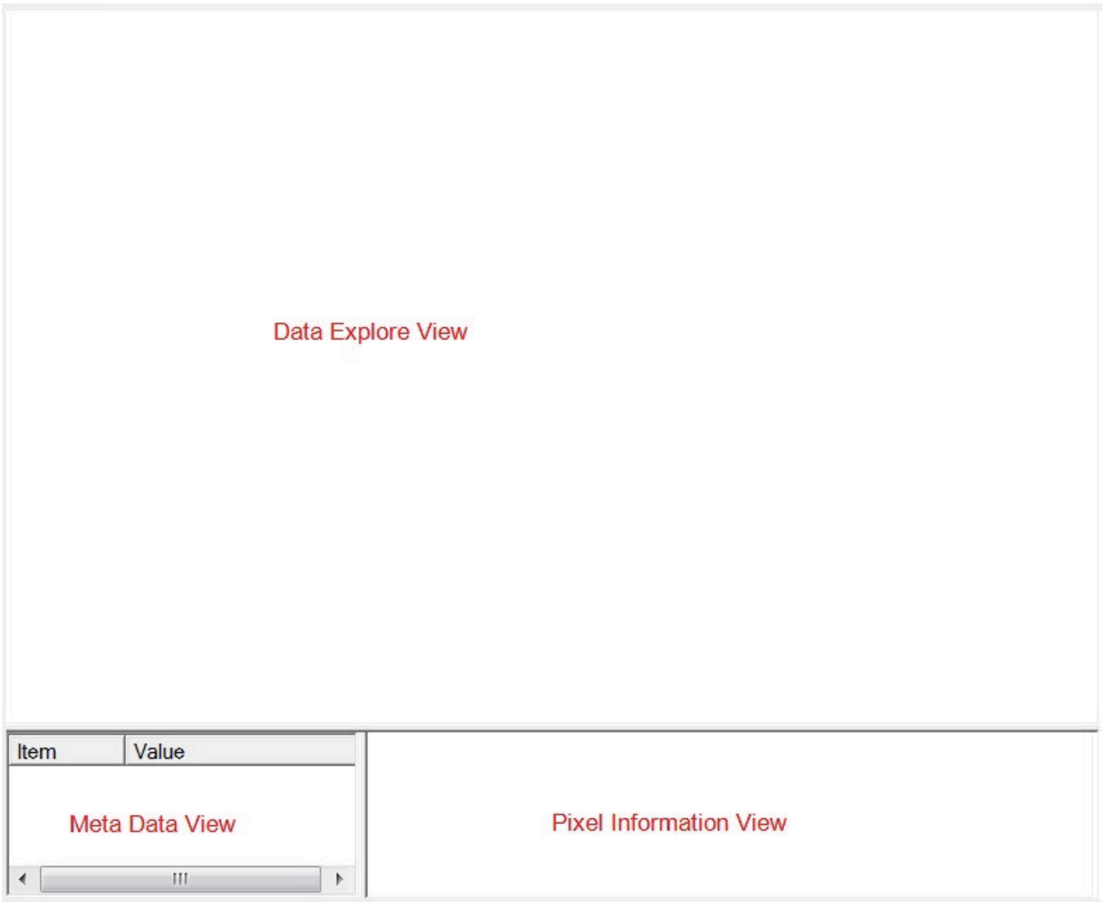
This section provides an instruction to the GIS database management in SoLIM Solutions. With the functionalities provided by the program, you can construct, display, and explore the environmental datasets of your study area. GIS database management consists of the following tools:

a. The “GIS Database” Node in the Project Panel



By manipulating this node and its sub-nodes, you can add and delete environmental data layers in the GIS database.

b. The GIS Database Views



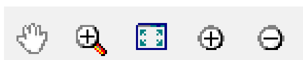
The GIS database view contains three sub-views:

the Data Explore View: provides a graphical view of the data layer currently selected.

the Pixel Information View: displays the value of pixels at or near the mouse pointer.

the Meta Data View: displays the header information (meta data) of the data layer currently selected.

c. The View Control Toolbars

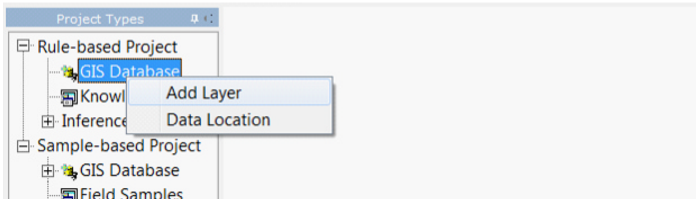


The toolbar provides view control functionalities, such as zooming and panning for you to explore the displayed data layer.

Add and Delete Data Layers

Add Data Layer(s)

1. Right click on the "GIS Database" node. In the pop-up menu, select "Add Layer"



2. In the add layer dialog, choose one or more data files. The recommended file format is .3dr format. However, SoLIM Solutions can convert other commonly used raster file format (e.g. TIFF, ERDAS IMAGE) to .3dr format on the fly. SoLIM Solutions can also rasterize commonly used vector file format (e.g. Shapefile) into .3dr format on the fly.

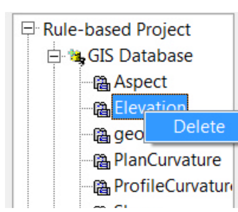
3. click on "OK", The data layer is added to the current project and copied into the GISData directory of the current project. If the layer is not valid, SoLIM will pop up an error dialog and the layer will not be added.

Tips

1. The default name of a data layer is the "data name" stored in the header of the .3dr file.
2. The name of a newly added data layer **must be different** from the names of **all of the existing** data layers in the project.

Delete a Data Layer

In the control panel, right click on the name of the data layer to be deleted. Click on "Delete".



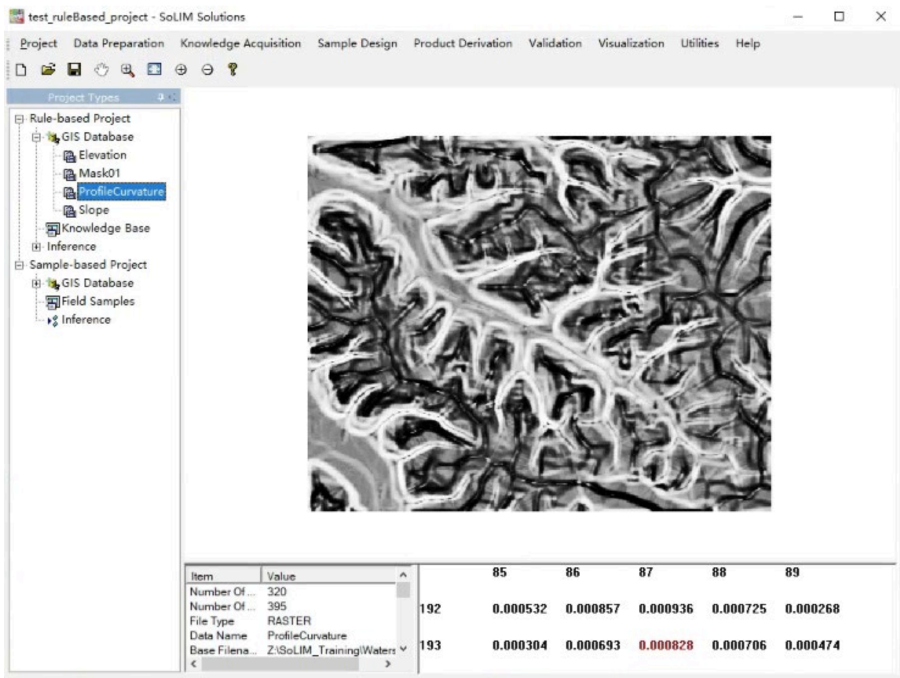
Tips

1. The data layer is only **removed** from the list. The data layer file still in the GISData directory.
2. If a layer is deleted, the rules which attach with this layer will be removed: the attached rule will be set to NULL.

Exploring Your Data

Display the Layer

Click on a data layer under "GIS Database" node, the layer will be displayed in the data explore view. Header information will be displayed in the meta data view.



View the Value of Pixels

In the data explore view, move the mouse to point at the pixel, the value of the pixel will be displayed in red color at the center of the pixel information view. The values of the pixels near the mouse pointer will also be displayed in black color. If you want to view more neighboring pixels, drag the edge of pixel information view to enlarge the view display area.

	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
208	-0.000884	-0.000632	-0.000271	-0.000117	-0.000362	-0.000892	-0.001462	-0.001813	-0.001935	-0.001979	-0.001863	-0.001642	-0.001444	-0.001350	-0.001425	-0.001495	-0.001428
209	-0.001056	-0.000978	-0.000558	-0.000039	0.000303	0.000334	-0.000117	-0.000785	-0.001286	-0.001518	-0.001503	-0.001328	-0.001183	-0.001239	-0.001479	-0.001777	-0.002062
210	-0.000863	-0.000921	-0.000834	-0.000490	-0.000352	-0.000480	-0.000556	-0.000597	-0.000605	-0.000604	-0.000569	-0.000394	-0.000199	-0.000114	-0.000275	-0.000845	-0.001853
211	-0.000664	-0.000614	-0.000487	-0.000272	-0.000245	-0.000391	-0.000491	-0.000584	-0.000675	-0.000710	-0.000695	-0.000615	-0.000456	-0.000116	0.000386	0.000845	0.000573

Zoom In

Click on ⊕ in the toolbar to zoom in the display.

Zoom Out

Click on ⊖ in the toolbar to zoom out the display.

Zoom in to a Rectangle

Click on in the toolbar, drag a rectangle in data explore view to zoom in the display to the specified rectangle.

Pan

Click on in the toolbar, hold the mouse button down while dragging in the direction you want to move the area, then release the button.

Show All

Click on in the toolbar to set the display to the original state.

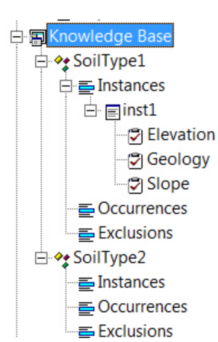
Knowledge Base Management

Before any inference can be performed, knowledge that captures the relationship between soil types and environmental variables must be obtained and organized into the knowledge base. The knowledge base:

- is organized as a hierarchical structure.
- has one or more soil types.
- contains global knowledge (expressed as instances) which takes effect on the whole mapping area and local knowledge (expressed as occurrences and exclusions) which takes effect only on limited area.
- the basic units are rules.

This section provides an instruction to the knowledge base management in SoLIM Solutions. With the functionalities provided by the program, you can construct and edit knowledge base. Knowledge base management is realized by the following tools:

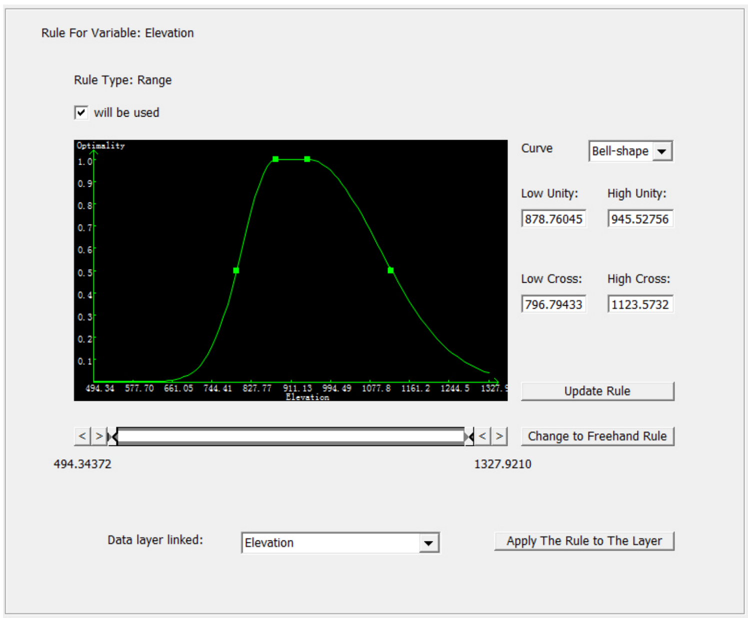
a. The "Knowledge Database" Node in the Control Panel



By manipulating this node and its sub-nodes, you can add, delete, edit and copy knowledge in the knowledge base.

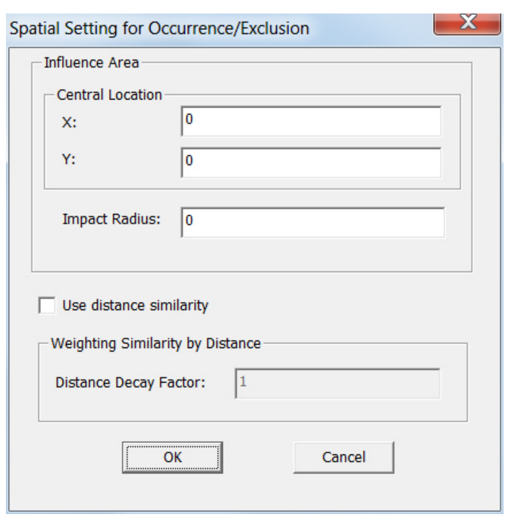
b. The Rule Editing View

It provides a graphical view to current selected rule. Almost all the editing work is done here. Different kinds of rules have different editing view. However, the basic components of the views are the same. The following picture shows the editing view for a Range Rule.



c. Spatial Setting Dialog

The spatial setting dialog is only for local knowledge (occurrences and exclusions) and is used to set their central locations, impact radii and distance decay factors.

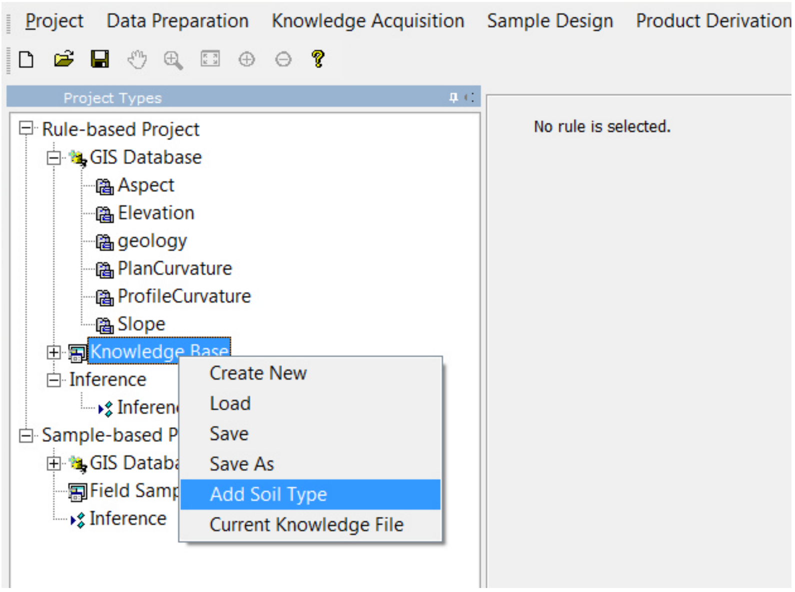


Soil Type

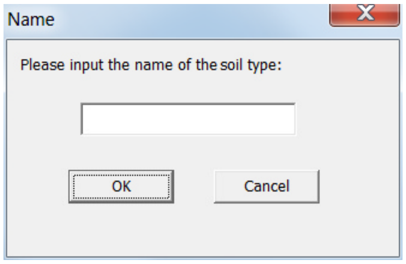
A knowledge base contains knowledge for one or more soil types. A soil type is a taxonomic class or user-defined category.

Add a Soil Type

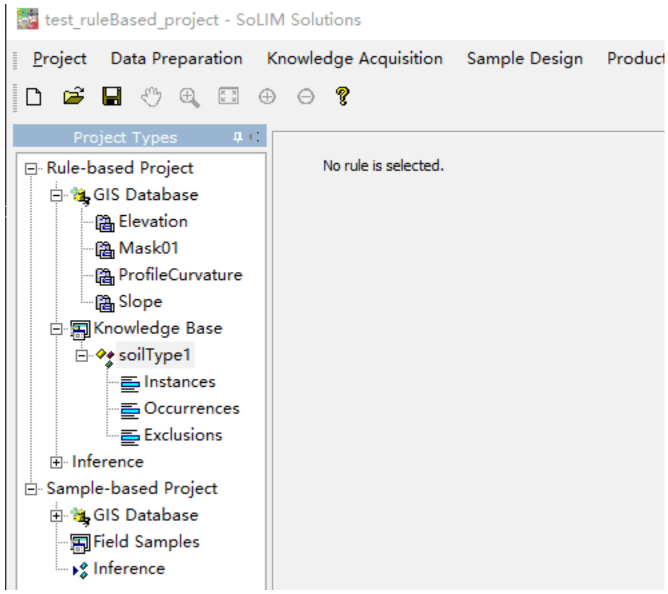
Right-click on the "Knowledge Base" node in control panel and select "Add Soil Type" from the pop-up menu.



This displays a dialog for you to specify the soil type name. Enter the soil type name in the text box. Be sure to include a unique ID before or after the soil type name. For example, "*SoilType1*" or "*23SoilType*", 1 and 23 are unique IDs. The reason for having a unique ID for each soil is that when the fuzzy membership are hardened to produce a soil type map, the unique ID in the soil name will be used as the ID for the soil type in the resulting hardened map.



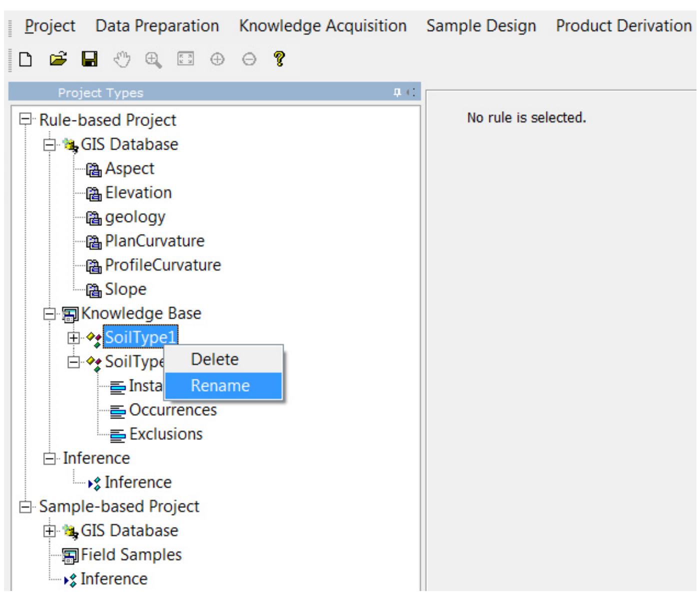
The new soil type will be appended to the end of the list in "Knowledge Base" Node. Unfold the newly created soil type node, you will see three sub-nodes: Instances, Occurrences, Exclusions. They are used to hold different kinds of knowledge.



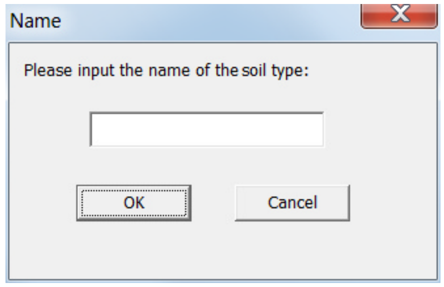
Tips: if a soil type name already exists, the program will bring up a warning dialog. You cannot add a soil type with the same name.

Rename a Soil Type

To change the name of soil type, right click on the soil type node you want to modify and select "Rename".



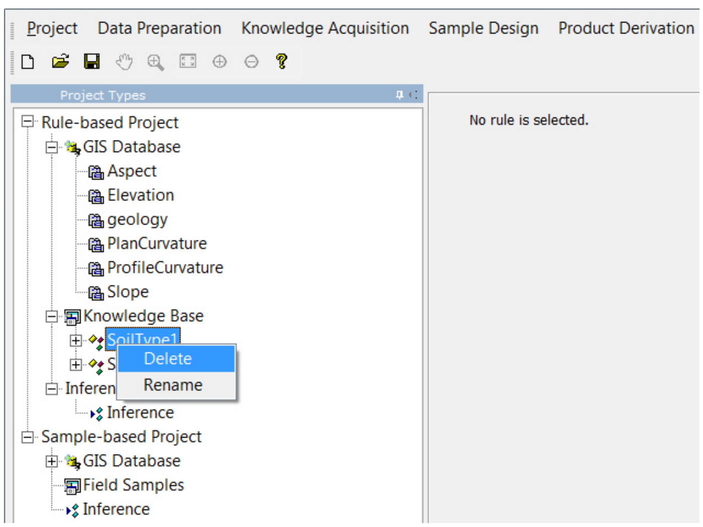
This displays a dialog for you to input the soil type name. Be sure to include a unique ID before or after the soil type name. For example, "SoilType1" or "23SoilType ", 1 and 23 are unique IDs.



Tips: if the soil type name already exists, the program will bring up a warning dialog. You cannot add a soil type with the same name.

Remove a Soil Type

Right-click on the soil type you want to remove and select "Delete" to remove the soil type.



Tips: This will permanently remove all instances, occurrences and exclusions of this soil type from the knowledgebase.

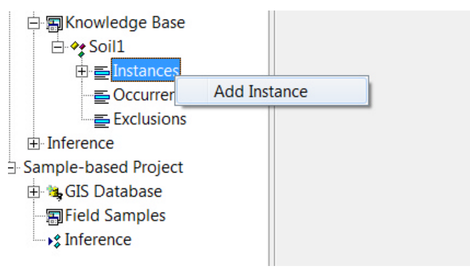
Instance

An instance is a representation of the soil scientist's knowledge of the relationship between a soil type and its environmental conditions characterized by topography, geology, climate, vegetation, and other environmental variables. It is not explicitly location-specific and is applicable to the whole mapping area, thus is considered to be the global knowledge. Note that a given soil type might appear in several instances, for example, a soil type may occur on both south facing and north facing slopes with different elevation and slope gradient conditions. These two environmental configurations form two instances of the soil. Technically, an instance eventually refers to an environmental configuration defined by one or more rules, each of which corresponds to one environmental variable and characterizes the relationship between the optimality of the soil type and the environmental variable.

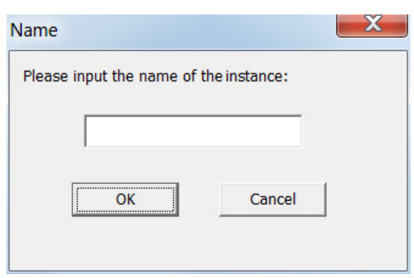
An instance integrates all the rules under it to infer the optimality value of this instance to a soil type. The inference method is discussed in detail in Soil Inference section.

Create a New Instance

Right-click on the "Instances" node under soil type node for which you want to create a new instance and select "Add Instance".



A dialog will appear for you to input the instance name.

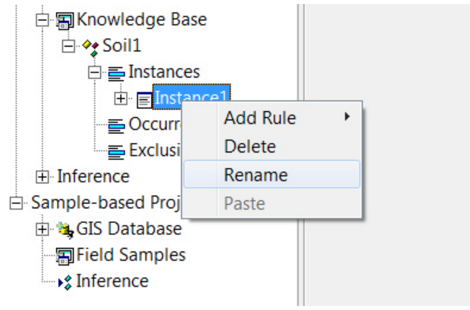


Be sure to use a unique name. After you input the name and click on OK, the new instance is appended to the end of "Instances" list under the current soil type. The newly-created instance contains no rule(s) , you can add ,delete, rename rules (See Rule Management).

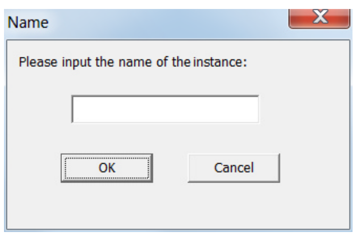
Tips: if the instance name already exists under a soil type, the program will bring up a warning dialog. You cannot add an instance with the same name.

Rename an Instance

To change the name of an instance, right click on the instance node you want modify and select "Rename".



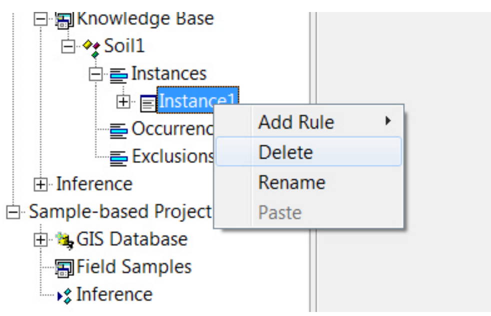
A dialog will appear for you to input the instance name. Be sure to use a unique name.



Tips: if the instance name already exists, the program will bring up a warning dialog. You should use a different name.

Remove an Instance

Right-click on the instance you want to remove and select "Delete" to remove the instance. Note: This will permanently remove all rules for this instance from the knowledgebase.

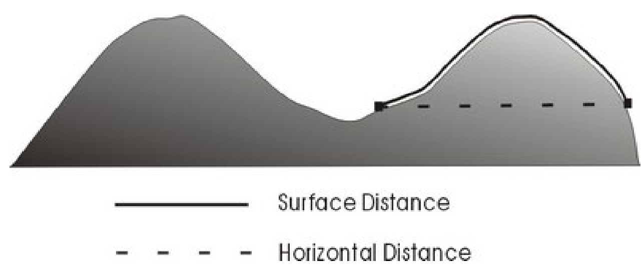


Occurrence

An occurrence is a **positive exception**, which means a particular soil will occur in places where the global knowledge does not cover. It is a representation of the soil scientist's knowledge of the relationship between a soil type and its environmental conditions characterized by topography, geology, climate, vegetation, and other environmental variables **within limited geographic area**. The area is called "influenced area" in SoLIM Solutions. In the influenced area, the relationship between soil type and the environment is different from other places. The influence area is defined by a central location and an impact radius. Technically, in SoLIM Solutions, the knowledge for an occurrence contains one or more rules (rule part) and a spatial setting. When creating occurrences, the soil scientist needs to define rules that characterizes the relationship between the optimality of the soil type and the environmental variables. Besides, they need to pinpoint the coordinates of central location and provide the impact radius and a distance decay factor which determines how the optimality value will change as the distance to central location changes.

Some concepts:

- Central Location: It defines the central point of an occurrence.
- Impact Radius: It is the distance for defining a buffer zone around the central point. If there is a DEM layer in GIS database (the program will look through the GIS database, if a layer with name "Elevation", "ELEVATION" or "elevation" is found, it will be taken as a DEM layer), the program will use the surface distance, that is, the distance along the terrain surface. This may result in an influenced area with an irregular shape. The unit of the impact radius is the unit used by the environmental database. Otherwise, it will use the horizontal distance between two locations.

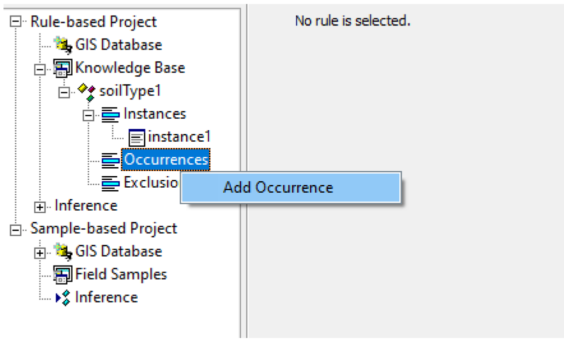


- Distance Similarity: It means that the closer a pixel to the central location of an occurrence, the more impact the occurrence tends to have on the optimality of the pixel to be inferred.
- Distance Decay Factor: This parameter is optional. If we use distance similarity, this parameter will decide how the distance to the central point impacts the optimality value.

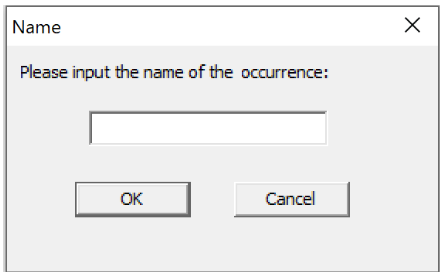
If a pixel falls into the influenced area of an occurrence, the program will integrate all rules under the occurrence and the spatial setting to infer the optimality value. The inference method is discussed in detail in Soil Inference.

Create a New Occurrence

Right click on the "Occurrences" node under the soil type node for which you want to create a new occurrence and select "Add Occurrence".



A dialog will appear for you to input the occurrence name.



Be sure to use a unique name. After you input the name and click on OK, the Spatial Setting dialog will appear. Edit the spatial setting and click on "OK", the new occurrence is appended to the end of the "Occurrences" list under the current soil type. The newly created instance contains no rule(s), you can add, delete, rename rules.

Tips: if the occurrence name already exists under a soil type, SoLIM will bring up a warning dialog. You cannot add an occurrence with the same name.

Rename an Occurrence

To change the name of an occurrence, right click on the occurrence node you want modify and select "Rename".

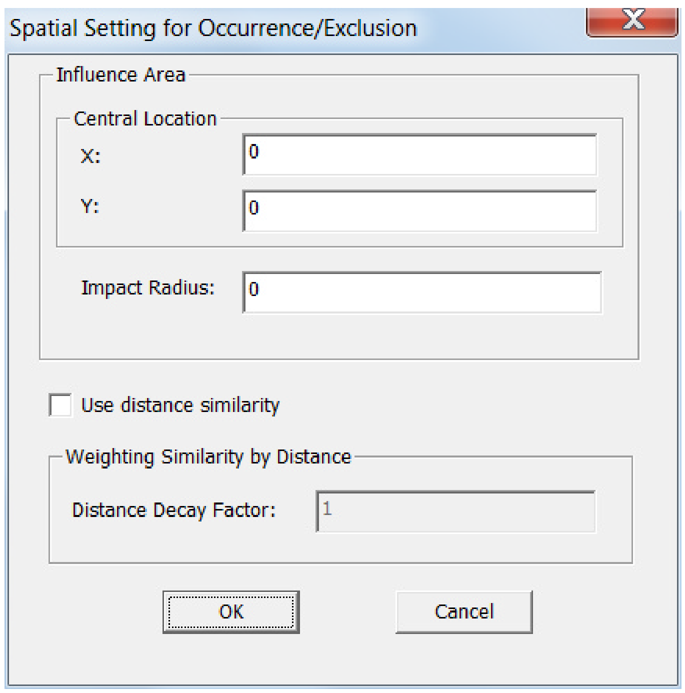
Tips: Be sure to use a unique name. if one occurrence name already exists, the program will bring up a warning dialog.

Remove an Occurrence

Right click on the occurrence you want to remove and select "Delete" to remove the occurrence. This will permanently remove all rules and spatial setting for this occurrence from the knowledgebase.

Edit Spatial Setting

When you create a new occurrence, the spatial setting dialog will appear to let you edit spatial settings for the occurrence. You can modify spatial setting at any time by click on the occurrence node which will also bring up the dialog.



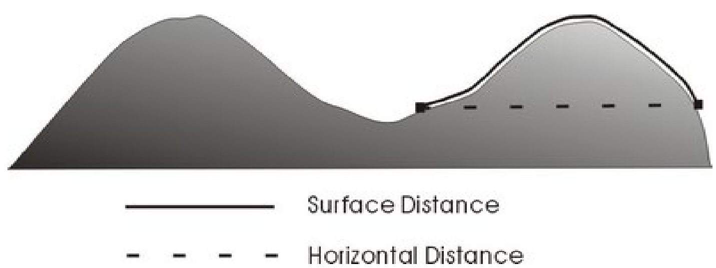
In the spatial setting dialog, specify *x* and *y* coordinates of central location of the current occurrence. Enter impact radius in the corresponding text box. If you want to perform a distance adjustment, check "Use distance similarity" and specify value for distance decay factor in the corresponding text box. The default distance decay factor is 1. If you do not want to perform distance adjustment, uncheck "Use distance similarity".

Exclusion

An exclusion is a negative **exception**, which means a particular soil will be very **unlikely to occur** in some places where the global knowledge does cover. It is a representation of the soil scientist's knowledge of the relationship between a soil type and its environmental conditions characterized by topography, geology, climate, vegetation, and other environmental variables **within limited geographic area**. The area is called "influenced area" in SoLIM Solutions. In the influenced area, the relationship between soil type and the environment is different from other places. The influenced area is defined by a central location and an impact radius. Technically, in SoLIM Solutions, the knowledge for an exclusion contains one or more rules (rule part) and a spatial setting. When creating exclusion, the soil scientist needs to define rules that characterizes the relationship between the optimality of the soil type and the environmental variable. Besides, they need to pinpoint the coordinates of central location and provide the impact radius and a distance decay factor which determines how the optimality value will change as the distance to the central location changes.

Some concepts:

- Central Location: It defines the central point of an exclusion.
- Impact Radius: It is the distance for defining a buffer zone around the central point. If there is a DEM layer in GIS database (The program will look through the GIS database, if a layer with name "Elevation", "ELEVATION" or "elevation" is found, it will be taken as a DEM layer), the program will use the surface distance, that is, the distance along the terrain surface. This may result in an influenced area with an irregular shape. The unit of the impact radius is the unit used by the environmental database. Otherwise, it will use the horizontal distance between two locations.



- Distance Similarity: It means that the closer a pixel to the central location of an exclusion, the more impact the exclusion tends to have on the optimality of the pixel to be inferred.
- Distance Decay Factor: the parameter is optional. If we use distance similarity, this parameter will decide how the distance to the central point impacts the optimality value.

If a pixel falls into the influenced area of an exclusion, the program will integrate all rules under the exclusion and the spatial setting to infer the optimality value. The inference method is discussed in detailed Soil Inference.

Create a New Exclusion

Right click on the soil type for which you want to create a new exclusion and select "Add exclusion". A dialog will appear for you to input the exclusion name.

Be sure to use a unique name. After you input the name and click on OK, the new exclusion is appended to the end of the "exclusion" list under current soil type. The newly-created exclusion contains no rule(s), you can add, delete, and rename rules.

Tips: If the exclusion name already exists under a soil type, SoLIM will bring up a warning dialog. You cannot add an exclusion with the same name.

Rename an Exclusion

To change the name of an exclusion, right click on the exclusion node you want to modify and select "Rename". Be sure to use a unique name.

Tips: If the exclusion name already exists, SoLIM will bring up a warning dialog.

Remove an Exclusion

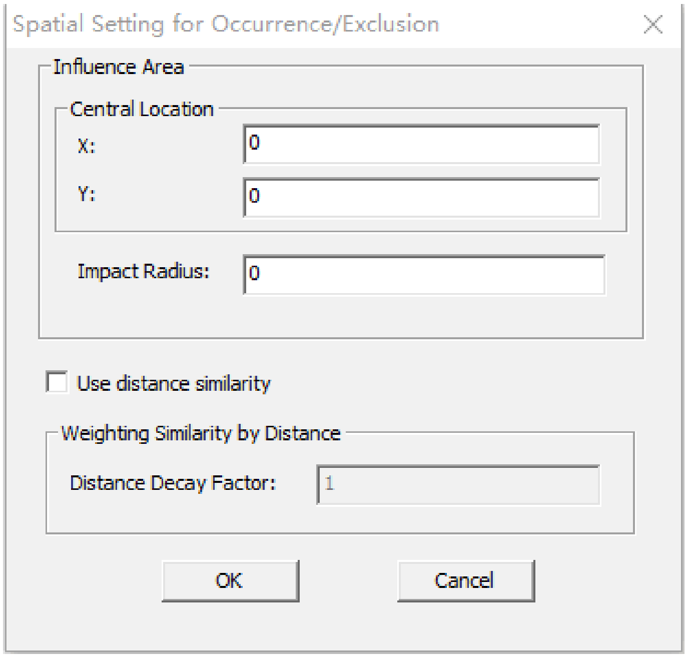
Right-click on the exclusion you want to remove and select "Delete" to remove the exclusion.

Note: This will permanently remove all rules and spatial setting for this exclusion from the knowledgebase.

Edit Spatial Setting

When you create a new exclusion, spatial setting dialog will appear to let you edit spatial setting information. You can modify spatial setting at any time by click on the exclusion node which will also bring up the spatial setting dialog.

In spatial setting dialog, specify *x* and *y* coordinates of central location of the current exclusion. Enter impact radius in the corresponding text box. If you want to perform a distance adjustment, check "Use distance similarity" and specify value for the distance decay factor in the corresponding text box. The default distance decay factor is 1. If you do not want to perform distance adjustment, uncheck "Use distance similarity".



Rule Management

A rule is a function that defines how changes in environmental condition affects the optimality of that environment for a soil type.

A rule is the basic unit of a knowledge database. Soil scientists may provide their knowledge in the form of optimality functions directly, or they may use words to express the knowledge on the relationship between soil type and environmental condition, or they may identify some locations as places where the soil is typical for the given soil type using a topographic map, DEM, or orthophoto. SoLIM Solutions provides different rule types to support different kinds of knowledge and organizes them in a uniform framework.

Rule management deals with the adding, deleting, renaming, and editing rules.

Different Rule Types

Currently, SoLIM Solutions provides five different kinds of rules to support different forms of knowledge

- Range Rule
- Freehand Rule
- Word Rule
- Enumerated Rule
- Point Rule

This section explains these different rule types, including the parameters they need and their usages.

Range Rule

Range rules are applicable to environmental features with interval or ratio values (e.g. temperature, slope). The basic idea is when certain condition is satisfied, the optimality value is 1, as the value deviates from the optimality value, the optimality decreases. This idea is expressed as the following function:

$$f(x) = e^{\left(\frac{x-b}{d}\right)^2 ln(0.5)}$$

$f(x)$: the optimality value

x: the environmental variable value in the location to be inferred

b: the environmental variable value at which the optimality for the soil type is the highest (the most ideal)

d: the difference between b and the environmental variable value at which the optimality value is 0.5 (soil is less typical of the specific soil type)

From the function above, some concepts can be derived:

Low Unity and High Unity:

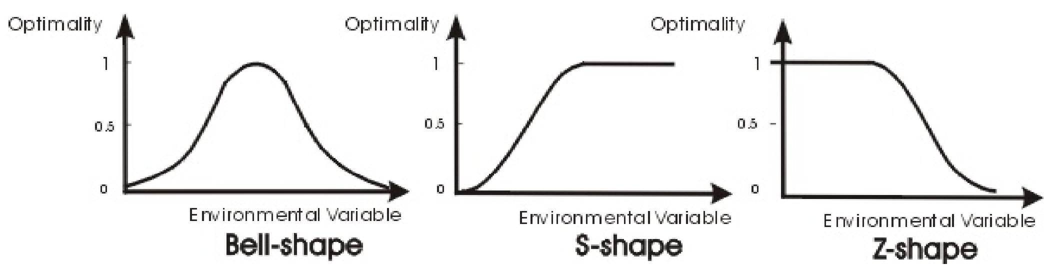
The highest optimality value (b) can be a single value or can be a range. Low unity is defined to be the "minimum b" while high unity is the "maximum b". So low unity and high unity being the same means b is a single point, otherwise, a range.

Low Cross and High Cross:

Low Cross and high cross are the environmental variable values when optimality value decreased to 0.5. Low Cross is smaller than b while High Cross is greater than b.

We can derive three kinds of curves from the basic function:

- Bell-shape curve: For bell-shaped curve, optimality value decreases when environmental variable value is lower than low unity or higher than high unity. Low unity and high unity can be the same or different values. If they are equal, there is no flat top. If the difference between low unity and low cross and the difference between high cross and high unity is the same, the curve is symmetric. Otherwise, it is unsymmetrical.
- S-shaped curve: S-shape curve can be created by setting the optimality value to 1 for x greater than b. It is a "the-higher-the-better" shape. The low unity and high unity are the same. High cross is not needed for S-shape curve. Optimality will be unity only if the environmental variable values are greater than low unity/high unity. For example, you may want a rule that indicates a particular soil is found at high elevations only. Elevations below low unity/high unity are sub-optimal, and elevations that are above low unity/high unity are optimal for the soil type.
- Z-shaped curve: Z-shape curve can be created by setting the optimality value to 1 for x less than b. It is a "the-lower-the-better" shape. The low unity and high unity are the same. Low cross in not needed for Z-shape curve. Z-shape curve is entirely analogous to the s-shape, except that the values lower than low unity/high unity are optimal.



To define a range rule, the curve type, the low unity, the high unity, the low cross and the high cross need to be specified by users.

Freehand Rule

Freehand rule is flexible and can express a relationship that is very complex. It is possible that none of bell-shaped, Z-shaped, and S-shaped can effectively express the relationship. For example, there might be more than one ideal conditions, so the curve may have more than one peak. To support this kind of knowledge, we need a freehand rule.

A freehand rule is used when soil scientists can provide some key points whose optimality value are known. If three or more key points are provided, the program will interpolate these points into a curve using cubic spline function.

When the key points are deleted, modified, or moved, the curve will change accordingly. Soil scientist can keep adjusting the curve by deleting or/and moving the points until they are satisfied with the result.

To define freehand rules, at least three key points should be provided so that SoLIM Solutions can interpolate them into a curve.

Word Rule

Sometimes, soil scientists can only provide qualitative description, that is, they may use words (such as "convex". "linear", "concave" for curvature variables) to express their knowledge. Word rule is designed for this kind of knowledge.

Word rule contains a word and its corresponding word library. The word is descriptive, such as "convex". "linear", "concave" etc., while the word library is a pool of references, where every word has a curve associated with it. So, when choosing a word, you can get a relationship curve from the word library.

To define a word rule, you need to firstly construct word library. A word library is recorded as a XML file. A typical word library file has a hierarchical structure. The first level is CurveLib, the second level is the environmental variable, and the third level is the description of the curve, including the word, key points number and coordinates of key points on the curve.

Here is a sample word library file:

```
<?xml version="1.0" encoding="UTF-8"?>
<CurveLib>
    <EnvAttri Name="Planform Curvature">
        <Curve>
          <Name>gibbous</Name>
          <NodeNum>4</NodeNum>
          <Coordinates>-0.33 0.2, -0.2 0.4, -0.05 0.8, 0.13 0.3</Coordinates>
        </Curve>
```

```
        <Curve>
          <Name>linear</Name>
          <NodeNum>5</NodeNum>
          <Coordinates>-0.25 0.2, -0.1 0.4, 0 0.5, 0.05 0.8, 0.15 0.9</Coordinates>
        </Curve>
    </EnvAttri>
    <EnvAttri Name="Profile Curvature">
        <Curve>
          <Name>linear</Name>
          <NodeNum>6</NodeNum>
        <Coordinates>-0.006 0.3, -0.003 0.5, 0.001 0.7, 0 0.8, 0.002 0.9, 0.003 0.3</Coordinates>
      </Curve>
      <Curve>
          <Name>concave</Name>
          <NodeNum>4</NodeNum>
          <Coordinates>-0.004 0.2, -0.0005 0.4, 0.001 0.8, 0.005 0.3</Coordinates>
        </Curve>
    </EnvAttri>
</CurveLib>
```

When a word library is prepared in the format that SoLIM Solutions can support, you need to choose a word from the library to define the word rule, and then choose one linguistic term.

Enumerated Rule

Enumerated rules are used for categorical environmental variables, e.g. geology, land use type. All the possible conditions when certain soil type occurs are listed. The conditions are expressed as items (integer numbers). If the actual environmental condition matches any of the items in this list, the optimality value yielded by this rule will be 1; otherwise, the value will be 0.

To define an enumerated rule, the items that will yield optimality value 1 should be provided.

Point Rule

Soil scientists sometimes are not able to give optimality values, but they know where the typical location is (a typical location is the place where soil is almost the typical case of a certain soil type). The value of environmental variable in the typical location can be viewed as the value that has the highest optimality value for the soil type. By providing typical locations, soil scientists describe their knowledge implicitly. Point rule is designed to support this kind of knowledge. The point was referred to as tacit point in earlier versions of SoLIM software.

Point rule is very similar to range rule. The basic function underlying them are the same:

$$f(x) = e^{\left(\frac{x-b}{d}\right)^2 ln(0.5)}$$

$f(x)$: the optimality values

x: the environmental variable value in the location to be inferred

b: the environmental variable value at which the optimality for the soil type is the highest (the most ideal)

d: the difference between b and the environmental variable value at which the optimality value is 0.5 (soil is less typical of the specific soil type)

The only difference between range rule and point rule is how *b* is determined: *b* of a range rule is directly specified by the user, whereas *b* of a point rule is identified by the inference engine according to the coordinate of the point specified. There is no low unity, high unity, low cross, or high cross for a point rule. Instead, it has four parameters: central X, central Y, left width and right width.

Central X and Central Y:

The coordinates of the point where the soil is the most typical. They will decide the value of *b*. It must be noted that in a point rule, *b* can only be a single value.

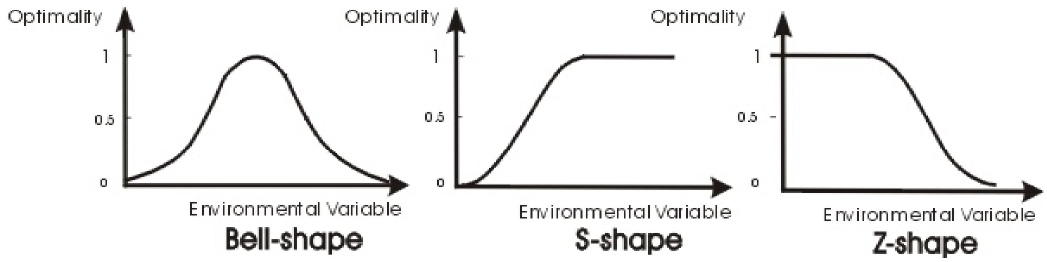
Left Width and Right Width:

Left width is the parameter *d* in the function when x<*b*, whereas right width is the parameter *b* when x > *b*.

Like range rule, point rule also has three types of curves.

- Bell-shape curve: For a bell-shaped curve, optimality values decrease away from the idea condition. If the left width and right are the same, the curve is symmetric. Otherwise, it is unsymmetrical.
- S-shaped curve: S-shape curve can be created by setting the optimality value to 1 for x greater than b. It is a "the-higher-the-better" shape. Right width is not needed for S-shape curve.

- Z-shaped curve: Z-shape curve can be created by setting the optimality value to 1 for x less than b. It is a "the-lower-the-better" shape. Right width in not needed for Z-shape curve.

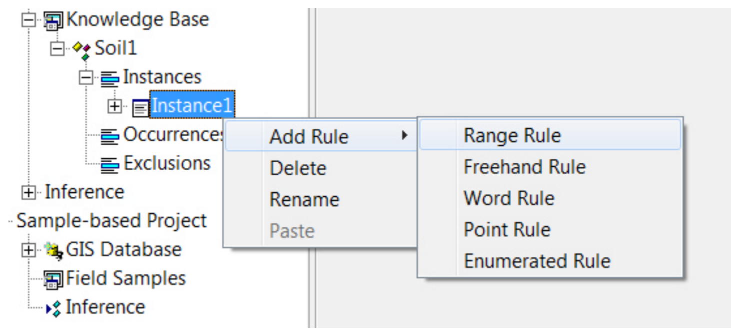


Point rule records the x-y coordinates (spatial location) of the typical point, and the left width and right width. Its optimality value is determined by the x-y coordinates rather than the direct input.

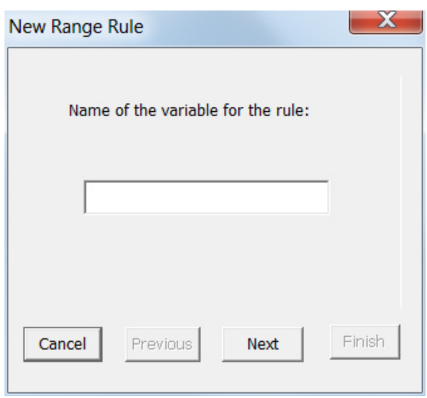
Add a Rule

Add a Range Rule

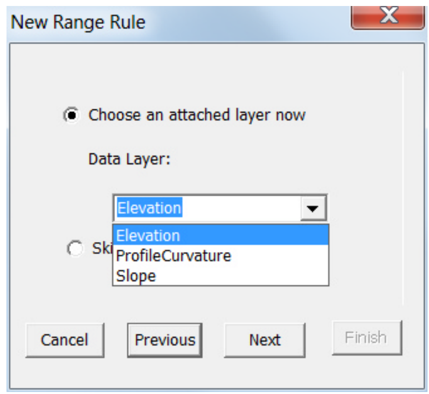
1. Right click on the instance/occurrence/exclusion which will hold the new rule. In the pop-up menu, select "Add Rule", then select "Range Rule".



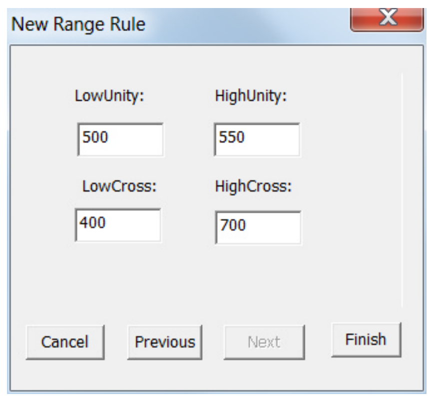
2. Enter the name of the variable for rule (e.g. elevation) in the text box. Click on "Next".



3. If you want to attach a layer now (recommended), check "Choose an attached layer now". The drop-down list will list all layers currently available in the GIS Database. Select the layer that you want to attach with the new rule. If you do not want to attach a layer now, check "Skip this step" and then click on "Next".



4. Select a curve type in the combo box. Then click on "Next".



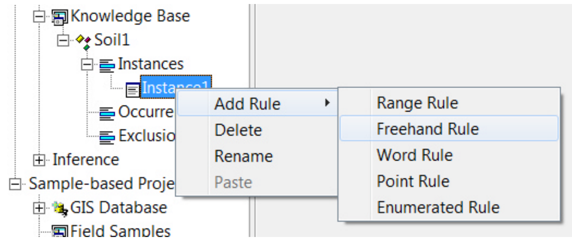
5. If you select an attached layer in step 3, the program will use the GIS layer to set default rule parameters and just click on "Finish" to finish adding the range rule. Otherwise, you should provide low unity, high unity, low cross, high cross and then click on "Finish" to finish adding the range rule.

Tips:

1. If you want to go back to the previous step, click on "Previous".
2. If you want to give up adding the rule, click on "Cancel".
3. It is not recommended to change the GIS layer during the add rule step. It is recommended to add rules to other layers after adding rules to one GIS layer.

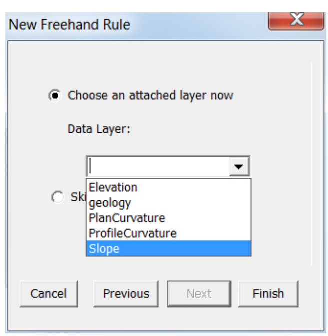
Add a Freehand Rule

1. Right click on the instance/occurrence/exclusion which will hold the new rule. In the pop-up menu, select "Add Rule", then select "Freehand Rule".



2. Enter the name of the variable for rule (e.g. elevation) in the text box. Click on "Next".

3. If you want to attach a layer now (recommended), check "Choose an attached layer now". The drop-down list will list all layers in current GIS Database. Select the layer that you want to attach with the new rule. If you do not want to attach a layer now, check "Skip this step" and then click on "Finish" to finish adding the new freehand rule.

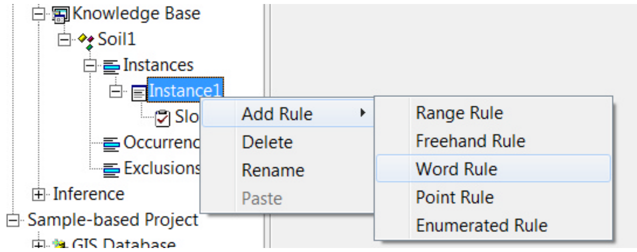


Tips:

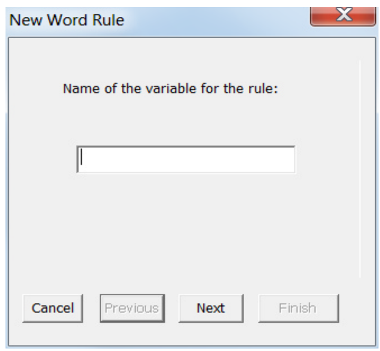
1. If you want to go back to the previous step, click on "Previous".
2. If you want to give up adding the rule, click on "Cancel".
3. It is not recommended to change the GIS layer during the add rule step. It is recommended to add rules to other layers after adding rules to one GIS layer.

Add a Word Rule

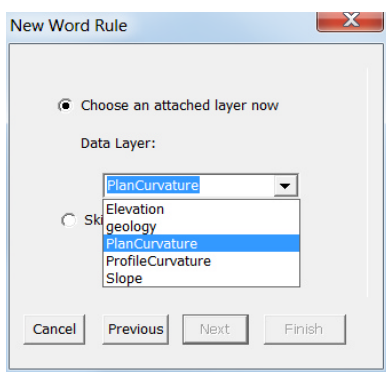
1. Right click on the instance/occurrence/exclusion which will hold the new rule. In the pop-up menu, select "Add Rule", then select "Word Rule".



2. Enter the name of the variable for rule (e.g. elevation) in the text box. Click on "Next".

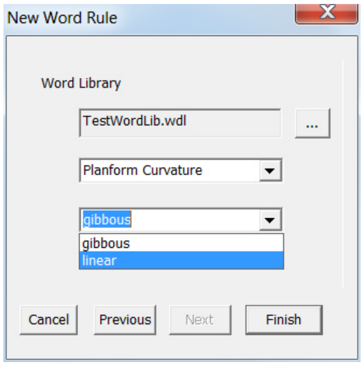


3. If you want to attach a layer now (recommended), check "Choose an attached layer now". The drop-down list will list all layers in current GIS Database. Select the layer that you want to attach with the new rule. Click on "Next". If you do not want to attach a layer now, check "Skip this step" and then click on "Next".



4. Select a word library (.lib file), Then the environmental variables will be listed. Select an environmental variable, then the words available in the word library expressing the knowledge on that environmental

variable will be listed. Select the one you need, then click on "Finish" to finish adding the new word rule.

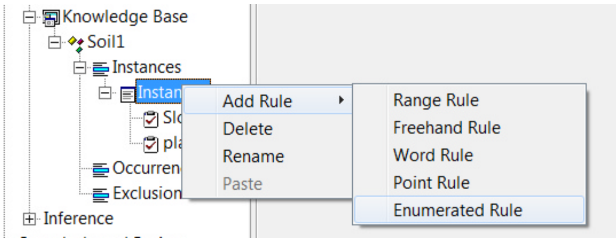


Tips:

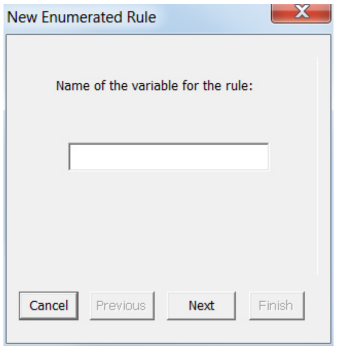
1. If you want to go back to the previous step, click on "Previous".
2. If you want to give up adding the rule, click on "Cancel".
3. It is not recommended to change the GIS layer during the add rule step. It is recommended to add rules to other layers after adding rules to one GIS layer.

Add an Enumerated Rule

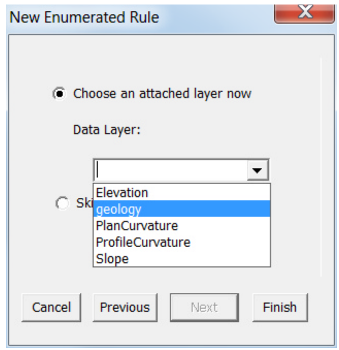
1. Right click on the instance/occurrence/exclusion which will hold the new rule. In the pop-up menu, select "Add Rule", then select "Enumerated Rule".



2. Enter the name of the variable for rule (e.g. geology) in the text box. Click on "Next".



3. If you want to attach a layer now (recommended), check "Choose an attached layer now". The drop-down list will list all layers in current GIS Database. Select the layer that you want to attach with the new rule and then click on "Finish" to finish adding the new enumerated rule. If you do not want to attach a layer now, check "Skip this step" and then click on "Finish" to finish adding the new enumerated rule.

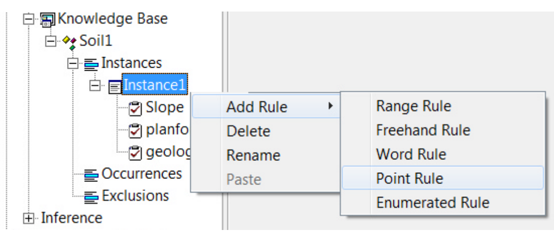


Tips:

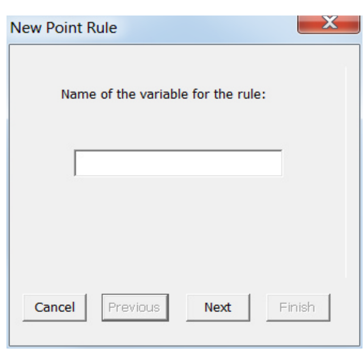
1. If you want to go back to the previous step, just click on "Previous".
2. If you want to give up adding the rule, click on "Cancel".
3. It is not recommended to change the GIS layer during the add rule step. It is recommended to add rules to other layers after adding rules to one GIS layer.

Add a Point Rule

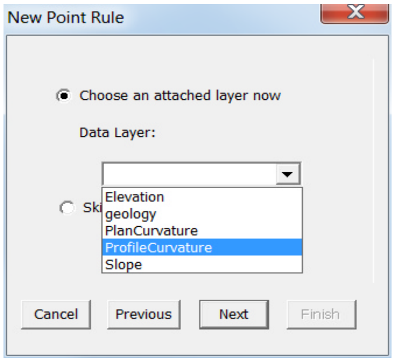
1. Right click on the instance/occurrence/exclusion which will hold the new rule. In the pop-up menu, select "Add Rule", then select "Point Rule".



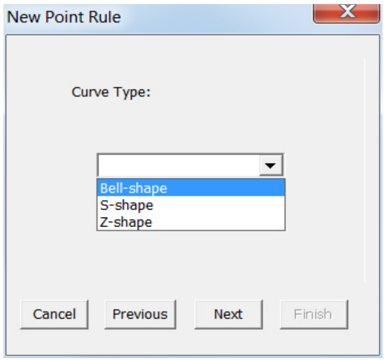
2. Enter the name of the variable for rule in the text box. Click on "Next".



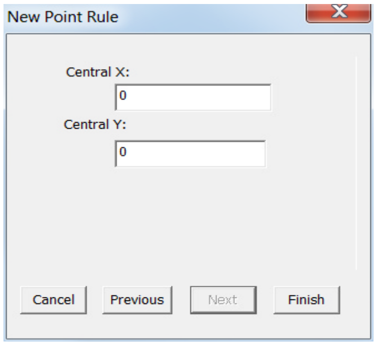
3. If you want to attach a layer now (recommended), check "Choose an attached layer now". The drop-down list will list all layers in current GIS Database. Select the layer that you want to attach with the new rule. If you do not want to attach a layer now, check "Skip this step" and then click on "Finish" to finish adding the new enumerated rule.



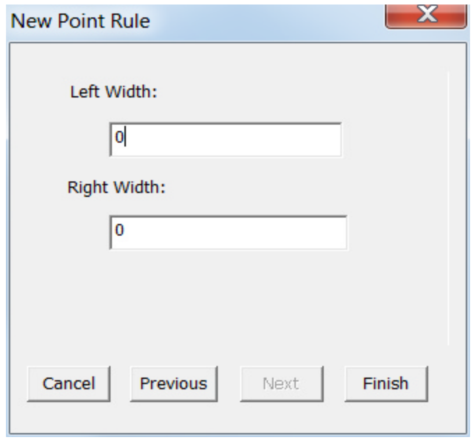
4. Select a curve type in the combo box. Then click on "Next".



5. Enter the coordinates (the spatial reference should be the same with the GIS database) of the central point in the text boxes. Click on "Next".



6. If you select an attached layer in step 3, as the program will use the GIS layer to set default rule parameters, click on "Finish" to finish adding the new point rule. Otherwise, you should provide left width, right width and then click on "Finish".



Tips:

1. If you want to go back to the previous step, click on "Previous".
2. If you want to give up adding the rule, click on "Cancel".
3. It is not recommended to change the GIS layer during the add rule step. It is recommended to add rules to other layers after adding rules to one GIS layer.

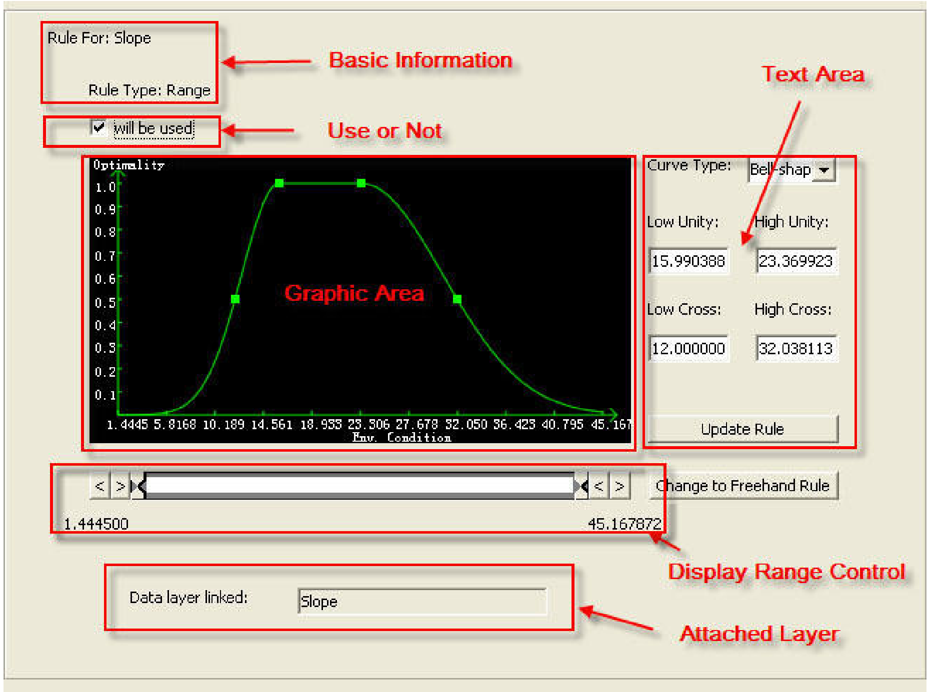
Edit a Rule

SoLIM Solutions provides interface for you to edit different kinds of rules conveniently. Many kinds of rules can be edited in both numerical (text) and graphical modes.

- Edit a Range Rule
- Edit a Freehand Rule
- Edit a Word Rule
- Edit an Enumerated Rule
- Edit a Point Rule

When you create a new rule or click on a rule in the control panel, you will see the rule editing view. Each kind of rules has a different editing view. However, the basic components of the views are the same. The following picture shows the editing view for Range Rule.

The rule editing view contains the an area to display basic information of the rule a check box to specify whether the rule will be used in the inference a graphic area to show the corresponding graph (curve) of the rule a text area to input parameters or display more detailed information of the rule a tool to control the display range an area to display and manipulate the attached layer of rule.

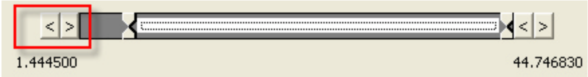


Specify Whether the Rule Is Used in the Inference

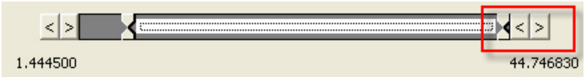
Check the box next to "will be used" to make the rule be used in the inference. Uncheck the box to disable the rule. The rule is editable only when it is in used.

Set the Display Range of the Graphic Area

Display range decides the maximum range on the x-axis of the graph area. To set the display range, use the arrows at each end of the **display range control** bar. On the left, click on "<" to decrease the low display limit, click on ">" to increase the low display limit.



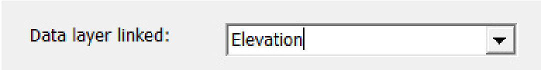
On the right, click on ">" to increase the high display limit, click on "<" to decrease the high display limit.



Current display range decides the range for the current x-axis in the graphic area. To set current displaying range, drag the handles on the display range control bar.

Select Attached Layer

If you attached a layer when creating the rule, the linked layer will be displayed.



Otherwise, the program will tell you "No data layer linked. Please set data layer".

You can attach a layer at any time by using the drop-down list. You have two options here:

1. choose one existing layer to attach.
2. choose one layer which is not in the GIS database. In this case, if the selected layer file is valid and the name of the layer is not any of these of the existing layers in GIS database, SoLIM will add the layer to GIS database, and the program will also show that the layer has already been attached with current rule. If the selected layer is valid, but there is already one layer with the same name in GIS database, the program will bring up a dialog to ask if you will replace existing layer with the new one. If you click on OK, it will replace the existing layer in the GIS Database, and the program will also show that the layer has already been attached with current rule. If the selected layer is invalid, the program will pop up an error dialog.

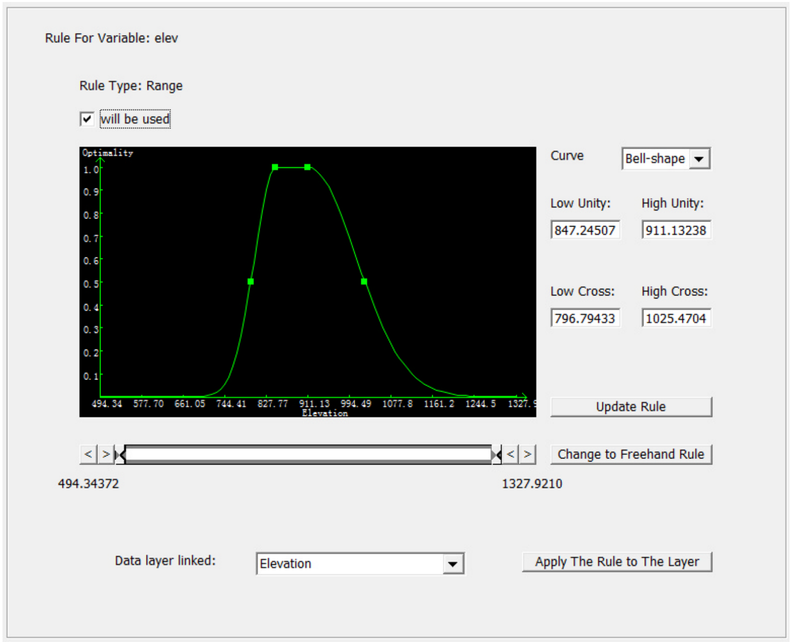
Tips: A rule will take part in inference only if it is attached with a layer.

Apply the Rule to The Layer

If the rule has an attached layer, you can see the effect of applying the rule to the layer.

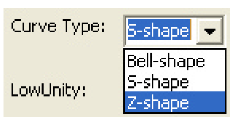
Edit a Range Rule

The interface for editing a range rule looks like this:



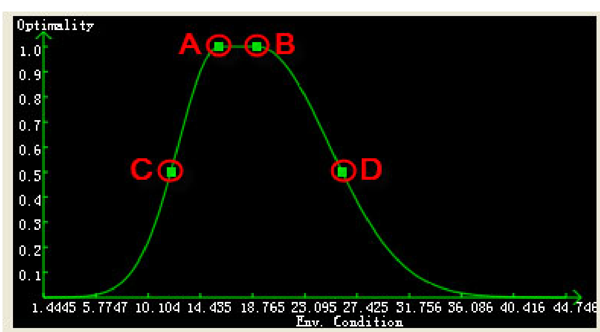
Specify Curve Type

Select a curve type from the drop-down list "Curve Type". The curve can be bell-shape, S-shape or Z-shape.



Edit the Rule

In the graphic area, the green curve is a graphic representation of the rule. There are some small square handles attached to the curve. The two square handles (A and B) on the top of the curve indicate the positions of low unity and high unity respectively; and the other two square handles (C and D) on the curve indicate the positions of low cross and high cross.



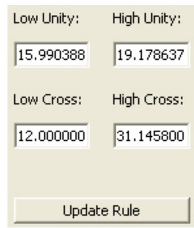
A: The location where the optimality is at unity and the environment value is low unity.

B: The location where the optimality is at unity and the environment value is high unity.

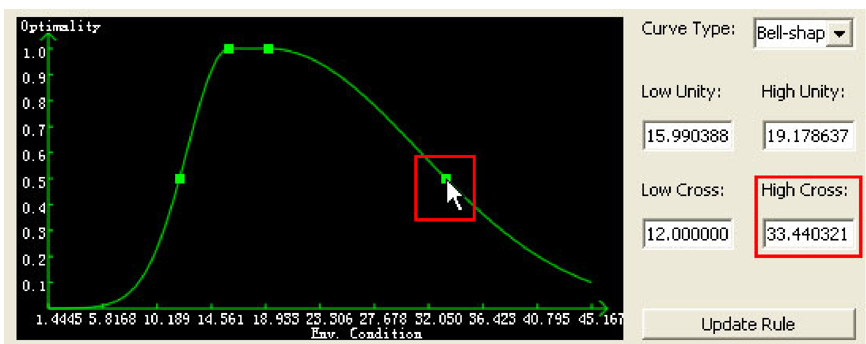
C: The location where the optimality is at half and the environment value is at low cross.

D: The location where the optimality is at half and the environment value is at high cross

You can directly enter values for low unity, high unity, low cross, high cross in the text area to adjust the rule. Click on "Update Rule" button to submit modification. This will also change the curve displayed in graphic area accordingly.



You can also use the mouse to adjust the rule in graphic area by click-and-drag. Put the cursor close enough to any of the square handles on the green curve, click on it, hold, and drag. Dragging a handle also changes the value in the corresponding field in the text area.

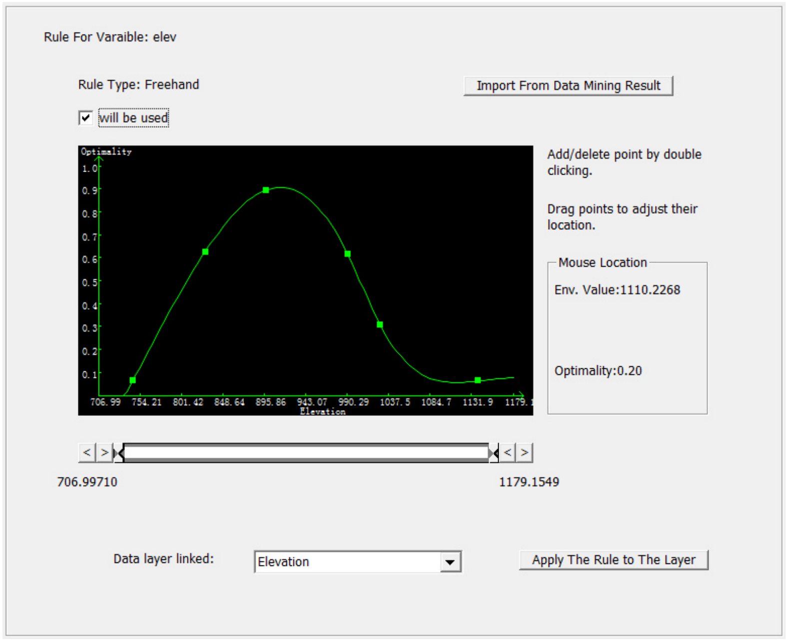


Change Range Rule to Freehand Rule

Click on "Change to Freehand Rule". Current range rule can be converted to freehand rule. You can edit the key points to edit the rule.

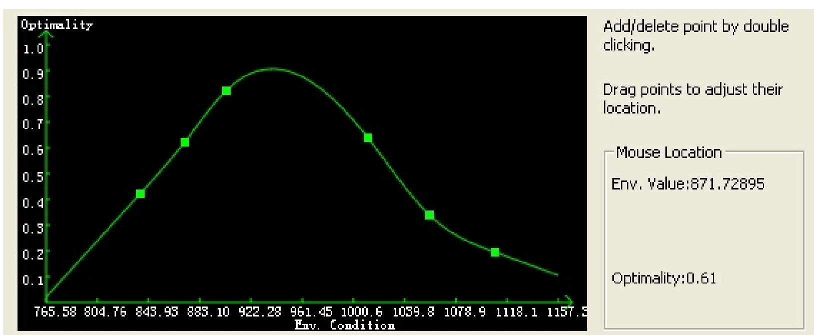
Edit a Freehand Rule

The interface for editing a freehand rule looks like this:



Edit the Rule

Freehand rule is edited by manipulating the graphic area.



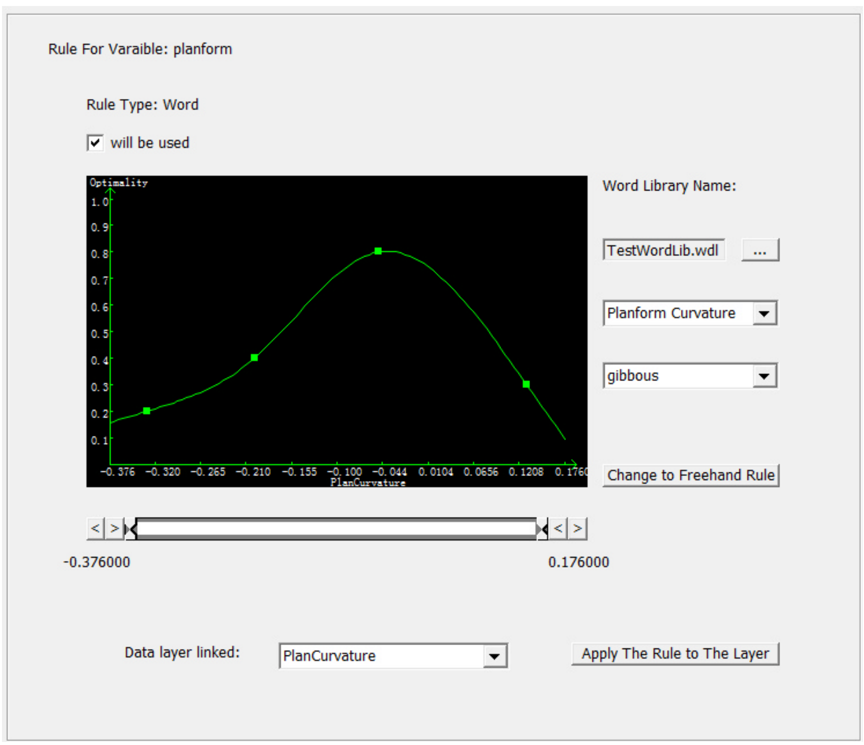
When you move you mouse in the graphic area, current mouse location is displayed in real time on the right to help you locate key points. You can edit these points with the following operations:

- To **add** a key point, double click at the location where the key point is to be added to.
- To **move** a key point, drag it to the location where you want to put it.
- To **remove** a key point, double click on it.

Tips: the interpolation method is cubic spline function. If the maximum optimality value derived from interpolation is greater than 1, it will be changed to 1 by the program. Similarly, if the minimum optimality value derived from interpolation is smaller than 0, it will be changed to 0 by the program.

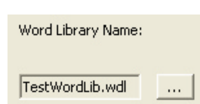
Edit a Word Rule

The interface for editing a word rule looks like this:

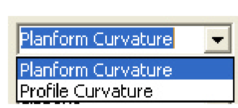


Edit the Rule

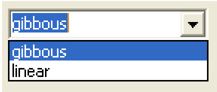
To define a word rule, you need to firstly construct a word library. After the word library file is prepared, load it by click on



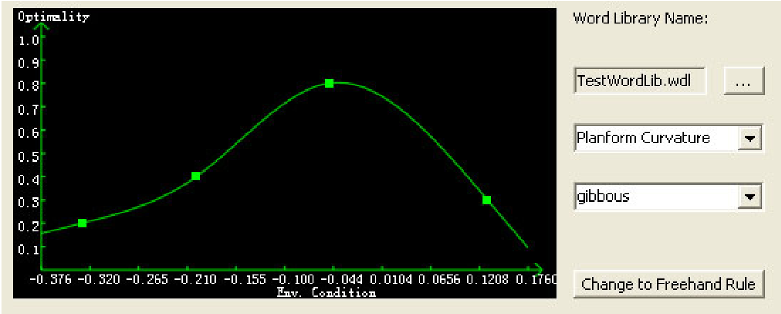
The environmental variables will be listed in the drop-down list. Select an environmental variable.



The words that express the knowledge will be listed.



Pick one word, the corresponding curve the word refers to will be displayed in graph area.

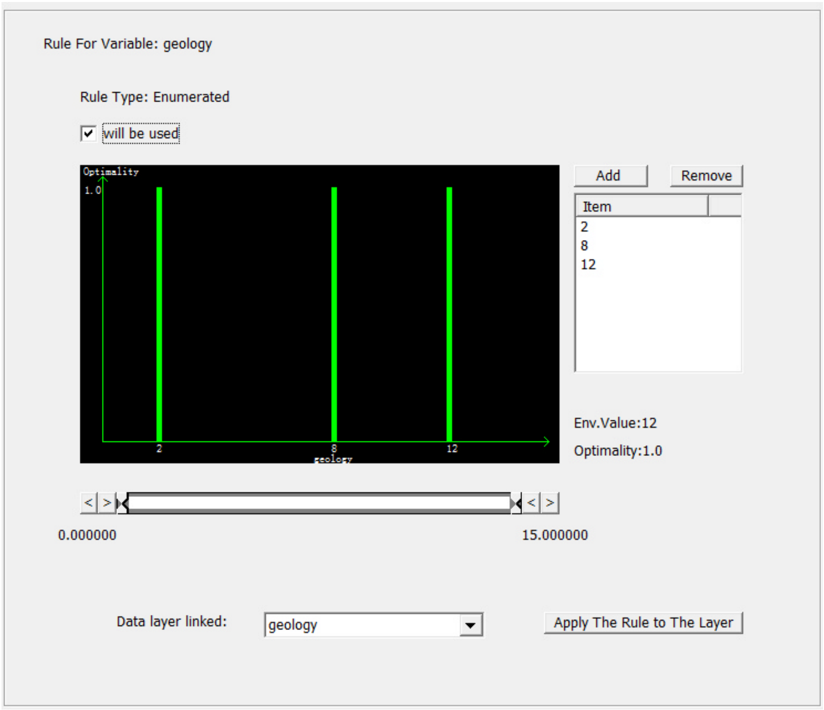


Change Word Rule to Freehand Rule

Click on "Change to Freehand Rule". Current word rule can be converted to freehand rule. You can edit the key points to edit the rule.

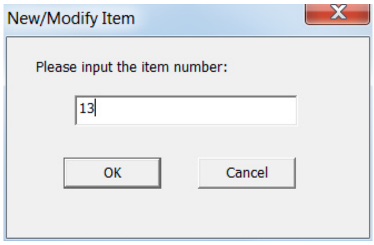
Edit an Enumerated Rule

The interface for editing a enumerated rule looks like this:

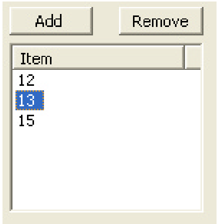


Edit the Rule

All of the current enumerated items are listed in the list box. To add an item, click on "Add". In the pop-up dialog, enter the item (an integer number). Click on "OK".

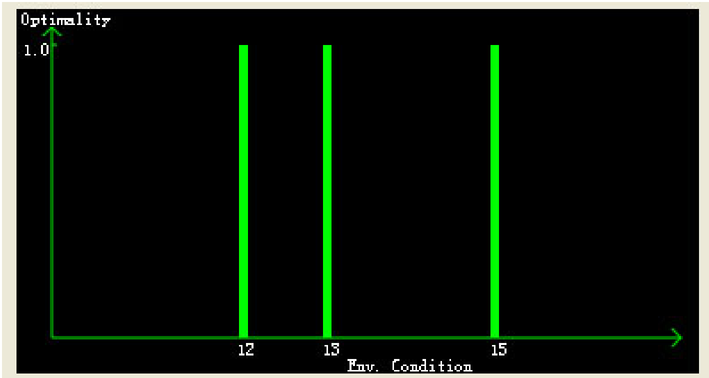


To remove an item, click on the item in the list box, then click on "Remove",

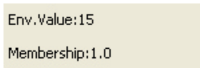


The editing will be reflected in the graphic area. You can also modify the rule in the graphic mode.

- To **change** the value of an item, drag the green bar to the position and release mouse.
- To **add** an item, double click on the location where the new item is to be added.
- To **remove** an item, double click on the bar that represents the item to be deleted.

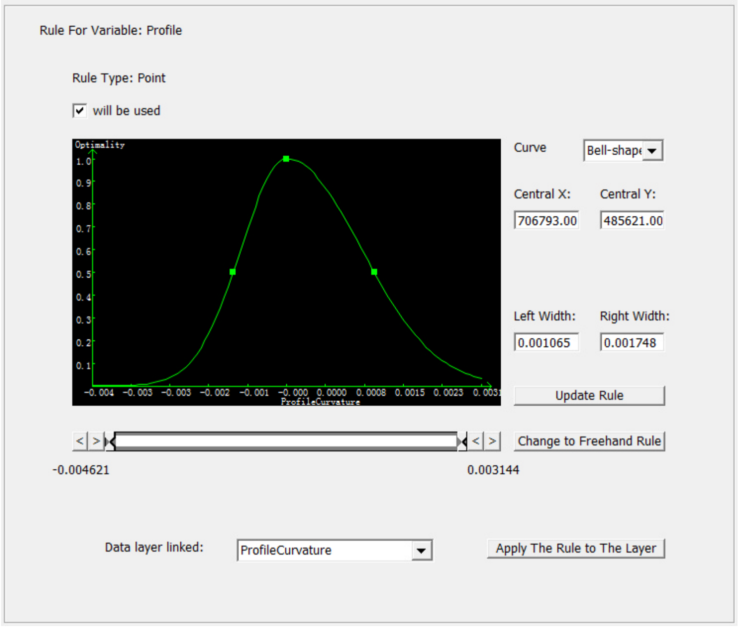


Current mouse location is displayed in real time on the right.



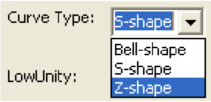
Edit a Point Rule

The interface for editing a point rule looks like this:



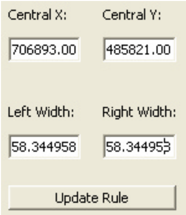
Specify Curve Type:

Select a curve type from the drop-down list "Curve Type". The curve can be bell-shape, S-shape or Z-shape.



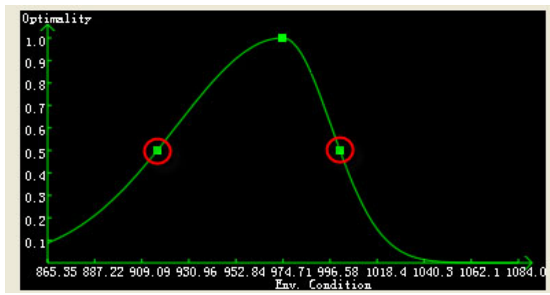
Edit the Rule

In the graphic area, the green curve is a graphic representation of the rule. There are some small square handles attached to the curve. The square handle on the top of the curve indicates the value at which the environmental variable is the most optimal for the soil type. The value is decided by the coordinates (central x, central y) of the central point. The two square handles on the sides of the curve indicate the values where optimality value is equal to 0.5. They are decided by left width and right width.



You can directly enter values for central x, central y, left width and right width in the text area to adjust the rule. When modifications are made, click on "Update Rule" button to submit. This will also cause the curve displayed in graphic area to change accordingly.

You can also use the mouse to adjust the rule in the graphic area in a click on-and-drag manner. For point rule, drag the handles for left and right width which are highlighted with right circles which will cause the values in the text area to change accordingly. It must be noted that the central point can only be modified by enter values in the text area.



Change Point Rule to Freehand Rule

Click on "Change to Freehand Rule". Current point rule can be converted to freehand rule. You can edit the key points to edit the rule.

Import Rule from Data Mining Result

In SoLIM project, Knowledge Miner.exe can extract knowledge from existing soil maps and express the knowledge as a set of curves. The format of one curve is like this:

1	*328.500000*	*0.140704*
2	*333.377777*	*0.245936*
3	*338.255554*	*0.379375*
4	*343.133331*	*0.585644*
5	*348.011108*	*0.846171*
6	*352.888885*	*0.997742*
7	*357.766663*	*0.873335*
8	*362.644440*	*0.540178*
9	*367.522217*	*0.227060*
10	*372.399994*	*0.062093*
11	*377.277771*	*0.011032*
12	*382.155548*	*0.001214*

To import a rule from data mining result, first you need to create a blank freehand rule. In the freehand rule editing view, click on "Import From Data Mining Result". In the pop-up dialog box, choose the curve file (.txt format). By doing this, the curve will be imported into SoLIM Solutions and changed into a freehand rule. So the data mining result can be used in soil inference.

Rename a Rule

1. Right click on the rule to be renamed. Select "Rename".
2. Enter the name in the dialog.
3. click on "OK" to finish renaming operation or click on "Cancel" to give up.

Tips: The name of the rule should be unique with in one instance/occurrence/exclusion.

Copy and Paste a Rule

Copy a Rule

Right click on the rule to be copied. In the pop-up menu, select "Copy", the rule will be copied into clipboard.

Paste a Rule

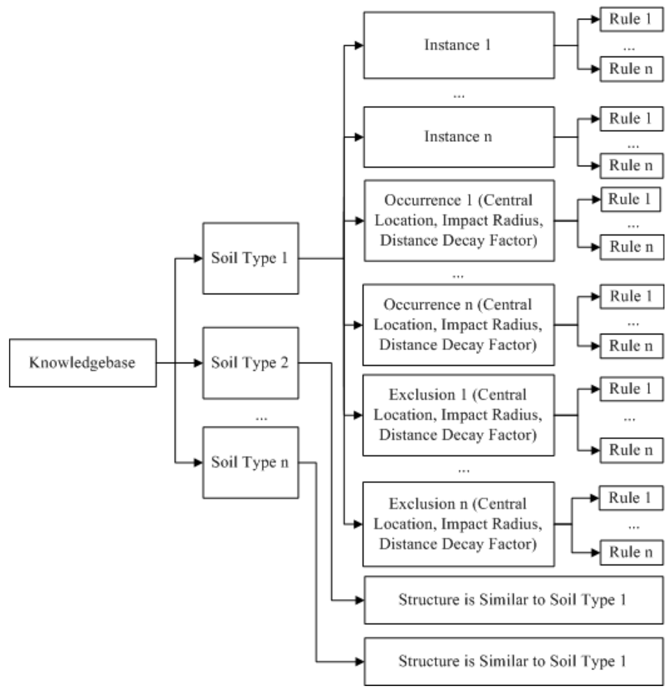
1. Right click on the instance/occurrence/exclusion which the rule is to be pasted to.
2. If there is a rule in clipboard, in the pop-up menu, the "Paste" menu item is enabled. Select "Paste" to paste the rule into the instance/occurrence/exclusion.

Delete a Rule

In the Control Panel, right click on the rule to be deleted, then select "Delete" in the pop-up menu.

Structure of Knowledge Base

A typical knowledge base has the following structure:



A knowledge base contains knowledge for one or more soil types. Under one soil type, there may be:

- one or more instances (global knowledge)
- one or more occurrences (local knowledge)
- one or more exclusions. (local knowledge)

Instance, occurrence and exclusions contain one or more rules.

Occurrences and exclusions have spatial setting (central point, impact radius, distance decay factor) associated with them.

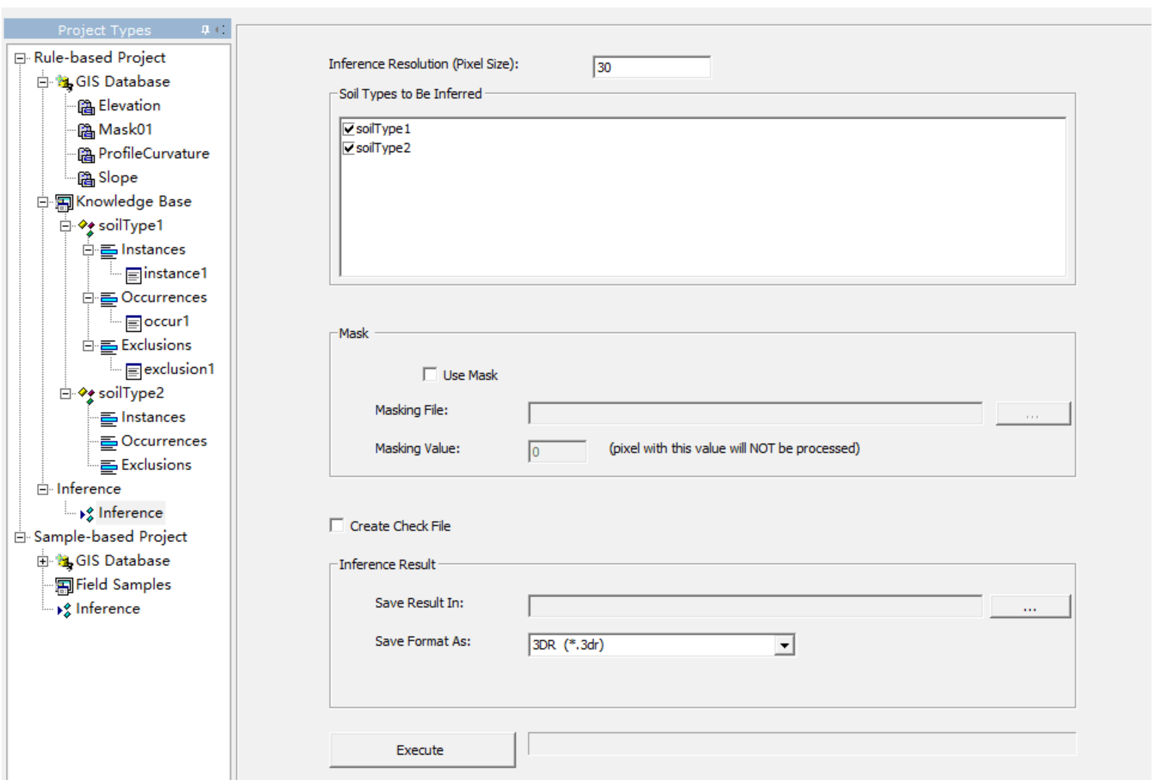
Soil Inference

The soil maps created by SoLIM Solutions are fuzzy membership maps indicating how similar the soils at different locations are to the most typical cases of the specified soil types. The resulting maps use the raster data model and are in .3dr format.

This section provides the instructions to operate the soil inference in SoLIM Solutions. With the functionalities provided by the program, you can choose soil types to be inferred, set mask file, create check files and run inference.

In the control panel, click on "Inference" node to unfold it. Under that node, click on "Inference", the view will shift to Inference View.

Soil inference is done through the following interface:



Inference Method

Inference is done pixel by pixel. For every pixel its membership (similarity) to soil type k is computed as the following:

Inference of Instances

SoLIM uses the MIN operation in fuzzy logic to integrate the optimality values derived from individual rules for a given instance of soil type *k*. In other words, it chooses the minimum value among those values as the overall output value for the instance. The theoretical basis for this option is the limiting-factor principle in ecology.

If there are more than one instances, the maximum of the values from instances will be chosen to be the final optimality value related to the global knowledge for soil type *k*.

Inference of occurrences

An occurrence has two parts: the rule part and the spatial setting. If the pixel falls into the influenced area of an occurrence, for the rule part SoLIM Solutions uses the MIN operation in fuzzy logic to integrate the optimality values derived from individual rules. This process is the same as the inference of a single instance stated in the "Inference of Instance" section above.

If distance similarity for this occurrence is disabled, then the minimum optimality value will be the overall output value for this occurrence. If the distance similarity is enabled, SoLIM Solutions uses the spatial setting to adjust the optimality value calculated using the rule part. The basic idea is that the geographically closer a location to an occurrence's central location, the more impact the occurrence will exert on the location, and the further a location to the occurrence's central location, the less impact the occurrence will exert on the location. The function of this adjustment is shown a below:

$$S_{ij,o}^{k\prime} = \frac{\frac{1}{(d_{ij,b})'} * S_{b,G}^{k} + \frac{1}{(d_{ij,o})'} * S_{ij,o}^{k}}{\frac{1}{(d_{ij,b})'} + \frac{1}{(d_{ij,o})'}}$$

The meanings of the symbols in the equation are described as follows:

$S_{ij,o}^{k\prime}$: the adjusted optimality value at location (*i, j*) calculated based on occurrence *o*.

$S_{ij,o}^{k}$: the optimality value calculated using the rule part of occurrence *o*.

b: when we draw a straight line that passes the central point and the pixel to be inferred, it will have two intersections with the boundary of the influenced area of occurrence o. Of the two points, b is the one that is on the same side with the pixel to be inferred.

$d_{ij,o}$: the distance (surface distance or horizontal distance) between the location (*i, j*) and occurrence o.

$d_{ij,b}$: the distance (surface distance or horizontal distance) between the location (*i, j*) and boundary point b.

$S_{b,G}^{k}$: the optimality value at boundary point *b* calculated based on the global knowledge (instances).

r: distance decay factor.

Inference of exclusions

Similar to an occurrence, an exclusion also has two parts: the rule part and the spatial setting. If the pixel falls into the influenced area of an exclusion, for the rule part, SoLIM Solutions uses the MIN operation in fuzzy logic to integrate the optimality values derived from individual rules. This process is the same as the inference of a single instance stated in the "Inference of Instance" section above.

If distance similarity for this exclusion is disabled, then the minimum optimality value will be chosen, and the overall output value for this exclusion is one minus the minimum optimality. If the distance similarity is enabled, SoLIM Solutions uses the spatial setting to adjust the optimality value calculated using the rule part. The basic idea is that the geographically closer a location to an exclusion's central location, the more impact the exclusion will exert on the location, and the further a location to the exclusion's central location, the less impact the exclusion will exert on the location. The function of this adjustment is shown as below:

$$S_{ij,u}^{k\prime} = \frac{\frac{1}{(d_{ij,b})^{r}} * S_{b,G}^{k} + \left(1 - S_{ij,u}^{k}\right) * \frac{1}{(d_{ij,u})^{r}}}{\frac{1}{(d_{ij,b})^{r}} + \frac{1}{(d_{ij,u})^{r}}}$$

The meanings of the symbols in the equation are described as follows:

$S_{ij,u}^{k\prime}$: the adjusted optimality value at location (*i, j*) calculated based on exclusion u.

$S_{ij,u}^{k}$: the optimality value calculated using the rule part of exclusion u.

b: when we draw a straight line that passes the central point and the pixel to be inferred, it will have two intersections with the boundary of the influenced area of exclusion u. Of the two points, b is the one that is on the same side with the pixel to be inferred.

$d_{ij,u}$: the distance (surface distance or horizontal distance) between the location (*i, j*) and exclusion u.

$d_{ij,b}$: the distance (surface distance or horizontal distance) between the location (*i, j*) and boundary point *b*.

$S_{b,G}^{k}$: the optimality value at boundary point *b* calculated based on the global knowledge (instances).

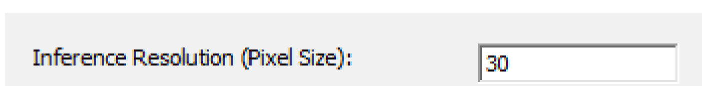
r: distance decay factor.

Derivation of Final Similarity Value (membership integration):

1. If the pixel does not fall into the influenced areas of any occurrences or exclusions, the maximum of the optimality values among all instances will be chosen to be the final similarity value for soil type *k* at the pixel location.
2. If the pixel falls into the influenced areas of one or more occurrences, but it does not fall into the influenced area of any exclusions, the maximum optimality value among all the occurrences will be the final similarity value for soil type *k* at the pixel location.
3. If the pixel does not fall into the influenced areas of any occurrences, but it falls into the influenced areas of one or more exclusions, the minimum optimality value from exclusions will be the final membership value for soil type *k* at the pixel location.
4. If the pixel falls into the influenced areas of one or more occurrences and at the same time it also falls into the influenced area of one or more exclusions, the program will firstly get the maximum optimality value from all occurrences and then choose the minimum optimality value from all exclusions. Between the two values, the smaller one will be the final membership value for soil type k at the pixel location.

Inference Resolution

You can specify the resolution (pixel size) of the inference result (fuzzy membership maps). The default resolution is the same as the resolution of the first layer in GIS Database.



Specify Soil Type to be Inferred

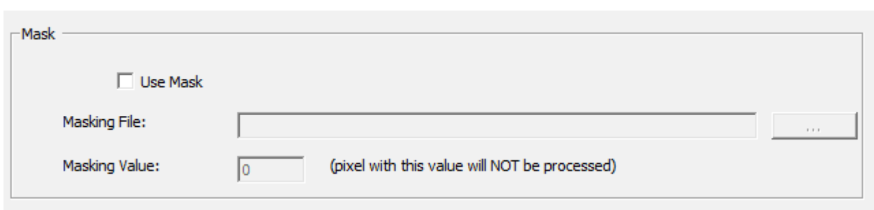
By default, SoLIM Solutions will create a set of fuzzy membership maps, one for each soil type in the current knowledge base.

In the inference view, the list on the top displays all soil types in the current knowledge base.

If you want the soil types to be inferred, check the boxes next to them. Otherwise, uncheck the boxes.

Mask Setting

This masking function is an "in or out" function. It simply classifies all the locations covered by the environmental database into two classes: locations that should be included in the inference and locations that should be excluded from the inference. The masking function is mainly for defining the mapping area when you want to set a special boundary for the output soil map, e.g., when you only want to map the soils within a watershed or specific areas. If the value of a pixel in the mask file equals to the specified Masking Value, the corresponding pixel in the output soil map will be assigned ‘Nodata’. Otherwise pixels will receive inference results.



Enable Mask:

In the Inference View, Check the box next to "Mask". This will enable the masking function and also activates the other options in the "Mask" area. Leaving the Mask box unchecked disables the masking function.

Set Masking File

Click on ... in the "Mask" area to open a masking file. A Masking file should be .3dr file of integer type and its extent should be the same with the first data layer in the current GIS database.

Set Masking Value

Input an integer number as masking value.

Check File

A check file is a map showing two pieces of information:

- the specific instance/occurrence /exclusion of the soil type that determines the final fuzzy membership at a location.
- the rule in the determining instance/occurrence/exclusion that determines the final fuzzy membership at a location.

A check file is a .3dr file and is named by adding "chk" at the end of the output file name for the soil type.

Each value in a check map has two parts: the integer part and the decimal part. The integer part is the ID of instance/occurrence/ exclusion, which indicates the determining instance/occurrence/exclusion at that location. The ID corresponds to the order of the instance/occurrence /exclusion in the soil type (1 for instance, 2 for occurrence, 3 for exclusion). If none of the existing instances, occurrences and exclusions has effect at a location, the integer part of the check file value at that location will be -1.

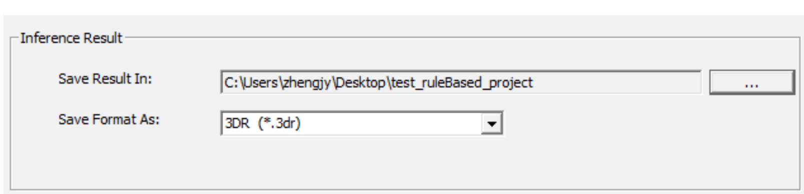
The decimal part is the determining rule's ID, which indicates the determining environmental variable at that location.

For example, suppose that a pixel in the check file has an output of 103.04. This value can be interpreted as follow: The integer part of the check file in this example is 103. "1" indicates that it is the instance that determines the final fuzzy membership. "03" indicates that for this location the third instance in the soil type is the determining instance. The decimal part of the check file in this example is "04", which means the fourth environmental feature is the determining environmental feature at this location.

Similarly, "205.03" means the 3rd rule of the 5 the occurrence is the determining rule. "301.05" means the 5th rule of the 1st exclusion is the determining rule.

Run Inference

In inference view, click on within "Inference Result"



A dialog will appear for you to create or choose output directory. Choose or create a new directory to host the output membership maps, click on "Ok", the selected directory will appear on the "Save Result " text box. Click on "Execute" to start the inference process.

If inference is completed without error, in the result directory you can see the fuzzy membership map and check file (if you choose to create check file) for every soil type that was specified in "Specify Soil Types to be Inferred".

3.2.2 Sample-based Project

Project Management Overview

A sample-based project is the unit on which SoLIM Solutions performs database management, field sample importing and soil inference. A complete definition for a project consists of three components:

1. **Project file(.sip):** This file contains project configurations and the link to the sample file used in a project.
2. **Environmental data layers (GIS database) containing raster data layers needed for inference:** The files must be in 3dMapper raster format (.3dr) in a subdirectory named "GISData". This subdirectory must be in the same directory as the project file.
3. **Field sample file (.csv):** This file records the locations (X, Y coordinates) of field samples and the soil properties and/or soil types at sample location.

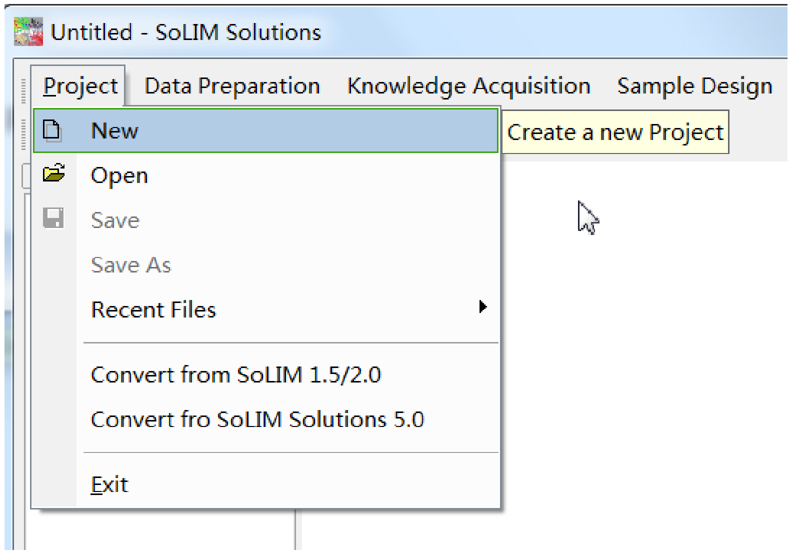
A project is stored in a project directory. Soil inference requires a GIS database and a field sample file.

In this section, you will find information about how to

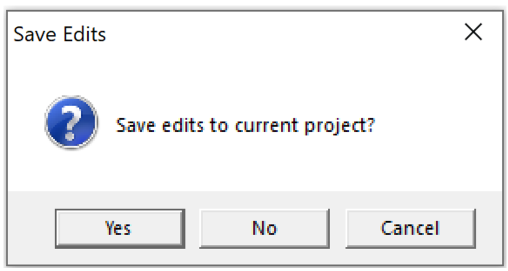
- Create a New Project
- Open an Existing Project
- Edit a Project
- Save a Project

Create a New Project

Select "Project->New" to create a new project.

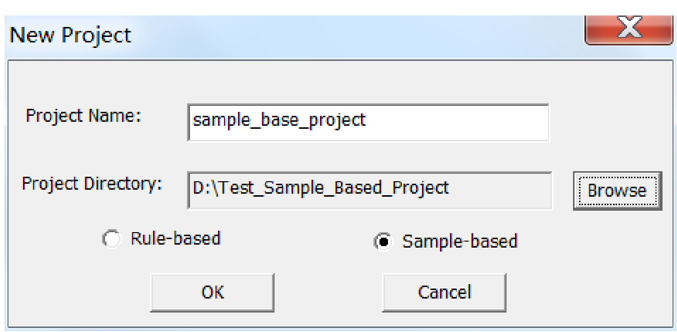


If you are editing another project, the program will remind you to save the current project.



Click on "Yes" to save it, "No" to skip or "Cancel" to return.

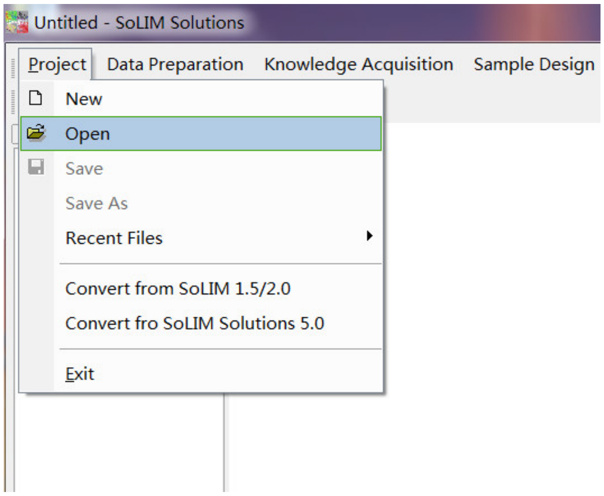
If you choose "Yes" or "No", the dialog below will appear. Specify project name and project directory. And choose "Sample-based".



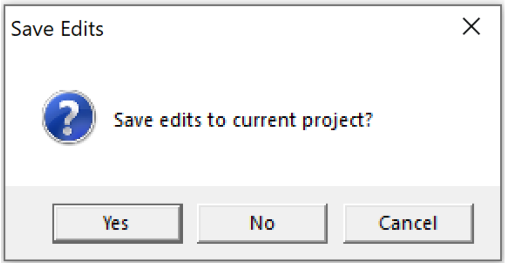
Open an Existing Project

If a SoLIM Solutions project has already been saved, you can open it and edit.

Select "Project->Open".



If you are editing another project, the program will remind you to save the current project.



Click on "Yes" to save it, "No" to skip or "Cancel" to return.

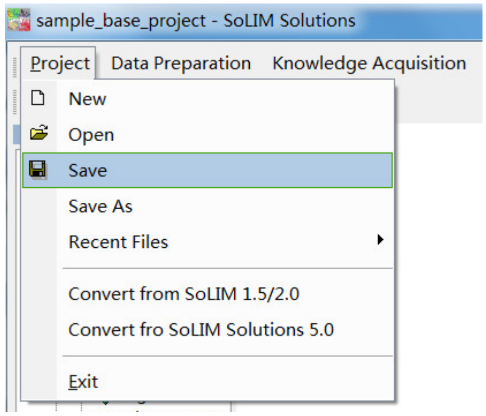
If you choose "Yes" or "No", the original project is closed. A dialog below will appear to let the user specify the project file (.sip) to be opened. After selecting a file, click on "Open". If the specified project file is valid, it will be opened. Otherwise, the program will pop up an error dialog.

Edit a Project

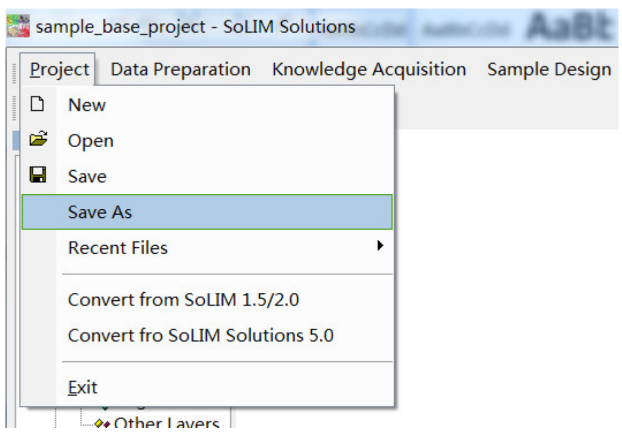
Please refer to GIS Database Management and Field Sample File Management.

Save a Project

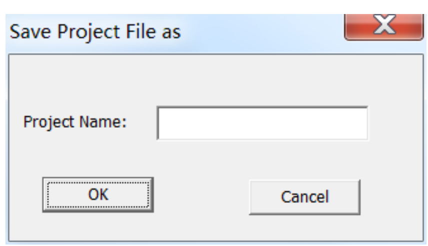
You can save a project at any time.



You can save a project to a file with another name by clicking "Project->Save As". "Save As" is useful when you need to save different versions of a project with different modification. In fact, they (these versions) can be treated as different projects with the GIS data layers in the same directory (GISData). This can save disk space if the majority of the GIS data layers are common to multiple projects.



In the dialog below, specify the name you want to use.



Click on "OK" and the program will save the project file (.sip) with the name you specified in the project directory.

GIS Database Management

Before any inference can be performed, the needed environmental data layers must be ready and properly referenced in the project file. The collection of these references to environmental data layers constitutes the GIS *database*. The GIS database:

- provides data on the *environmental variables* used in the soil inference; and
- defines the spatial extent of the resulting soil maps.

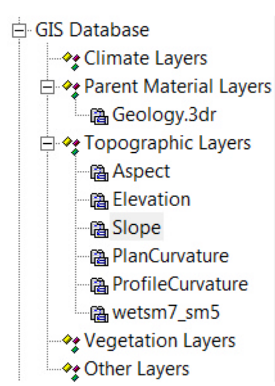
If the data layer for an environmental variable is not referenced or does not exist, the variable will not be used in the inference.

Environmental data layer must be in the **same coordinate system**.

The native file format for SoLIM Solutions is 3dr format. However, SoLIM Solutions can convert other commonly used raster file format (e.g. TIFF, ERDAS IMAGE) to .3dr format on the fly. SoLIM Solutions can also rasterize commonly used vector file format (e.g. Shapefile) into 3dr format. See the File Converters in the Utilities menu for how.

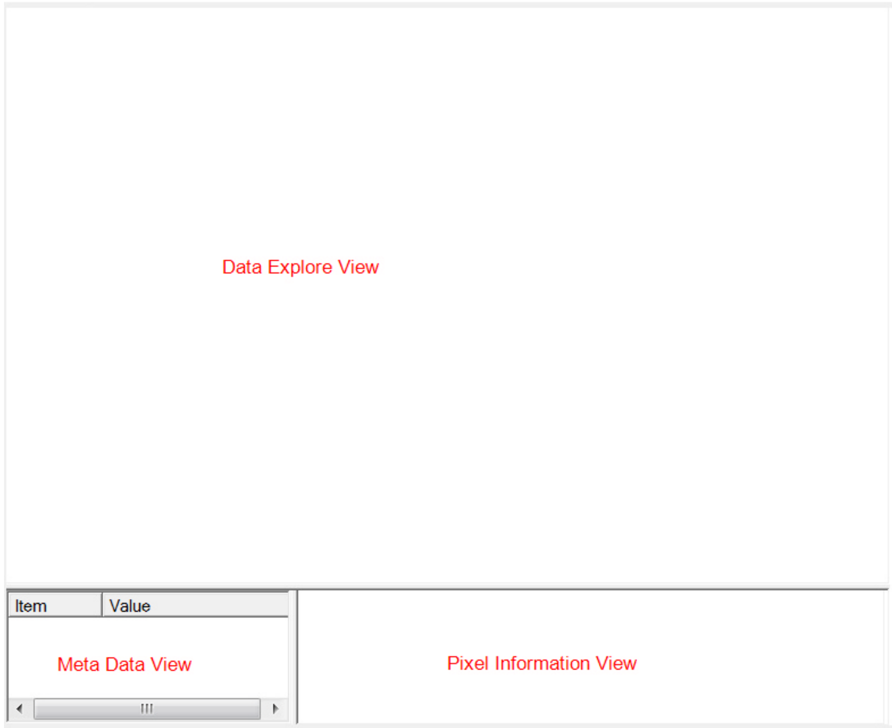
SoLIM Solutions automatically checks the spatial extents and resolutions of the data layers when they are loaded. If the spatial extents of the data layers do not match, SoLIM Solutions only perform soil inference over the common area covered by all data layers. Now data layers with different resolutions can be mixed and used in SoLIM Solutions.

The "GIS Database" Node and its Sub-nodes in the Project Panel



By manipulating this node and its sub-nodes, you can add and delete environmental data layers in the GIS database.

The GIS Database Views:



The GIS database view contains three sub-views:

- *The Data Explore View*: provides a graphical view of the data layer currently selected.
- *The Meta Data View*: displays the header information (meta data) of the data layer currently selected.
- *The Pixel Information View*: displays the value of pixels at or near the mouse pointer.

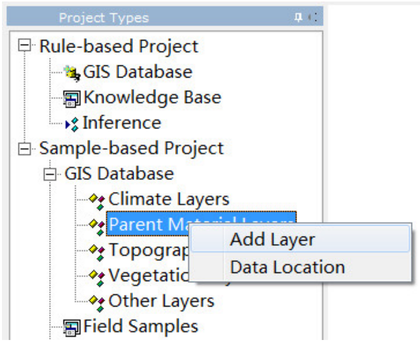
The View Control Toolbars

The toolbar provides view control functionalities, such as zooming and panning for you to explore the displayed data layer.

Add and Delete Data Layers

Add Data Layer(s)

1. Right click on any sub-node (e.g. "Topographic Layers") of "GIS Database" node. In the pop-up menu, select "Add Layer".



2. In the add layer dialog, choose one or more data files. The recommended file format is 3dr format. However, SoLIM Solutions can convert other commonly used raster file format (e.g. TIFF, ERDAS IMAGE) to .3dr format on the fly. SoLIM Solutions can also rasterize commonly used vector file format (e.g Shapefile) into 3dr format.

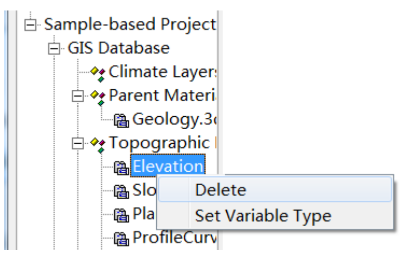
3. Click on "OK", the data layer is added to the current project and copied into the *GISData* directory of the current project. If the layer is not valid, SoLIM will pop up a error dialog and the layer will not be added.

Tips:

1. The default name of a data layer is the "data name" stored in the header of the .3dr file.
2. The name of a newly added data layer MUST BE DIFFERENT from the names of ALL OF THE EXISTING data layers in the project.

Delete a Data Layer

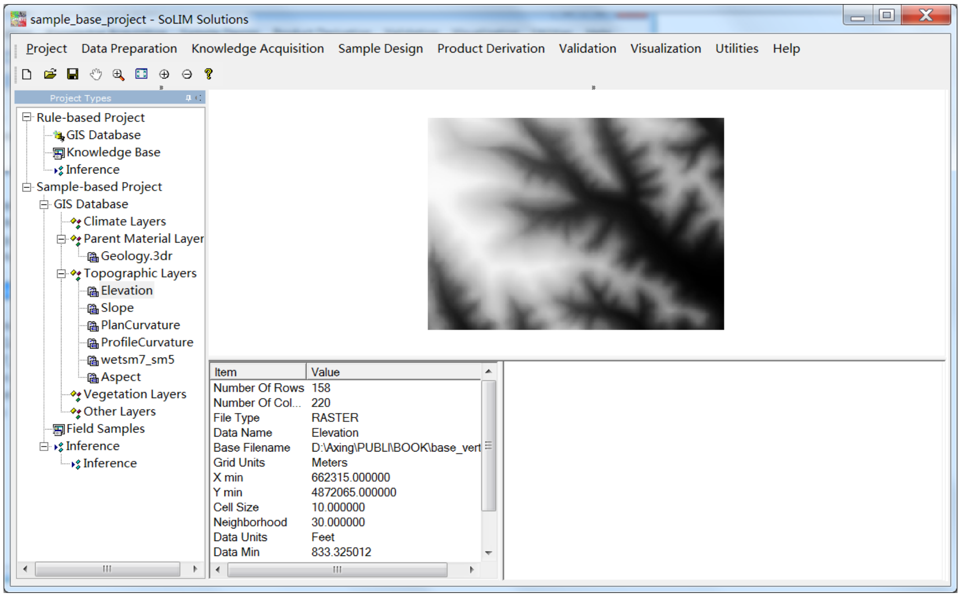
In the control panel, right click on the name of the data layer to be deleted. Click on "Delete".



Tips: The data layer is only REMOVED from the list. The data layer file still in the *GISData* directory.

Exploring Your Data

Click on a data layer in the left project panel, the layer will be displayed in the data explore view. Header information will be displayed in the meta data view.



View the Value of Pixels

In the data explore view, move the mouse to point at the pixel, the value of the pixel will be displayed in red color at the center of the pixel information view. The values of the pixels near the mouse pointer will also be displayed in black color. If you want to view more neighboring pixels, drag the edge of pixel information view to enlarge the view display area.

	134	135	136	137	138
59	910.61004	911.63501	912.04504	913.07000	914.50500
60	904.66503	905.28002	905.69000	906.91998	908.15002
61	899.33496	899.53997	899.74499	900.56500	901.58996
62	896.87500	896.46496	896.26001	896.26001	896.26001
63	897.69494	896.66998	896.05499	895.23498	894.20996
64	899.94995	898.92498	898.10498	896.66998	895.44000

Zoom In

Click on ⊕ in the toolbar to zoom in the display.

Zoom Out

Click on 🔍 in the toolbar to zoom out the display.

Pan

Click on in the toolbar, hold the mouse button down while dragging in the direction you want to move the area, then release the button.

Show All

Click on in the toolbar to set the display to the original state.

Field Sample File Management

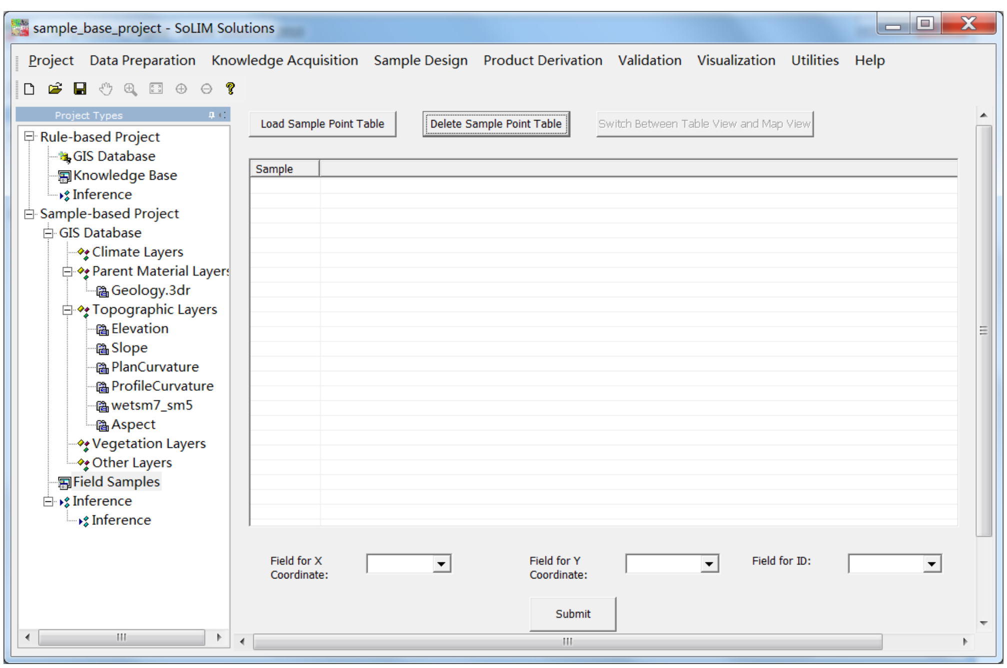
Before any inference can be performed, field sample points must be obtained and organized into a .csv file. The field sample file

- must contains columns for x and y coordinates.
- must have at least one column for soil property or soil type.

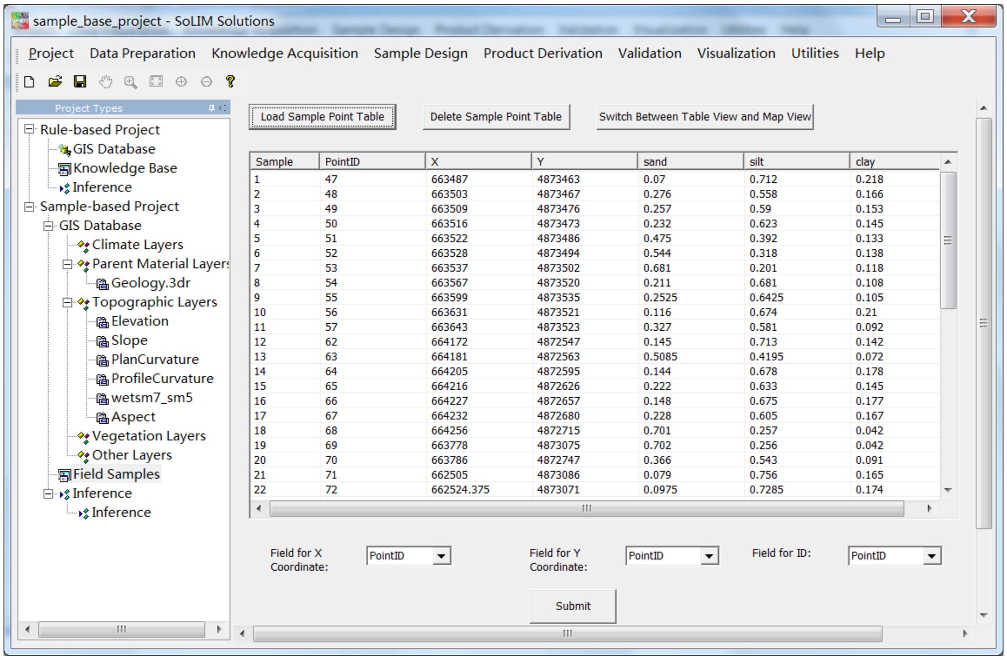
This section provides an instruction to the field sample file management in SoLIM Solutions. With the functionalities provided by the program, you can import and view sample points.

Load Field Sample File

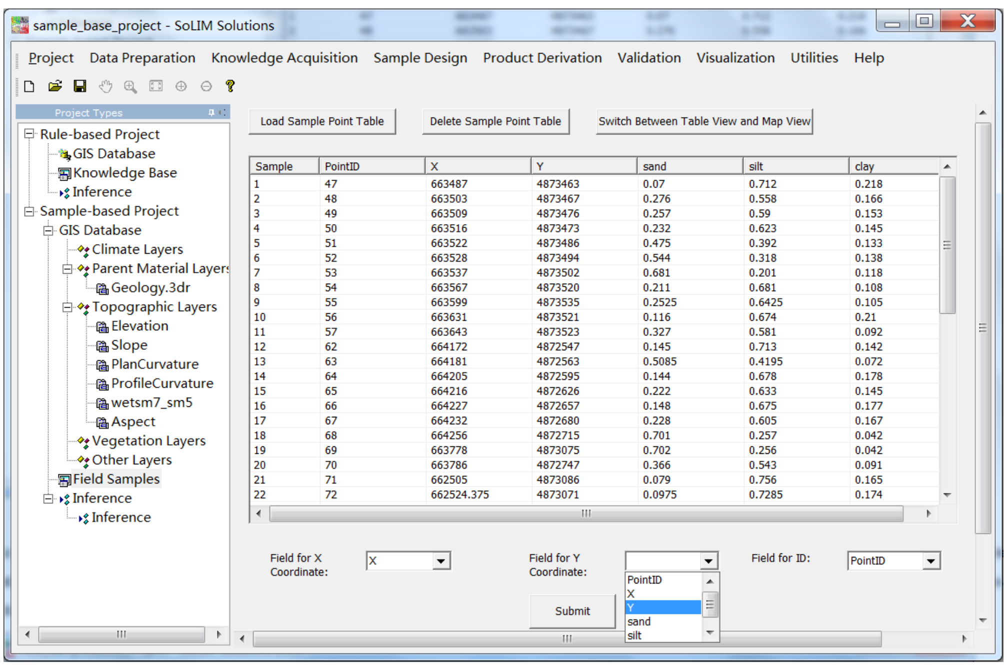
Left click "Field Samples" node in the left project panel. The interface below will show up for you to import field points.



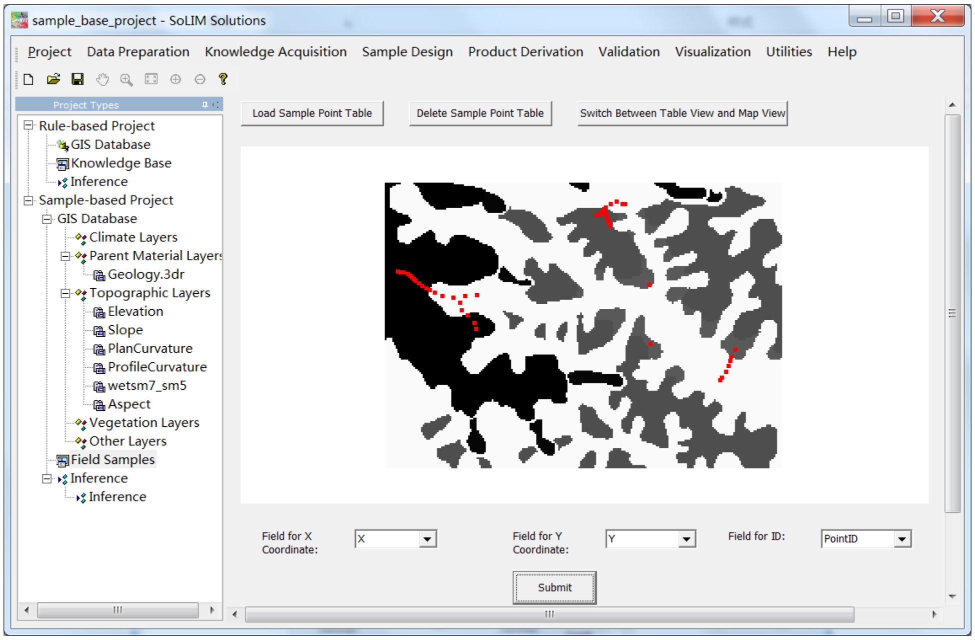
Click "Load Sample Point Table" to add the .csv file to our project.



Specify the columns that record X coordinate and Y coordinate and ID by selecting column names from drop-down list.



Click "Submit", you will see the distribution of the sample points in map view. The red points represent the field sample point in the study area. It should be noted that **at least one layer** needs to be in the GIS Database. Otherwise, you cannot view the distribution in map view.



View Sample Points

You can always switch between table view and map view by clicking "Switch between table view and map view".

Map View:

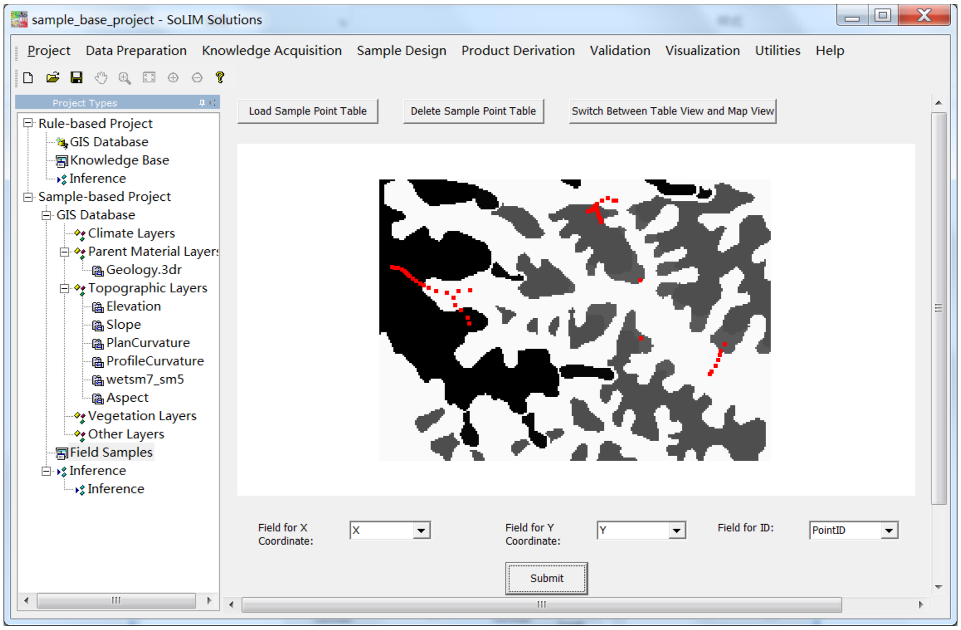
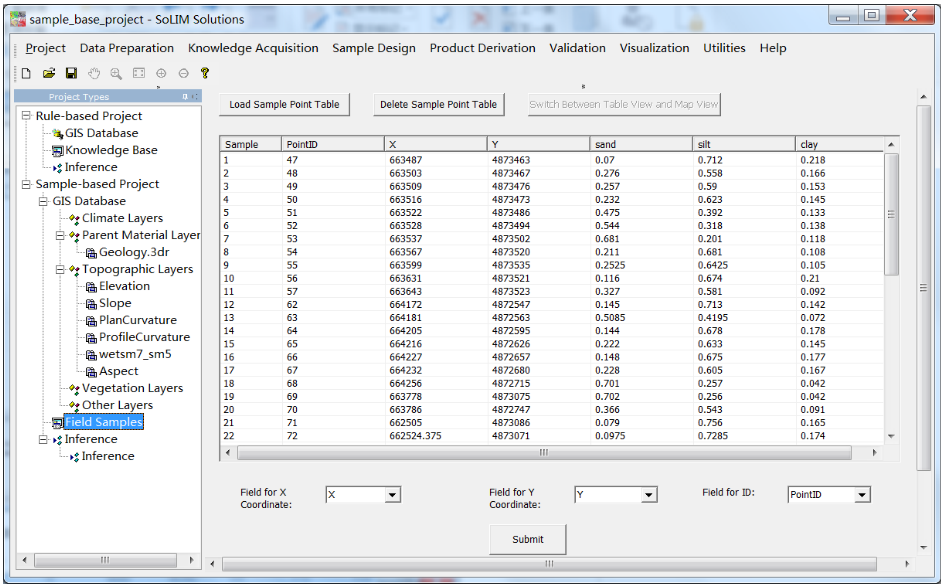


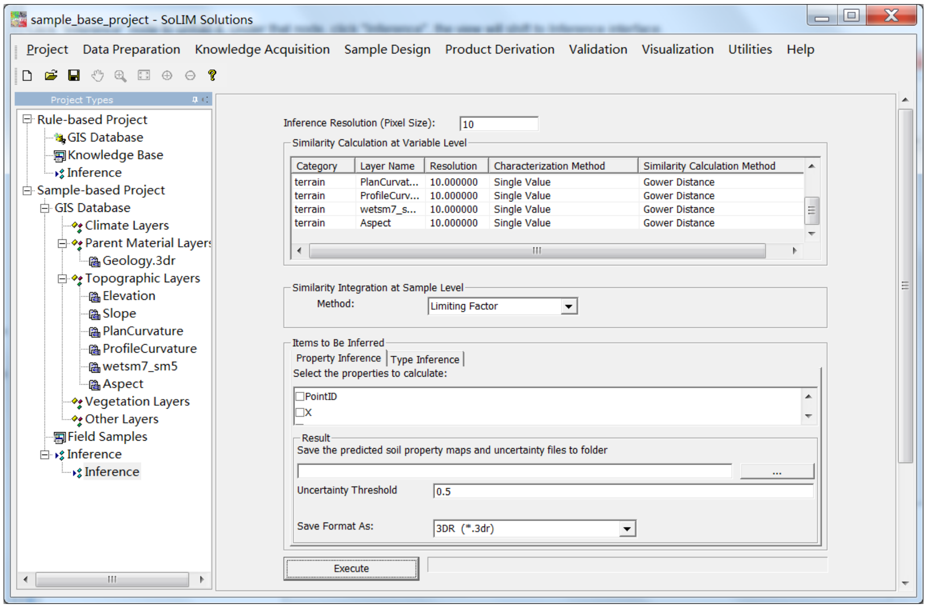
Table View:



Soil Inference in Sample-Based Project

SoLIM Solutions can be used to infer **soil property** and **soil type**. This section provides the instructions to operate the soil inference in sample-based project. With the functionalities provided by the program, you can set similarity calculation methods, choose soil properties or types to be inferred and run inference. The results use the raster data model and are in *.3dr* format.

In the project panel, click on "Inference" node to unfold it. Under that node, click on "Inference", the view will shift to Inference View. Soil inference is done through the following interface:



Inference Method

Each field sample in the field sample file reflects an underlying relation between soil and its relative environmental conditions, and this relation would recur over the space. It is assumed that locations with similar environmental conditions will have similar soil property/type. Therefore, each sample can be considered representative over locations with similar environmental conditions. That is, each sample has **individual representativeness**. Moreover, the representativeness level of an individual sample on an unsampled location can be approximated by the environmental similarity between them. Based on this concept, the soil property/type on unsampled locations can be predicted by referring to environmentally similar samples.

Before any inference can be done, similarity between each pixel and each field sample should firstly be estimated. In SoLIM Solutions, similarity calculations are conducted at two levels: variable level and sample level.

Similarity Calculation at Variable Level:

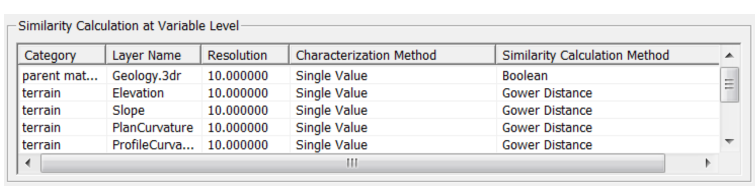
The type of methods for calculating similarity on variable level depends on the type of the environmental variable.

Single Value

If the variable is nominal or order, Boolean function should be used. For example, Boolean function is suitable for determining parent material similarity. That is, if the parent material of the pixel to be inferred and the field sample location are the same, the similarity value is 1, otherwise, the similarity value is 0.

If the variable is interval or ratio, Gower Distance can be used.

Both Boolean function and Gower Distance are **single value** similarity calculation methods.


Similarity Calculation at Variable Level

Category	Layer Name	Resolution	Characterization Method	Similarity Calculation Method
parent mat...	Geology.3dr	10.000000	Single Value	Boolean
terrain	Elevation	10.000000	Single Value	Gower Distance
terrain	Slope	10.000000	Single Value	Gower Distance
terrain	PlanCurvature	10.000000	Single Value	Gower Distance
terrain	ProfileCurva...	10.000000	Single Value	Gower Distance

Similarity Calculation at Sample Level:

Similarity calculation at sample level, in fact, is integrating similarities at variable level. SoLIM Solutions provides three integration methods:

Limiting Factor

It assumes that the least favorite environmental condition determines soil property on a given site, so the minimum similarity among the five environmental categories is used as the final similarity. This is the default method.

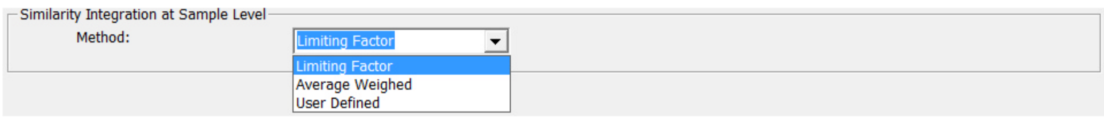
Weighted-Average

If each environment category has same significance, this integration method can be used.

User Defined

If this method id selected, users can customize weights of different environmental variables.

Users can select different method from the drop-down list.



Once similarity calculation methods are determined, inference can be conducted. Inference methods for soil property and soil type are different.

Soil Property Inference and Uncertainty Quantification:

Then soil property of each pixel is predicted by integrating the soil property values at sample points using the calculated similarities as weights.

Besides soil property prediction, the predicting uncertainty can also be quantified. Predicting uncertainty on an unsampled location would be high if there are no samples environmentally similar with this location. That is, if similarities with all samples are all very low, the uncertainty of the soil property at certain pixel will be high.

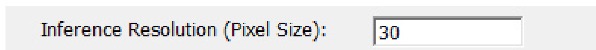
It should be noted that for one pixel, if the uncertainty is higher than the uncertainty threshold, the soil property will be set to *NoData*.

Soil Type Inference:

For certain soil type, all the field samples which have that soil type are selected and the environmental similarity between a pixel and selected sampled location are measured. Then the maximum similarity is assigned to the pixel for the soil type. Therefore, the final results are a set of similarity files, one for each soil type.

Inference Resolution

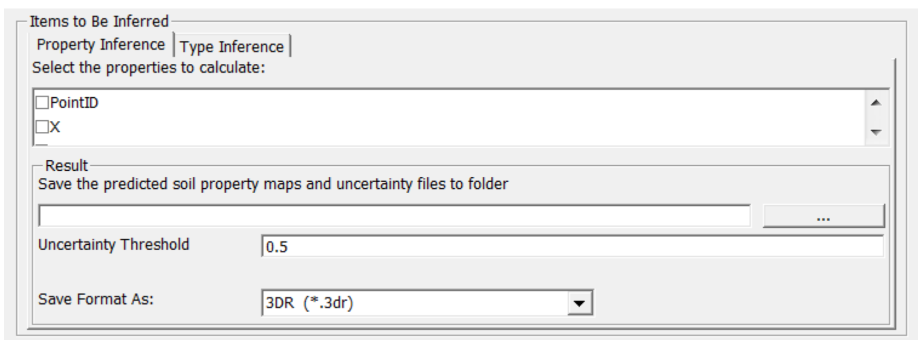
You can change the resolution (pixel size) of the inference result. The default resolution is the same as the largest pixel size of all the GIS data layers.



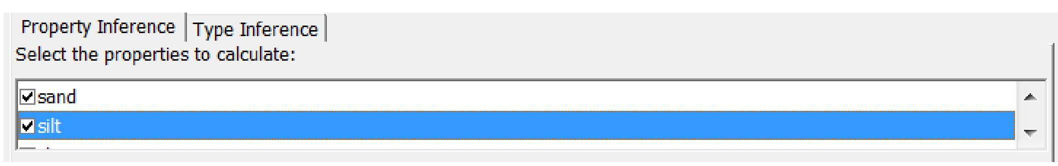
Soil Property Inference

Specify Soil Properties to be Inferred

If you want to infer soil property, make sure you select "Property Inference" tab.



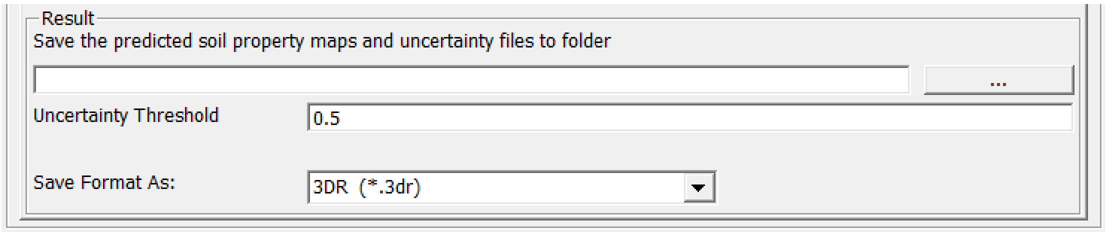
The list displays all the columns in the current field sample file.



Check the boxes next to the soil property you want to predict. Otherwise, uncheck the boxes. Note that you can check multiple soil properties.

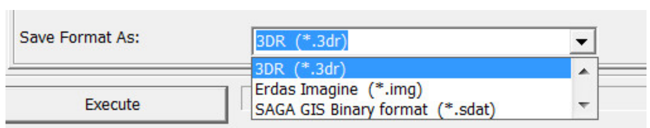
Run Inference

The results can be set in the "Result" box.



Choose the folder to save the predicted soil property map(s) and uncertainty file by clicking on .... A dialog will appear for you to create or choose output directory. Choose or create a new directory to host the output soil property map(s) and uncertainty file, click on "Ok", the selected directory will appear on the "Save the predicted soil property maps and uncertainty file to fold" text box. Specify uncertainty threshold in the corresponding text box. The default threshold is 0.5. Note the value should be between 0 and 1.

We prepare multiple choices for the format of the output data, such as 3DR default format (.3dr), Erdas Imagine format (.img), SAGA format (.sdat), TIFF/BigTIFF/GeoTIFF format (.tif). You can select one of formats as the output data format.



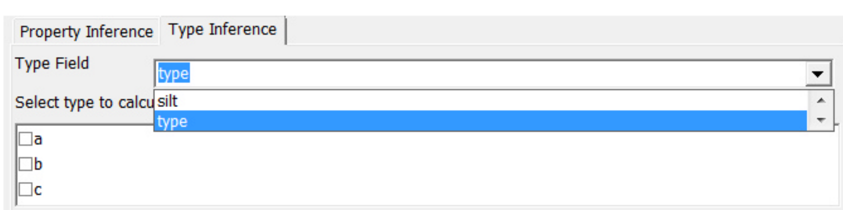
Click on "Execute" to start the inference process. If inference is completed without error, in the folder you just specified you can see the soil property map(s) and the uncertainty file.

Soil Type Inference

Specify Soil Types to be Inferred

If you want to infer soil property, make sure you select "Type Inference" tab.

You firstly need to specify which column (Type Field) stores soil type names in the field sample file.



Then the list will display all the different soil types in the column you just specified.

Check the boxes next to the soil type for which you want to generate similarity file. Otherwise, uncheck the boxes. Note that you can check multiple soil types.

Run Inference

In inference view, click on ... within "Result". A dialog will appear for you to create or choose output directory. Choose or create a new directory to host the output membership maps, click on "Ok", the selected directory will appear on the "Save membership files to folder" text box.

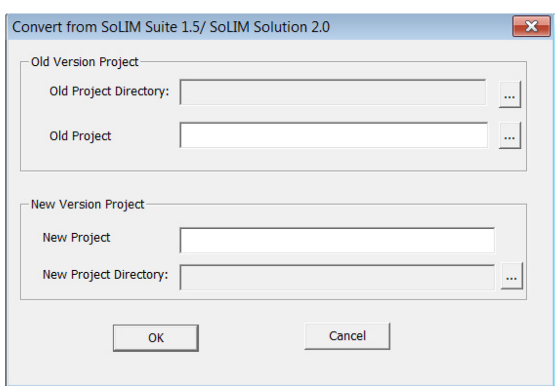
Click on "Execute" to start the inference process. If inference is completed without error, in the result folder you can see the similarity file for each soil type you specified.

3.2.3 Import Project

If you have SoLIM projects from SoLIM Suite 1.5/SoLIM Solutions 2.0 or SoLIM Solutions 5.0 you can import these projects into SoLIM Solutions. There is no need to re-create existing SoLIM projects.

Import from SoLIM Suite 1.5/SoLIM Solutions 2.0

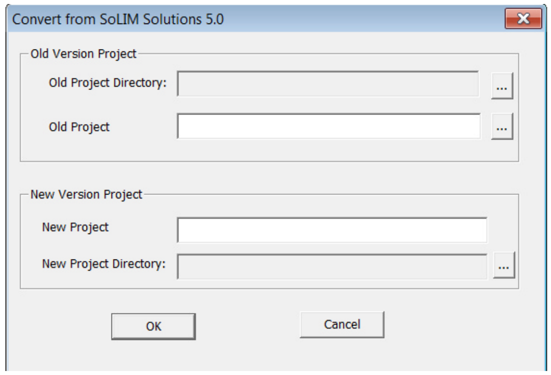
Select "Project->Convert from SoLIM 1.5/2.0". This will bring up the dialog below.



Fill in the required fields and click on "OK".

Import from SoLIM Solutions 5.0

Select "Project->Convert from SoLIM 5.0". This will bring up the dialog below.



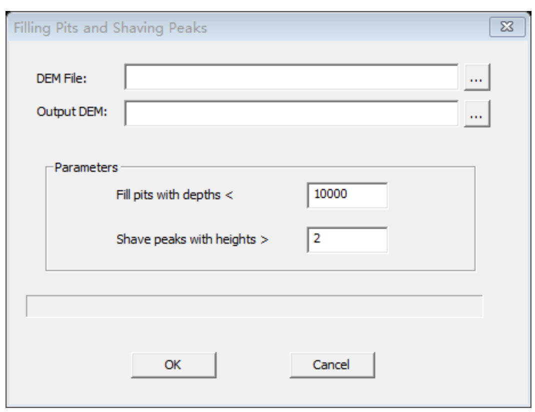
Fill in the required fields and click on "OK".

3.3 Data Preparation

3.3.1 Terrain Analysis

Fill and Shave DEM

Select "Data Preparation->Fill and Shave DEM". Specify the parameters, the algorithms and the output file.



Output: ***the DEM being filled and shaved.***

DEM File

The DEM data of the area.

Fill pits with depth <

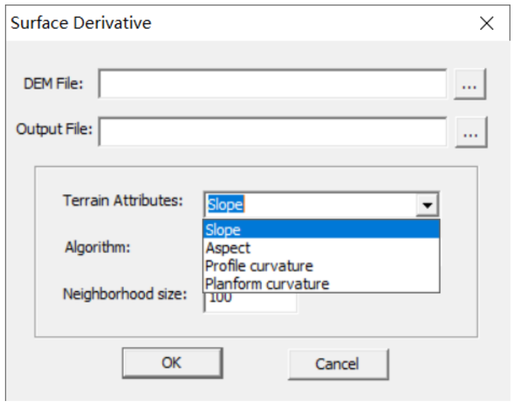
The pits with its depth less than the setting value will be filled.

Shave peeks with height >

The peeks with its height larger than the setting value will be shaved.

Surface Derivatives

Select "Data Preparation->Surface Derivative". Specify the parameters, the algorithms, and the output file.



Output: *Slope*

Gradient (result is in percent (tan(degree)))

Output: *Aspect*

The orientation of the line of steepest descent (Result is in degree)

Output: *Profile curvature*

Slope profile curvature

Output: *Planform curvature*

Contour curvature

DEM file

The DEM data of the area.

Neighborhood size

The size of the neighborhood.

You can select different algorithms to do this. Three algorithms are available: *Evans-Young, Horn, Zevenhergen-Thorn.*

Algorithm: ***Evans-Young's (Evans, 1972; Young, 1978; Evans, 1979)***

The essence of the Evans–Young method consists of the following parts: The non-smooth land surface in the vicinity of a given point is replaced by a "sheet" of a smooth surface (the second-order polynomial) that is fitted to the real surface by a least squares criterion. After that, this smooth surface ("sheet") is used for calculations of MVs that are expressed by first and second partial derivatives.

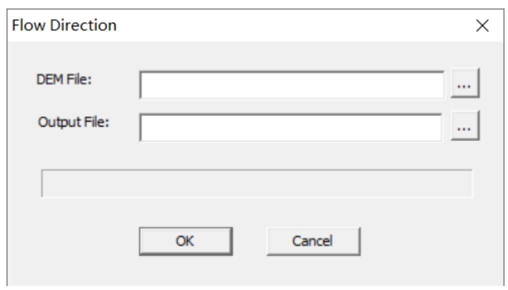
Algorithm: ***Horn's***

Algorithm used by ArcInfo. It uses square-shape neighborhood. When using circular neighbor, this method becomes equivalent to Evans' (1972).

Algorithm: ***Zevenbergen-Thorne's (Zevenbergen and Thorne, 1987)***

Flow Direction

Select "Data Preparation->Flow Direction". Specify the parameters and the output file.



Output: ***Flow Direction***

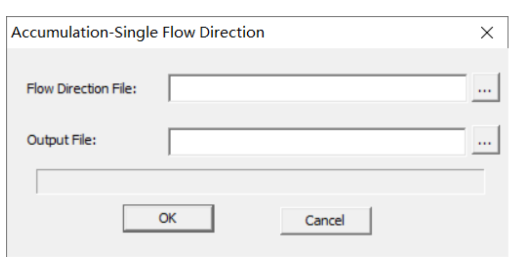
Creates a raster of flow direction from each cell to its steepest downslope neighbor.

DEM file

The filled DEM data of the area.

Flow Accumulation-Single Flow Direction

Select "Data Preparation->Flow Accumulation-Single Flow Direction". Specify the parameters and the output file.



Output: ***Flow Accumulation***

Flow accumulation is an indirect way of measuring the drainage areas (in units of grid cells).

Method: ***Singe Flow Direction***

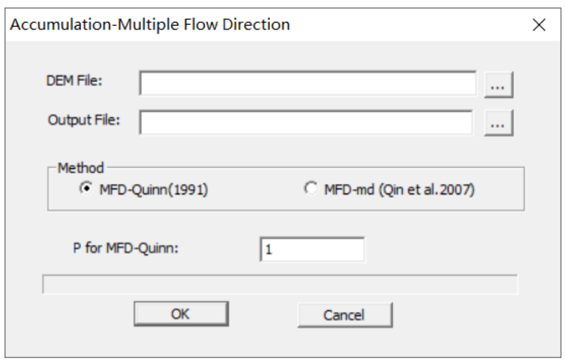
A single flow direction algorithm assumes that subsurface flow occurs only in the steepest downslope direction from any given point.

Flow Direction File

A raster of single-flow direction file of the area.

Flow Accumulation-Multiple Flow Direction

Select "Data Preparation->Flow Accumulation-Multiple Flow Direction". Specify the parameters, the algorithms, and the output file.



Output: ***Flow Accumulation***

Flow accumulation is an indirect way of measuring the drainage areas (in units of grid cells).

Method: ***Multiple Flow Direction***

A multiple flow direction algorithm assumes that subsurface flow occurs in all downslope directions from any given point.

DEM file

The filled DEM data of the area.

Algorithm: ***MFD-Quinn (also known as FD8, Quinn et al., 1991)***

MFD with a fixed flow partitioning exponent.

P for MFD-Quinn

The flow partitioning exponent.

Algorithm: ***MFD-md (Qin et al., 2007)***

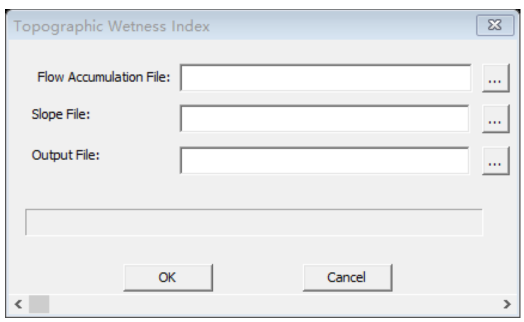
The maximum downslope gradient was chosen as the local topographic attribute for the flow-partition scheme. Use a linear function of the maximum downslope gradient as f(e).

$$f(e) = \begin{cases} p_l \, ; (e \leq e_{min}) \\ \dfrac{e - e_{min}}{e_{max} - e_{min}} \times (p_u - p_l) + p_l \, ; (e_{min} < e < e_{max}) \\ p_u \, ; (e \geq e_{max}) \end{cases}$$

where *e* is tangent *b* with *b* being the maximum downslope gradient; $f(e)$ is the flow partition function; p_u and p_l are the upper and lower bounds of $f(e)$ and are used as the *p* values representing completely divergent and convergent flows, respectively; and e_{min} and e_{max} are the *e* values which are associated with p_l and p_u, respectively. The typical values for the bounds (i.e. p_l and p_u) are 1.1 and 10.

Topographic Wetness Index

Select "Data Preparation-> Topographic Wetness Index". Specify the parameters and the output file.



Output: ***Topographic Wetness Index (Quinn et al., 1995)***

The topographic wetness index (TWI, ln(a/tanβ)), which combines local upslope contributing area and slope, is commonly used to quantify topographic control on hydrological processes.

Flow Accumulation File

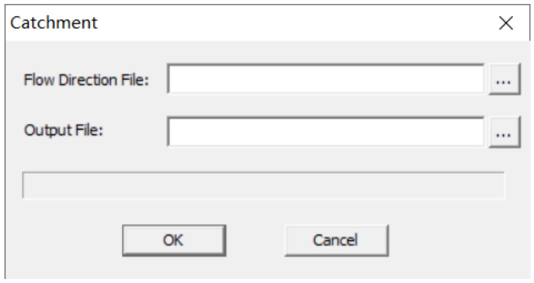
The flow accumulation (area unit) of the area.

Slope File

The slope in tangent (TAN) of the area.

Catchment

Select "Data Preparation->Catchment". Specify the parameters and the output file.



Output: ***Catchment Areas***

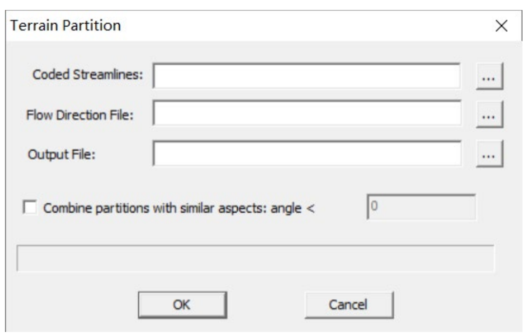
Delineate catchment areas

Flow Direction File

The flow direction data generated by "Flow Direction" modular in "Data Preparation".

Terrain Partition

Select "Data Preparation->Terrain Partition". Specify the parameters and the output file.



Output: ***Terrain Partition***

Divide the terrain into slope partitions.

Coded Streamlines

The streamlines data of the area, which can be generated by "Streamline" modular in "Data Preparation".

Flow Direction File

The flow direction data generated by "Flow Direction" modular in "Data Preparation".

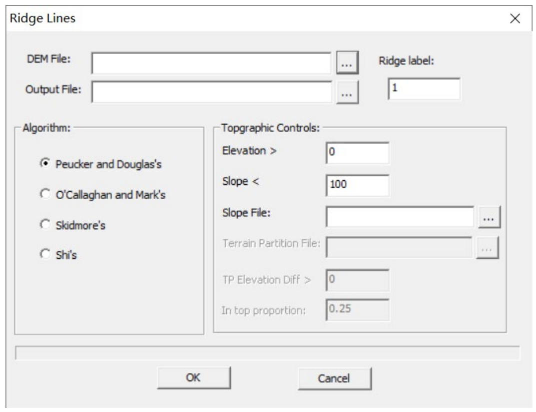
Combine partitions with similar aspect: angle <

If the difference between the two partitions is less than the value, combine them.

Terrain Positions

Ridgeline

Select "Data Preparation->Terrain Positions->Ridgelines". Specify the parameters, the algorithms and the output.



Output: *Ridge Lines*

Ridge label

The label you want to assign to the ridge lines.

DEM file

The DEM data of the area.

Four algorithms are available: Peucker and Douglas's; O'Callaghan and Mark's; Skidmore's and Shi's. You can define the steepness of the ridge line. To filter out noise in the flat drainage way, you need to set a threshold on Elevation.

Algorithm: *Peucker and Douglas's (Peucker and Douglas, 1975)*

The methods all use local analysis of the data; they are “parallel” in the sense that the results at each point do not depend on the results already obtained at other points.

> *Elevation >*
>
> If the elevation at a point is lower than the threshold, skip the point.
>
> *Slope Gradient >*
>
> If the slope gradient at a point is smaller than the threshold, skip the point.

Slope file

The slope gradient data of the area, which can be generated by "Surface Derivatives" modular in "Data Preparation".

Algorithm: *O'Callaghan and Mark's (O'Callaghan and Mark's, 1984)*

The method handles artificial pits introduced by data collection systems and extracts only the major drainage paths. Its performance appears to be consistent with the visual interpretation of drainage patterns from elevation contours.

Elevation >

If the elevation at a point is lower than the threshold, skip the point.

Slope Gradient >

If the slope gradient at a point is smaller than the threshold, skip the point.

Slope file

The slope gradient data of the area, which can be generated by "Surface Derivatives" modular in "Data Preparation".

Algorithm: *Skidmore's (Skidmore, 1991)*

Elevation >

If the elevation at a point is lower than the threshold, skip the point.

Slope Gradient >

If the slope gradient at a point is smaller than the threshold, skip the point.

Slope file

The slope gradient data of the area, which can be generated by "Surface Derivatives" modular in "Data Preparation".

Algorithm: *Shi's*

Elevation >

If the elevation at a point is lower than the threshold, skip the point.

Slope Gradient >

If the slope gradient at a point is smaller than the threshold, skip the point.

Slope file

The slope gradient data of the area, which can be generated by "Surface Derivatives" modular in "Data Preparation".

Terrain partition file

The terrain partition data of the area, which can be generated by the "Terrain Partition" modular in "Data Preparation".

TP Elevation Diff >

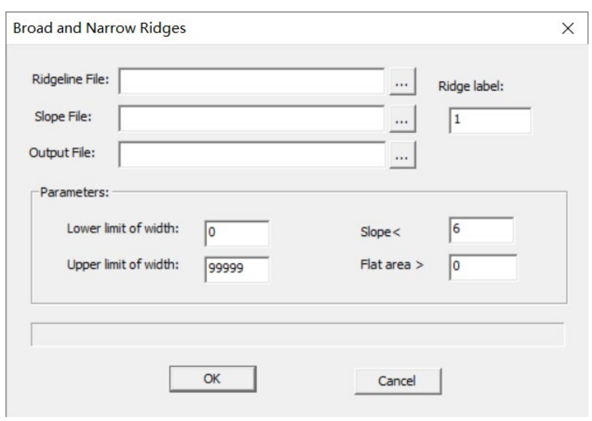
The minimum difference of elevation between the terrain partitions.

In top proportion

The proportion of the pixels in use in a terrain partition.

Broad/Narrow Ridgelines

Select "Data Preparation->Terrain Positions->Broad/Narrow Ridgelines". Specify the parameters and the output file.



Output: ***to classify the ridgelines based on the user-defined parameters with fuzzy logic***

Ridge label

The code of the ridge line.

Ridgeline file

The ridge line data of the area, which can be generated by the "Ridge Line" modular in "Data Preparation".

Slope File

The slope gradient data generated by "Surface Derivatives" modular in "Data Preparation".

Lower limit of width

The width of the ridgeline will not be less than the value.

Slope <

If the slope gradient of a point is smaller than the threshold, skip the point.

Upper limit of width

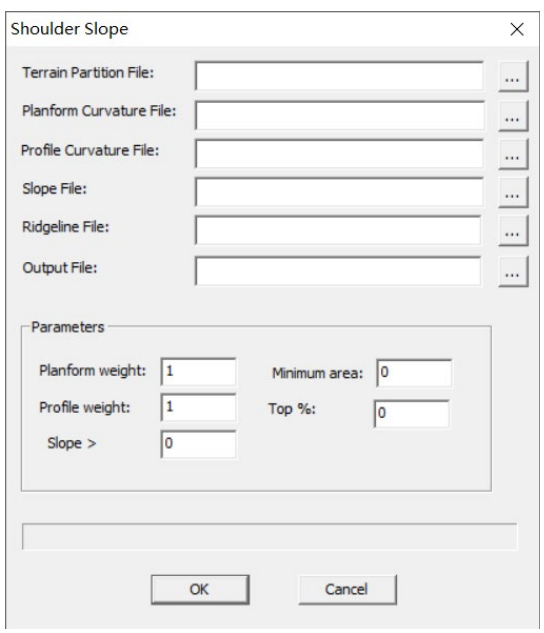
The width of the ridgeline will not be larger than this value.

Flat Area >

If the flat area of a ridgeline is smaller than this threshold, skip the ridgeline.

Shoulder

Select "Data Preparation->Terrain Positions->Shoulder". Specify the input file and the output file.



Output: *Shoulder Slope*

Terrain Partition File

The terrain partition data generated by "Terrain Partition" modular in "Data Preparation".

Platform Curvature File

The terrain partition data generated by "Surface Derivatives" modular in "Data Preparation".

Profile Curvature File

The profile curvature data generated by "Surface Derivatives" modular in "Data Preparation".

Slope Gradient File

The slope gradient data generated by "Surface Derivatives" modular in "Data Preparation".

Ridgeline File

The ridgeline data generated by "ridgeline" modular in "Data Preparation".

Platform weight

The weight of planform curvature in an integrated curvature index.

Minimum area

If the area of a partition is smaller than this threshold, skip the partition.

Profile weight

The weight of profile curvature in an integrated curvature index.

Top%

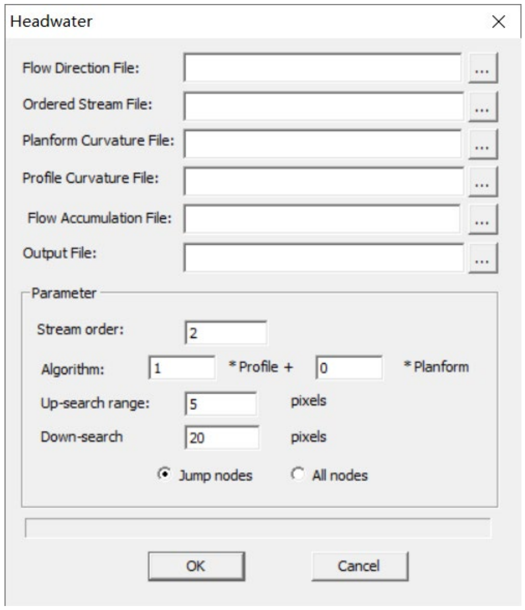
The percentage of the pixels which will be used in a terrain partition.

Slope >

If the slope of a point is smaller than this threshold, skip the point.

Headwater

Select "Data Preparation->Terrain Positions->Headwater". Specify the input file and the output file.



Output: ***Headwater***

The origin of a river.

Flow Direction File

The flow direction data generated by "Flow Direction" modular in "Data Preparation".

Ordered Stream File

The ordered stream data of the area, which can be generated by "Streamline" modular in "Data Preparation".

Platform Curvature File

The terrain partition data generated by "Surface Derivatives" modular in "Data Preparation".

Profile Curvature File

The profile curvature data generated by "Surface Derivatives" modular in "Data Preparation".

Flow Accumulation File

The upper drainage area data of the area.

Stream order

The order of the streamlines to exact the headwater.

Platform weight

The weight of planform curvature in an integrated curvature index.

Profile weight

The weight of profile curvature in an integrated curvature index.

Up-search range

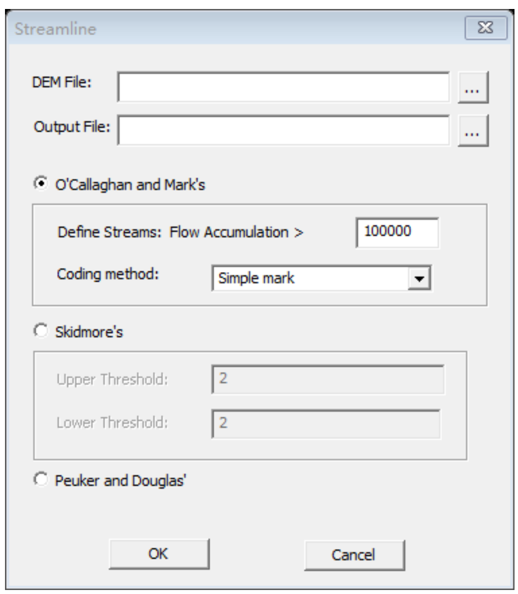
The number of iterations when searching upward.

Down-search range

The number of iterations when searching downward.

Streamline

Select "Data Preparation->Terrain Positions->Streamline". Specify the input file and the output file.



Output: *Streamline*

The streamlines in the area.

DEM file

The DEM data of the area.

Algorithm: ***O'Callaghan and Mark's (O'Callaghan and Mark's, 1984)***

Define Streams: Flow Accumulation >

When the UDA of a pixel is larger than the threshold, the pixel is part of a streamline.

Coding Method

The way to label the streamlines.

Algorithm: ***Skidmore's (Skidmore, 1991)***

Upper Threshold

The threshold to determine that a pixel is in a streamline.

Lower Threshold

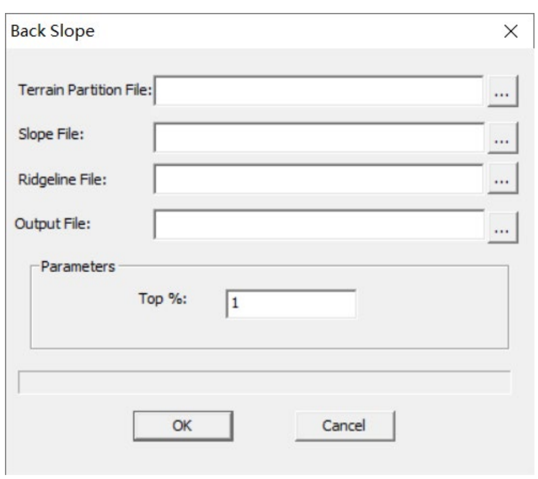
The threshold to determine that a pixel is not a streamline.

Algorithm: ***Peucker and Douglas's (Peucker and Douglas, 1975)***

The methods all use local analysis of the data; they are "parallel" in the sense that the results at each point do not depend on the results already obtained at other points.

Back Slope

Select "Data Preparation->Terrain Positions->Back Slope". Specify the input file and the output file.



Output: ***Back Slope***

Terrain Partition File

The terrain partition data generated by "Terrain Partition" modular in "Data Preparation".

Slope Gradient File

The slope gradient data generated by "Surface Derivatives" modular in "Data Preparation".

Ridgeline File

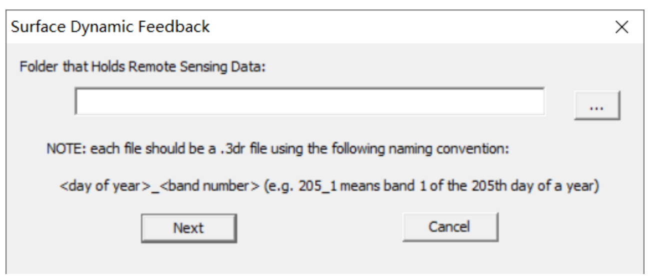
The ridgeline data generated by "ridgeline" modular in "Data Preparation".

Top%

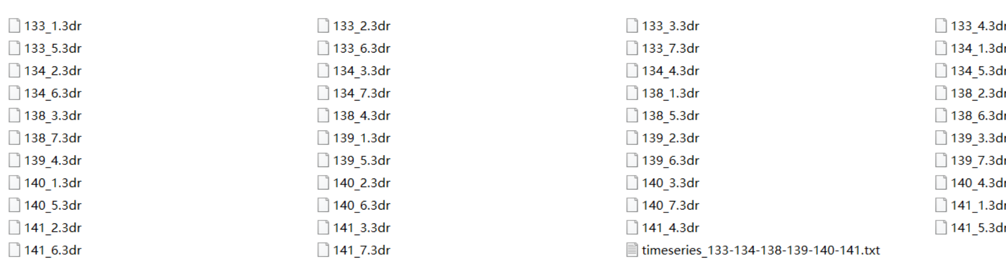
The percentage of the pixels in use in a terrain partition.

3.3.2 Remote Sensing Analysis

Select "Data Preparation-> Remote Sensing Analysis". Specify the folder that holds Remote.

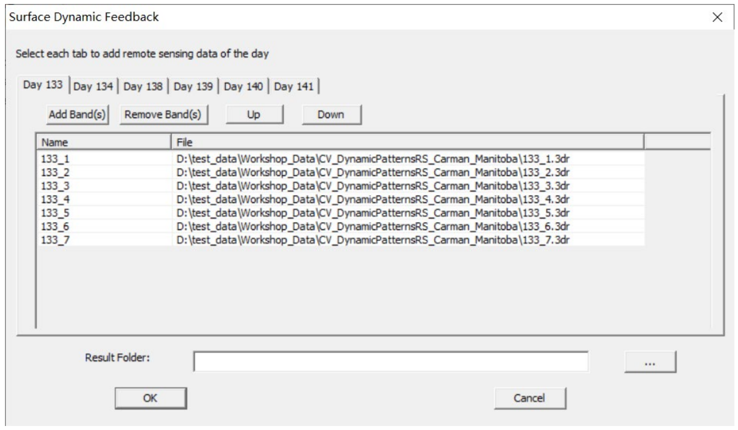


The structure of the example data is shown below:



Click Next

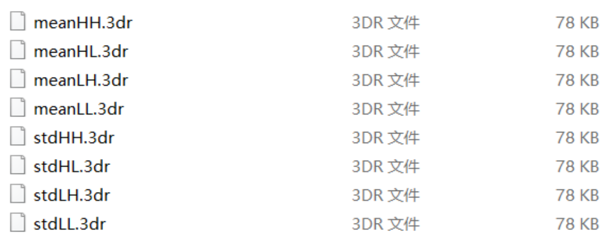
SoLIM Solutions will parse the files in the folder and organize them based on days.



Specify the result folder and click "OK"

Output: ***Soil Covariates***

The generated covariates will be put into the result folder.

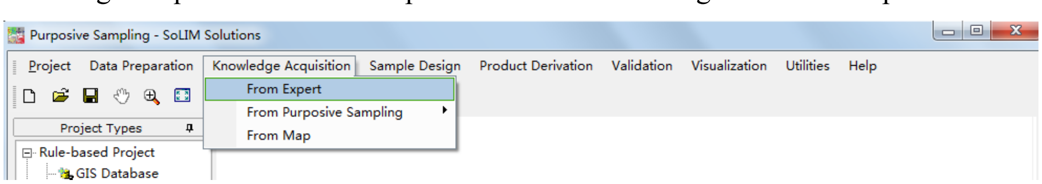


In the result folder, there will be eight new files generated from wavelet analysis. They are used to summarize the spatial-temporal response surface along different directions (e.g. horizontal direction, diagonal direction). You can visualize them in SoLIM Data Viewer.

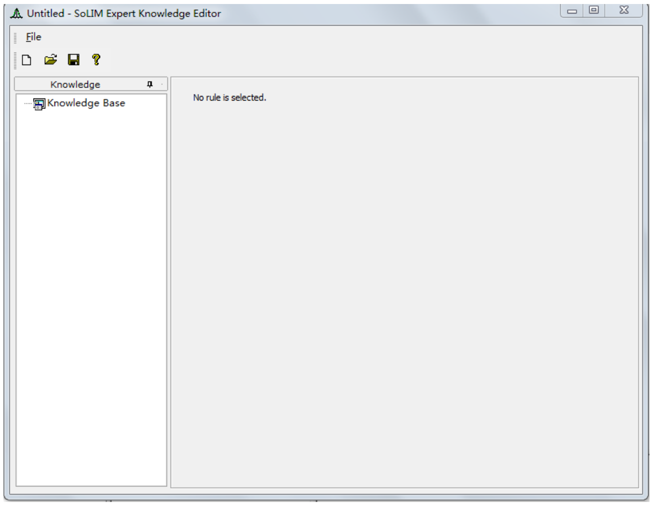
3.4 Knowledge Acquisition

3.4.1 From Expert

Select "Knowledge Acquisition-> From Expert " to extract knowledge base from experts.

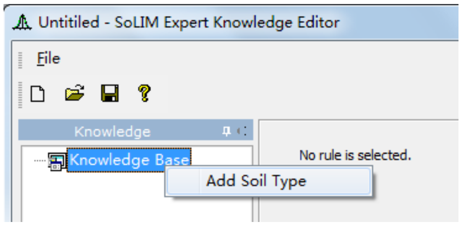


The window for expert knowledge editor will pop up.

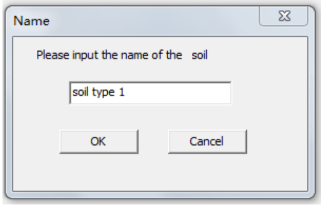


Add Soil Types:

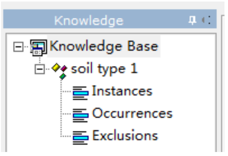
In the left project panel, right click on "Knowledge Base" node and select "Add Soil Type".



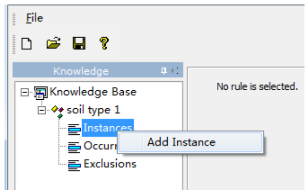
This displays a dialog box for you to specify the name of the soil type. For other soil types, use different IDs. Enter "soil type 1" and click on "OK".



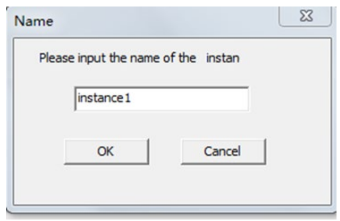
A new soil type is added into the knowledge base. Unfold the soil type node, you will see three sub-nodes are create: Instances, Occurrences, Exclusions. They are used to hold different kinds of knowledge (see Soil Mapping Using SoLIM for discussion on the different kinds of knowledge).



Based on the soil-environment knowledge for soil type 1 (**"South facing slopes with linear planform curvature, sandstone geology (indicated by the number 15 in the geology data layer) and slope gradient greater than 12%".)** there is only one environmental configuration under which soil type 1 occurs. The environmental configuration takes effect in the whole mapping area, so only one instance is needed to represent the knowledge (global knowledge) in the knowledge base. Right click on the "Instances" node under the "soil type 1" node and select "Add Instance". Right click "Instances" and select "Add Instance" and click "OK".



This will display a dialog to let you input the instance name. Enter "Instance1" and click on "OK", a new blank instance will be created.



Under the "Instances" node, you can see the newly created instance.

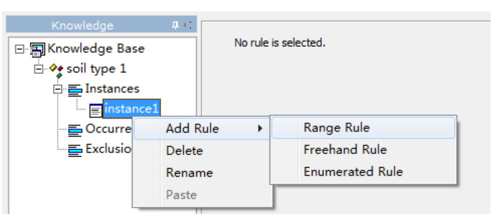
Create Rules:

Three environmental variables are used in the soil-environment knowledge for this soil type. Thus, the next task is to create rules for each environmental variable.

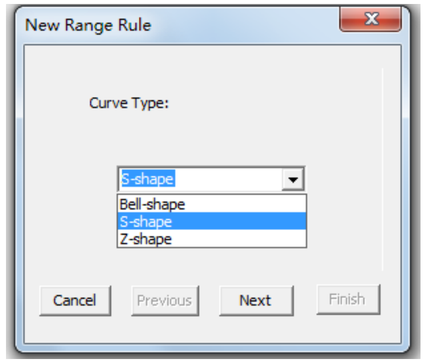
Rule for Slope:

We can use range rule to express the knowledge on slope.

Right click on the "Instance1" node. In the pop-up menu, select "Add Rule" and then select "Range Rule".



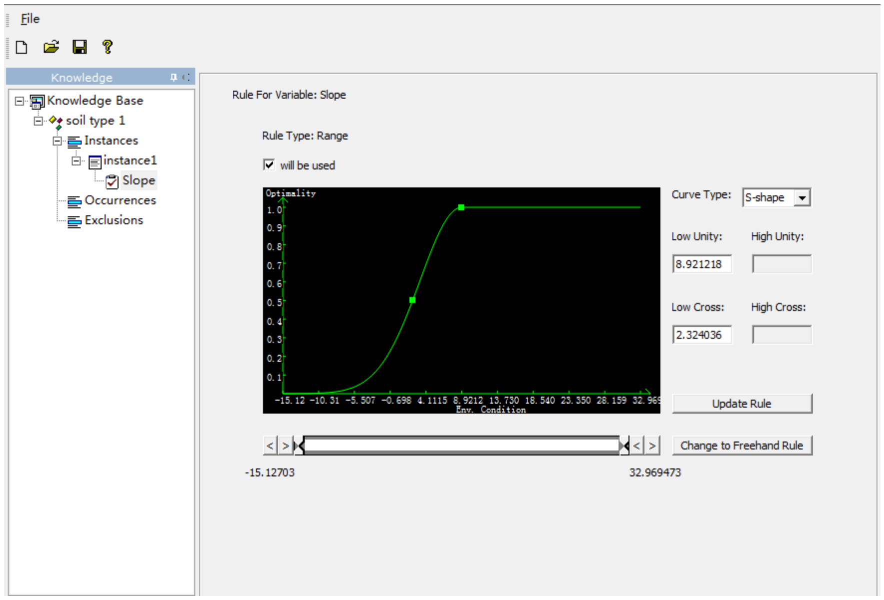
Based on the soil environmental conditions, Soil Type 1 occurs on slopes steeper than 12%. This calls for an S-shaped curve which says that with the increase of slope the membership value for being this soil type also increases. Click on the "Curve Type" drop down list and select "S-shaped", then click on "Next".



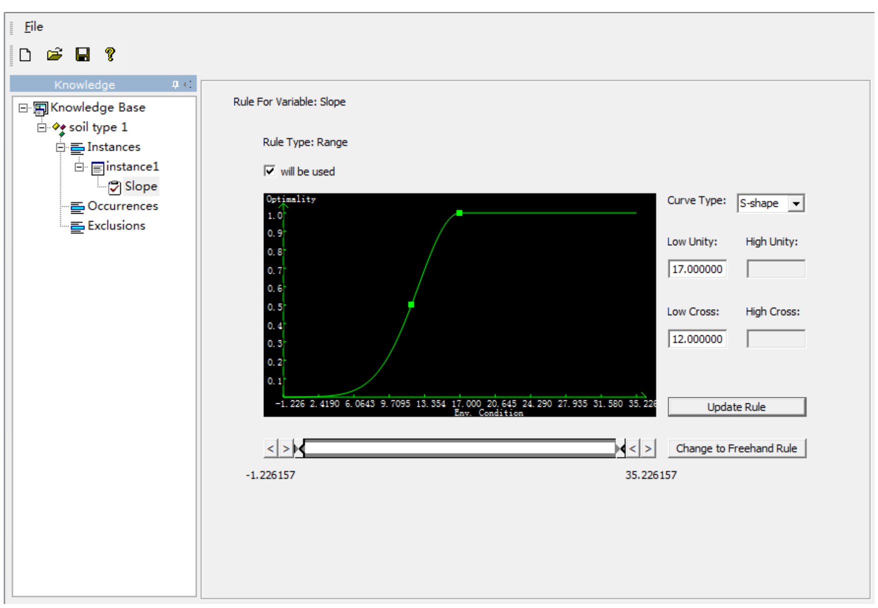
we need to determine the low unity and low cross. Low unity in this case represents the ideal slope value for the soil to be Soil Type 1. Let assume in this case low unity to be 9.0%. Low cross is set to be 3.0%. When low unity and low cross are all entered, then click on "Next".

Enter the name of the variable for the rule. Here we type in "Slope" then click on " Finish ".

Click on "Finish", a default S-shape curve is created for you. The view will switch to rule editing interface.



We can customize the newly created rule to better represent the knowledge on slope. The current rule as shown above represents that the soil occurs in areas with slope gradient greater than 2.324%. To make it represent greater than 12%, we need to determine the low unity and low cross. Low unity in this case represents the ideal slope value for the soil to be Soil Type 1. This is another piece of knowledge that soil scientists need to provide. Let us assume in this case low unity to be 17%. Low cross (the environmental condition at which the membership for this soil type is 0.5) is set to be 12%. When low unity and low cross are all entered, click on "Update Rule" to submit. This brings up the interface shown below:

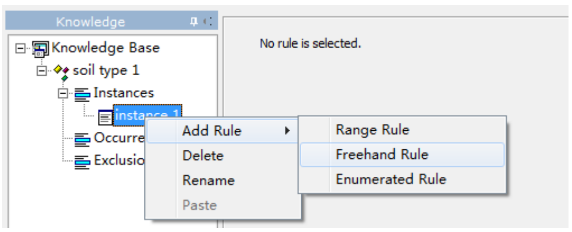


You can also click on and drag the green handles (dots) on the curve to adjust the rule. Please try to drag the handles respectively and see their individual impacts on curve and the values in text boxes. You can also click the "Apply The Rule to The Layer" button to see the effect of the rule.

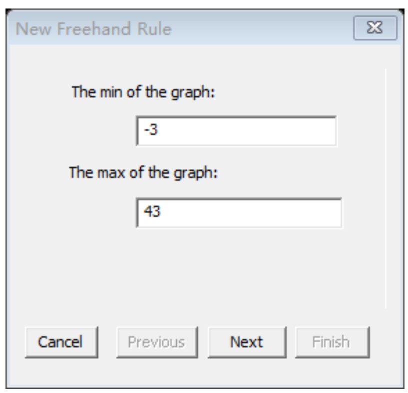
Rule for Gradient:

We can use Freehand rule to express the knowledge on gradient.

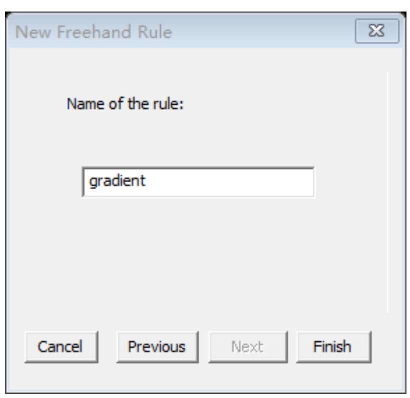
Right click on the "Instance1" node. In the pop-up menu, select "Add Rule" and then select "Freehand Rule".



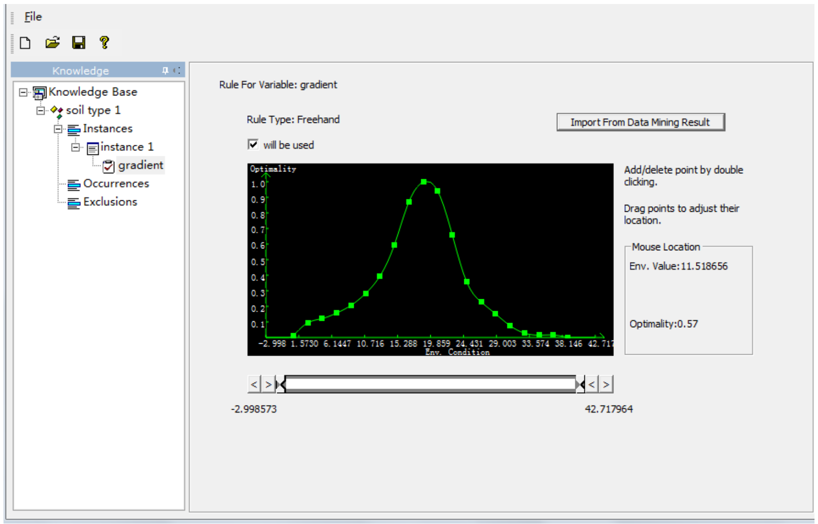
we need to determine the min and max of the graph. Let assume in this case the min of the graph to be -3.0. the max of the graph is set to be 43. When the min and max of the graph are all entered, then click on "Next".



Enter the name of the rule then click on " Finish ".

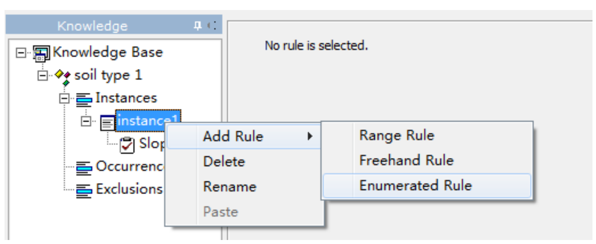


Click “Import From Data Mining Result” and specify the knowledge curve file (.txt file). The specified curve will be imported.

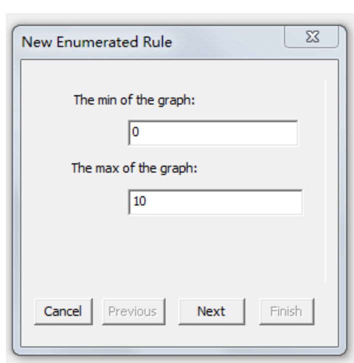


Rule for Geology:

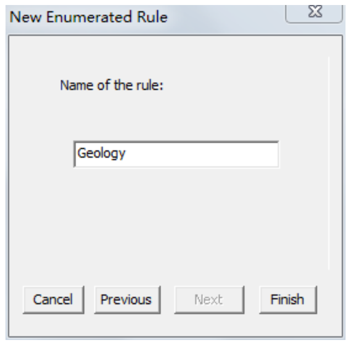
Geology is different from other environmental variables because it is a categorical variable. You should create an enumerated rule for it. The way to create enumerated rule is very similar to that of creating a range rule. Right click on the "Instance1" node. In the pop-up menu, select "Add Rule" and then select "Enumerated Rule".



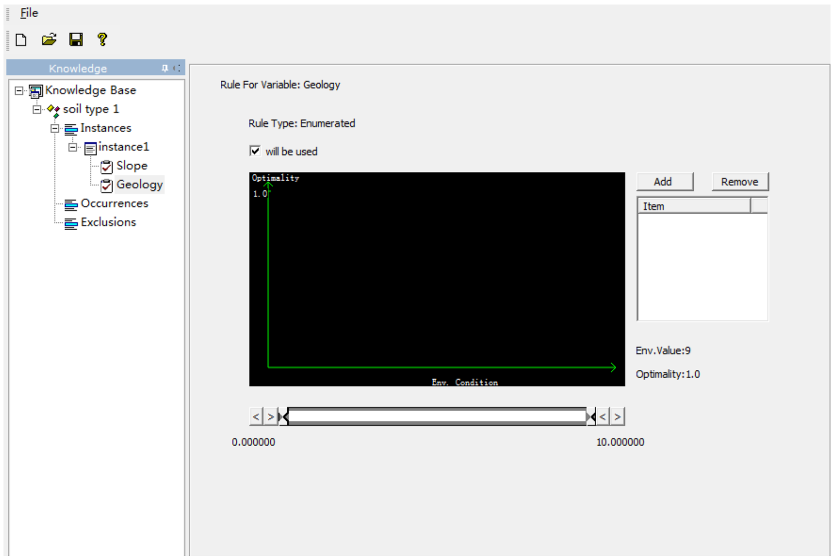
we need to determine the min and max of the graph. Let assume in this case the min of the graph to be 0. the max of the graph is set to be 10. When the min and max of the graph are all entered, then click on "Next".



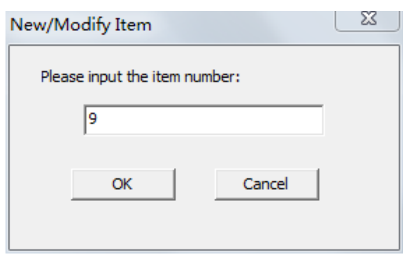
Enter the name of the rule then click on " Finish ".



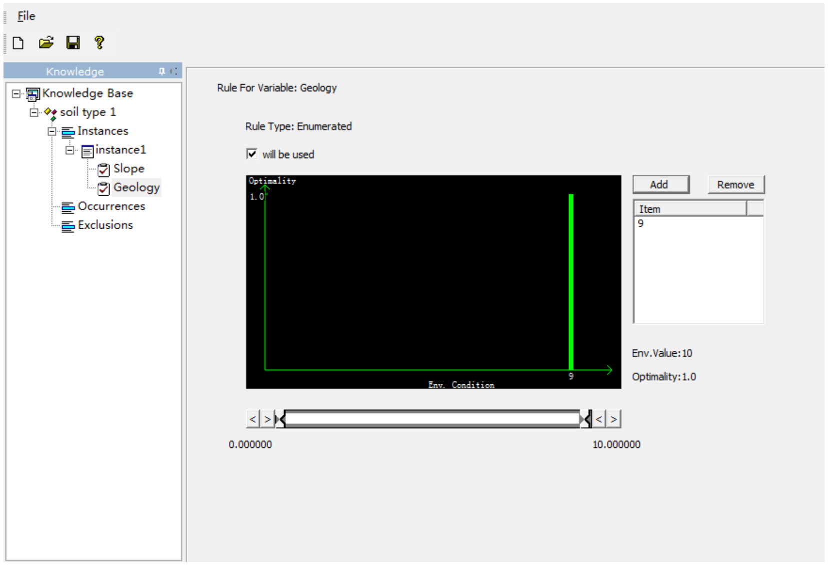
A blank enumerated rule will be created for you. The view will switch to rule editing interface.



In the editing interface, click on "Add". In the pop-up dialog, enter 9.



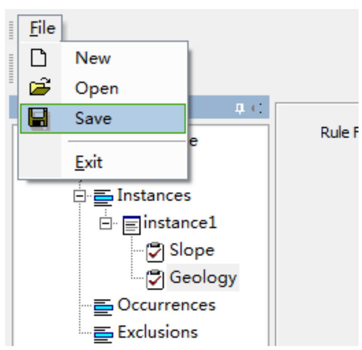
A green bar whose x coordinate is 9 will appear in the graphic area and 15 will also be added into the list.



Now we have encoded the knowledge on soil environmental conditions as rules. You can repeat the process for the other soil types.

Save the Project:

Let us save the project through "File->Save".

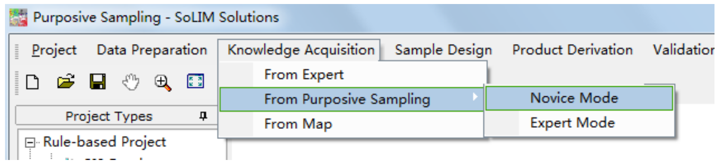


The project configuration and knowledge base will be saved into the project directory. At this point your project has been created and saved. You can use the control panel to navigate to the rules that have been created and make changes to them.

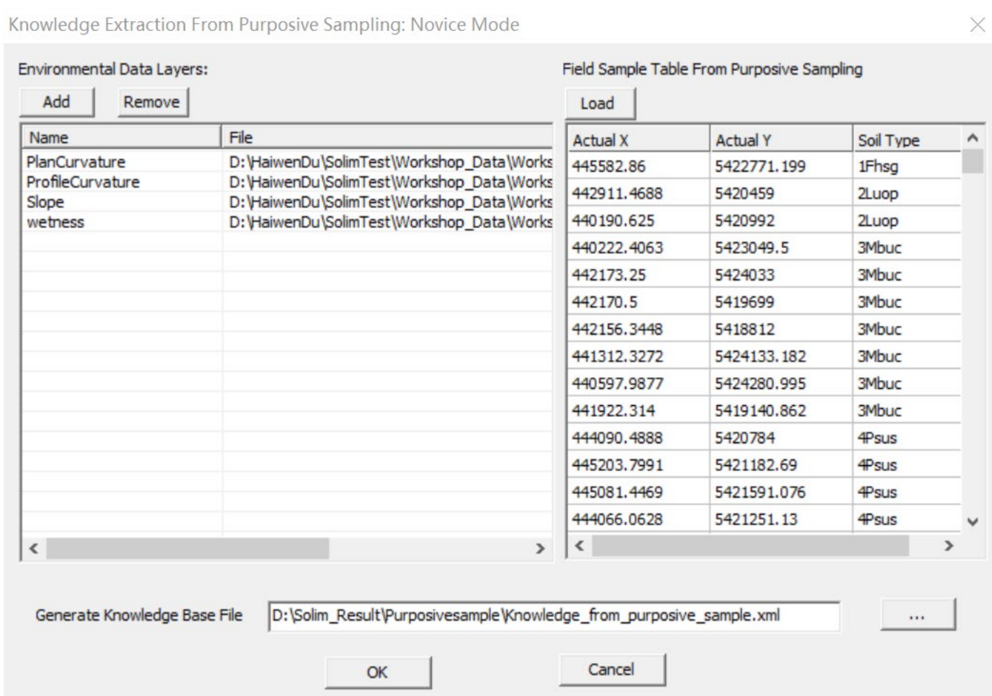
3.4.2 From Purposive Sampling

See [Purposive Sampling] for the design of purposive samples.

Select "Knowledge Acquisition-> From Purposive Sampling -> Novice Mode" to extract knowledge base from field sample table.

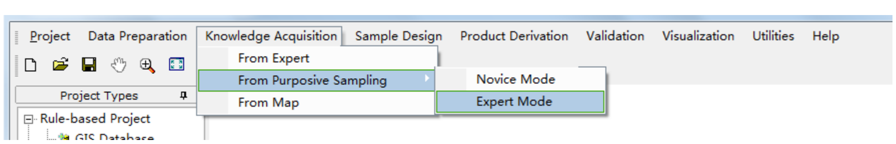


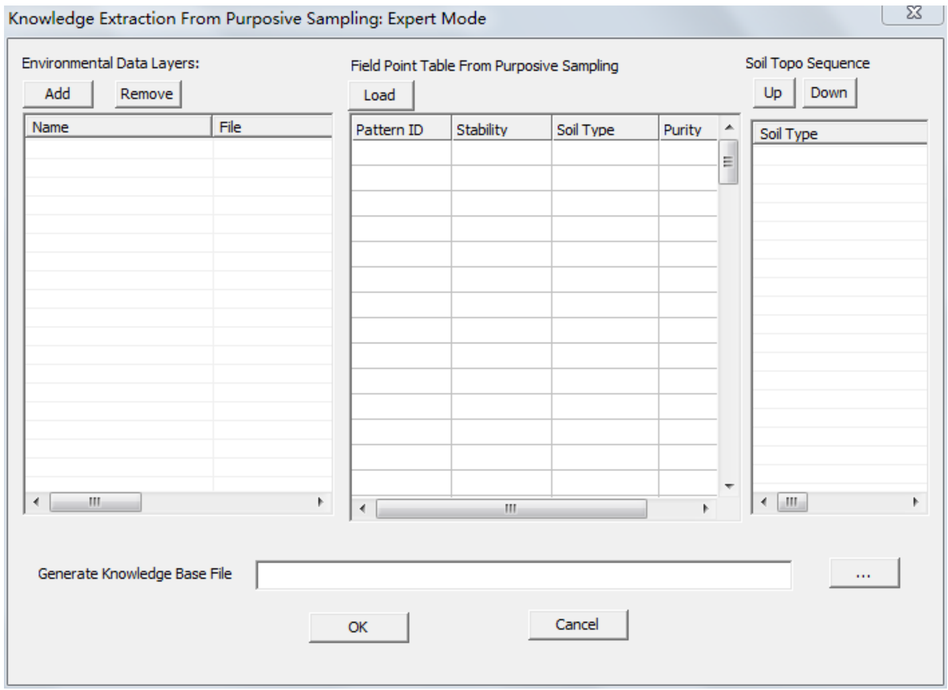
“Add” to load environmental layers from directory. Those layers are the same as the layers you used in sampling design. Load field sample table by clicking “Load” button. Specify the output knowledge base file.



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

Select "Knowledge Acquisition-> From Purposive Sampling -> Expert Mode" to extract knowledge base from field sample table.

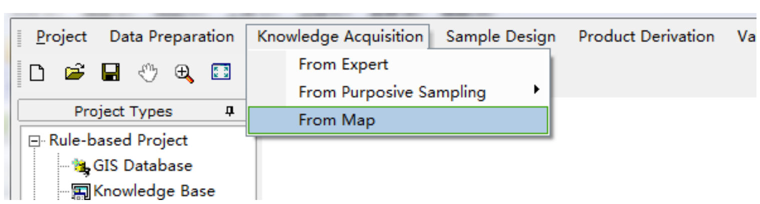




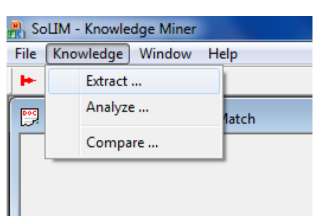
Click "Add" to load environmental layers from directory. Those layers are the same as the layers you used in sampling design. Load field sample table by clicking "Load" button. Adjust the sequence of soil type upward and downward. Specify the output knowledge base file. Click “OK”. A knowledge base file will be created.

3.4.3 From Map

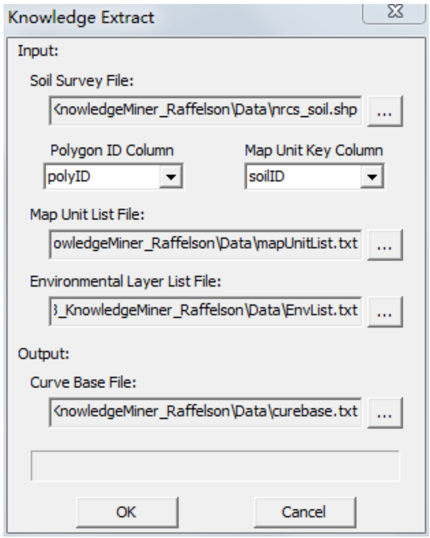
Select “Knowledge Acquisition->From Map” to start SoLIM-Knowledge Miner.



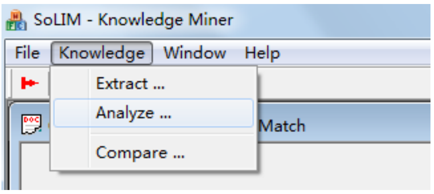
Start by hitting the button  or go to “Knowledge --> Extract…”.



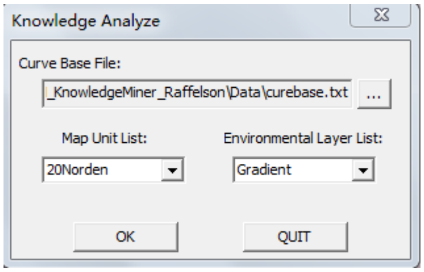
In the pop-up window, specify the path to the Soil Survey File. Choose the field (polyID) for polygon ID and the field (SoilID) for Map Unit Key in the following dropdown lists. Specify the path to the Map Unit List File and Environmental Layer List File. Set the name and the saving path of the output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



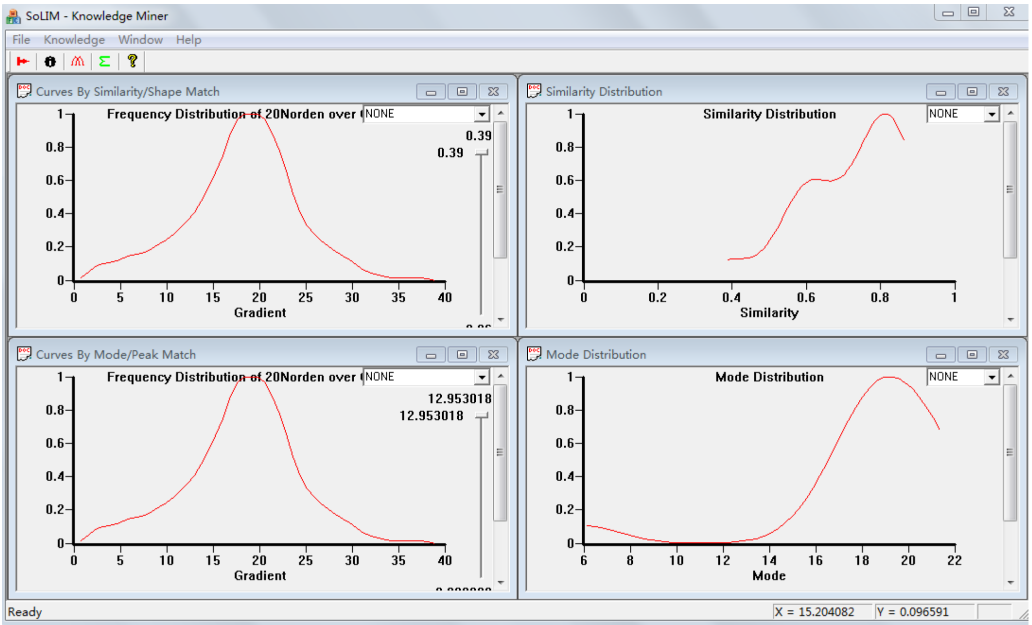
Knowledge analysis is normally performed for every combination of map unit and environmental data layer. Go to "Knowledge --> Analyze …" or hit the button on the tool bar to start the knowledge analysis interface.

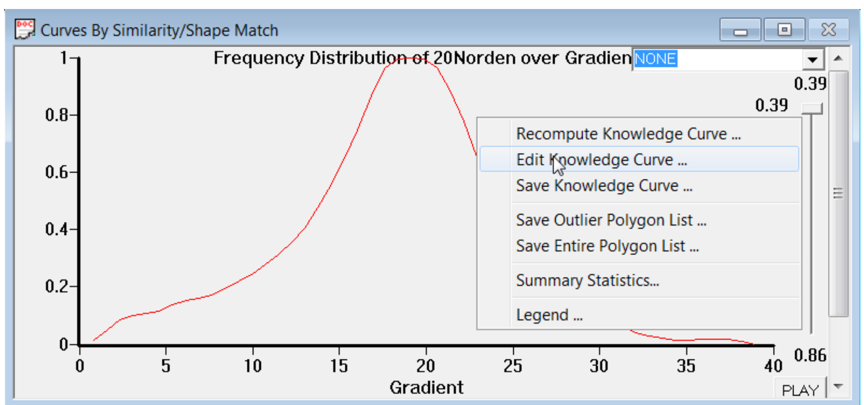


In the pop-up window "Knowledge Analyze" window, specify the path to the curve base file generated in "Extracting Knowledge" (step 1). Select a map unit and an environmental layer to be analyzed.

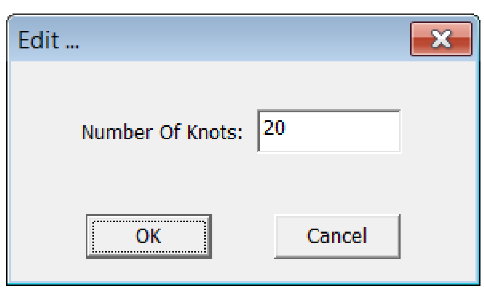


Click on "OK" and you will see the frequency distribution of the soil type you choose over the specified environmental variable displayed in the top-left panel. This curve can be used as the membership curve to the soil type over the specified environmental variable. This curve is editable. Right click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

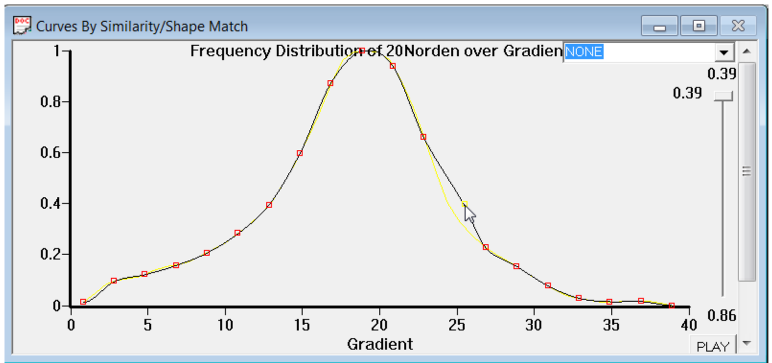




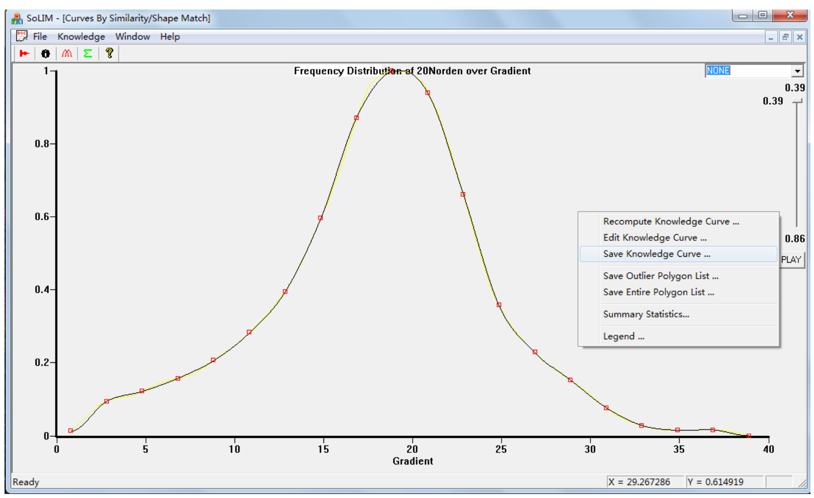
You may set the number of knots used to define the shape of the curve:

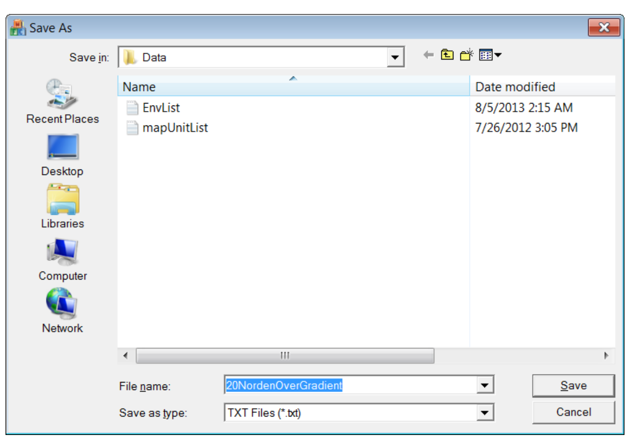


Click on OK and you will see the knots showing up along the curve. You can drag the knot(s) to adjust the shape of the curve.

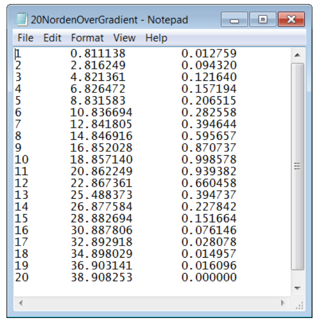


When you finish editing, you can save your edits. Right click on the curve and choose “Save Knowledge Curve”, the curve will be saved in a .txt file as following:





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.


20NordenOverGradient - Notepad

File Edit Format View Help

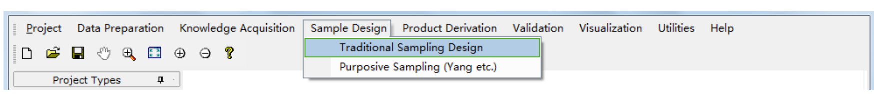
1	0.811138	0.012759
2	2.816249	0.094320
3	4.821361	0.121640
4	6.826472	0.157194
5	8.831583	0.206515
6	10.836694	0.282558
7	12.841805	0.394644
8	14.846916	0.595657
9	16.852028	0.870737
10	18.857140	0.998578
11	20.862249	0.939382
12	22.867361	0.660458
13	25.488373	0.394737
14	26.877584	0.227842
15	28.882694	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

3.5 Sample Design

- Traditional Sampling Design: generated samples with different sampling schema
- Purposive Sampling: purposive sample design based on Fuzzy c-Means (FCM) clustering of environmental data

3.5.1 Traditional Sampling Design

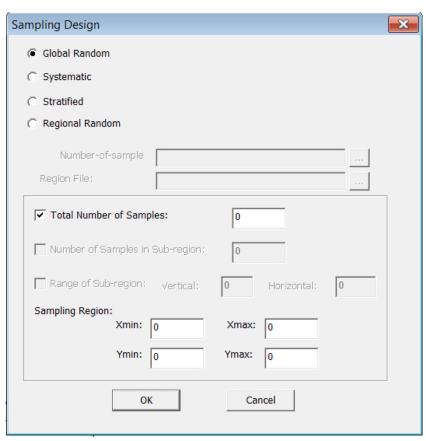
Select "Sample Design->Traditional Sampling Design" to generate a set of samples using user-defined sampling scheme:



Four sample strategies are available:

Method: *Global Random*

Every pixel in the entire area has an equal chance of being selected. You need to define the total number of samples and the sampling region.



Total Number of Samples

The total number of samples to be generated.

Xmin and *Xmax*

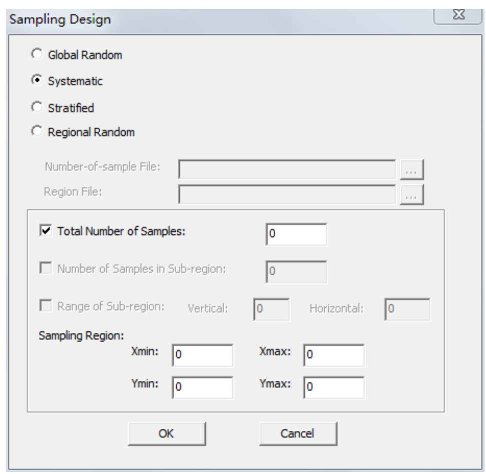
The minimum and maximum value of X coordinate value.

Ymin and *Ymax*

The minimum and maximum value of Y coordinate value.

Method: *Systematic*

After you defined the sample size, every Nth record is selected from a list of pixels. As long as the list does not contain any hidden order, this method is as good as the random sampling method. Its only advantage over the global random sampling method is simplicity.



Total Number of Samples

The total number of samples to be generated.

Xmin and *Xmax*

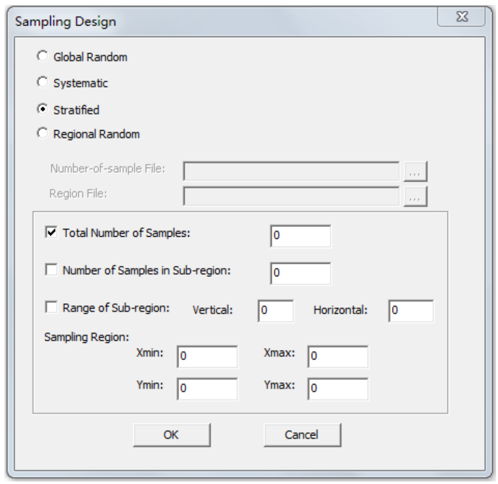
The minimum and maximum value of X coordinate value.

Ymin and *Ymax*

The minimum and maximum value of Y coordinate value.

Method: *Stratified*

Besides number of samples and sampling region, you have to define the number of samples in sub-region and range of sub-region. The program will randomly generate the same number of samples for each sub-region.



Total Number of Samples

The total number of samples to be generated.

Number of samples in Sub-region

The number of samples in each sub region.

Vertical

The vertical length of a sub region.

Horizontal

The horizontal length of a sub region.

Xmin and *Xmax*

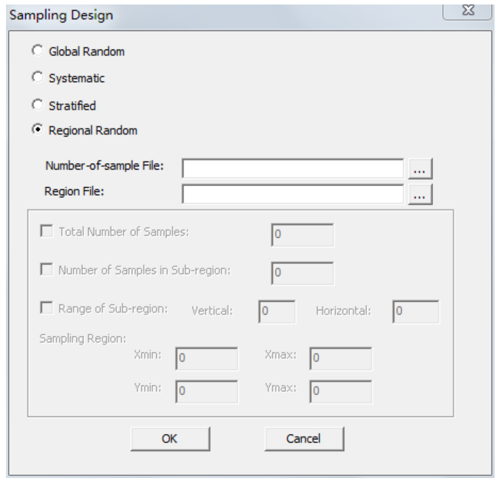
The minimum and maximum value of X coordinate value.

Ymin and *Ymax*

The minimum and maximum value of Y coordinate value.

Method: *Regional Random*

Unlike stratified sampling, you can have the program generate different number of samples for different sub-regions. The definition of regions is an ArcGIS Coverage ungenerated polygon text file.



number of samples file

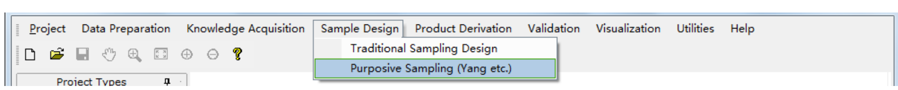
A text file to tell the program how many samples to be generated for each polygon.

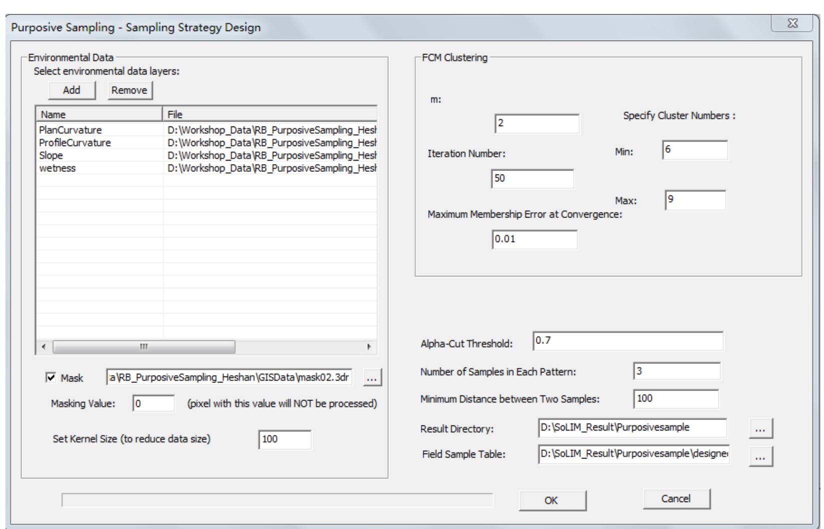
Region File

an ArcGIS Coverage ungenerated polygon text file.

3.5.2 Purposive Sampling

Select "Sample Design-> Purposive Sampling (Yang et al., 2012)" to generate suggested sample locations:





Environmental Data

The data of environmental covariates. Add the environmental layers by clicking “Add” button.

Mask (optional)

If you want to use mask, check the box next to "Mask" and specify a mask file and masking value.

Kernel Size

As FCM clustering is computation-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.

FCM Clustering

In this block, basic parameters for FCM clustering are set.

m

The fuzzifier parameter.

Iteration number

Maximum membership error at convergence

Specify cluster numbers: Min

Minimum cluster number.

Specify cluster numbers: Max

Maximum cluster number

Alpha-cut value

The value above which a fuzzy membership value can be regarded as high fuzzy membership.

Number of samples for each pattern

The number of samples needed for each detected pattern

Minimum distance between two samples

Result directory

The directory that holds all intermediate results.

Field sample table

The results of samples designed with purposive sampling.

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.

3.6 Product Derivation

- Property Map: create a soil property map from the soil fuzzy membership maps
- Hardened Map: Produce a soil class map by hardening the individual soil fuzzy membership maps.

3.6.1 Property Map

Select "Product Derivation->Property Map" to create soil property map. SoLIM Solutions uses the linear additive weighting function to estimate the local soil property value.

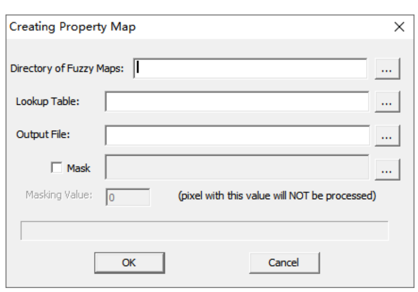
$$V_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{k} \cdot V^{K}}{\sum_{k=1}^{n} S_{ij}^{k}}$$

V_{ij}: the estimated soil property value at (*i, j*)

S_{ij}^{k}: the fuzzy membership value for *k*th soil at (*i, j*)

V^{K}: the representative property value for *k*th soil.

This approach requires that you prepare a lookup table to tell the program the representative property values for all soils:



Output File

The property map.

Directory of Fuzzy Maps

The directory where the soil fuzzy membership maps are stored.

Lookup Table

The file containing the representative property values for all soils. A lookup table is like:

soil type name1	*value1*
soil type name2	*value2*
......	......

Soil type name is soil fuzzy membership file name (without .3dr suffix) in "Result Directory" and values are the representative property values for those soils.

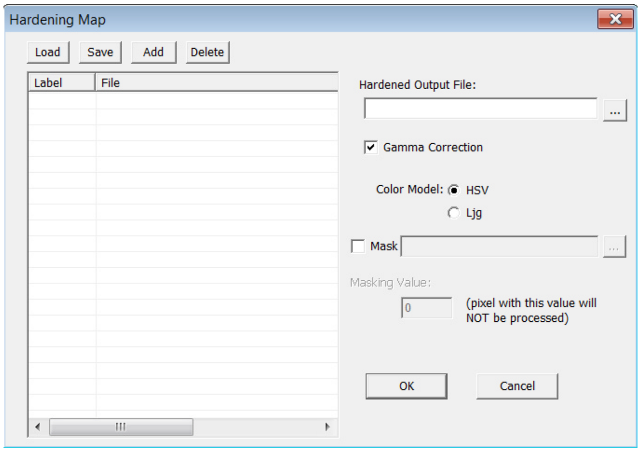
Mask

The mask to define the extent of mapping area (optional).

3.6.2 Hardened Map

Making a hardened map from a set of soil fuzzy membership maps assigns each pixel the ID of the soil type which has the maximum fuzzy membership value at the site. This ID is the unique ID (number) in the soil type name used in the knowledge file.

Select "Product Derivation->Hardened Map"



- click on "Add" to add a soil fuzzy membership map to the list.
- click on "Delete" to remove the selected soil fuzzy membership map.
- click on "Save" to save the list into a list file for future use. The list file record the labels and corresponding soil fuzzy membership maps.
- click on "Load" to load an existing list.

Hardened output file.

Gamma correction and color model options are used to set the color effect for the output files when they are visualized in 3dMapper.

Mask (optional)

If you want to use mask, check the box next to "Mask" and specify a mask file.

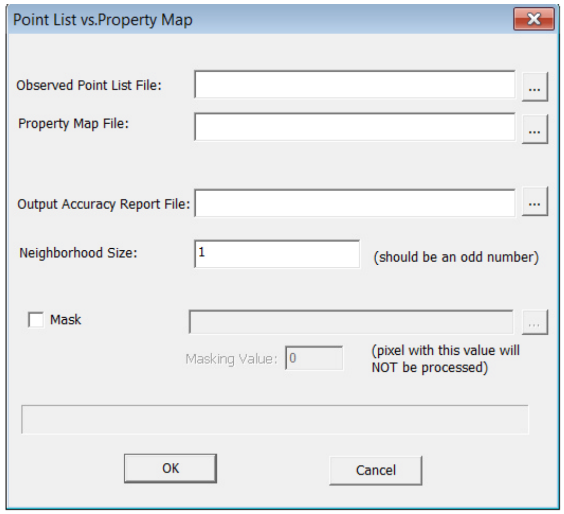
Click on "OK" to harden soil fuzzy membership maps in the list. If the hardening process is finished successfully, three files should be created in the in the same directory. One is the hardened map, the other two files are entropy map and exaggerated uncertainty map, respectively. For example, if your hardened output file is "HardenUnMap.3dr", the entropy file is named "HardenUnMapEnt.3dr" and the exaggerated uncertainty file is named "HardenUnMapExg.3dr". The entropy map and exaggerated uncertainty map are used to indicate the uncertainty at each pixel when assigning a soil class to that location. For detailed information, please refer to Appendix D: SoLIM Publications.

3.7 Validation

- Property Validation: create accuracy report for property map
- Type Validation: validate soil type inference result with different forms of input

3.7.1 Property Validation

Select "Validation->Property Validation", you can create accuracy report for property map.



Observed point list file

The information about the sample locations. A point file has the following format:

PointID	***X***	***Y***	***PropertyValue***
index1	*X1*	*Y1*	*property1*
index2	*X2*	*Y2*	*property2*
……	……	……	……

The first row contains the column headings. The first column *PointID* contains the IDs assigned to the sample points. *X* and *Y* are the coordinates for the points. *PropertyValue* contains the observed property values at sample locations.

Property map

The map of soil property in .3dr format.

Neighborhood size

The size of window over which the mean properties will be retrieved as the inferred property value.

Mask (optional)

If you want to use mask, check the box next to "Mask" and specify a mask file.

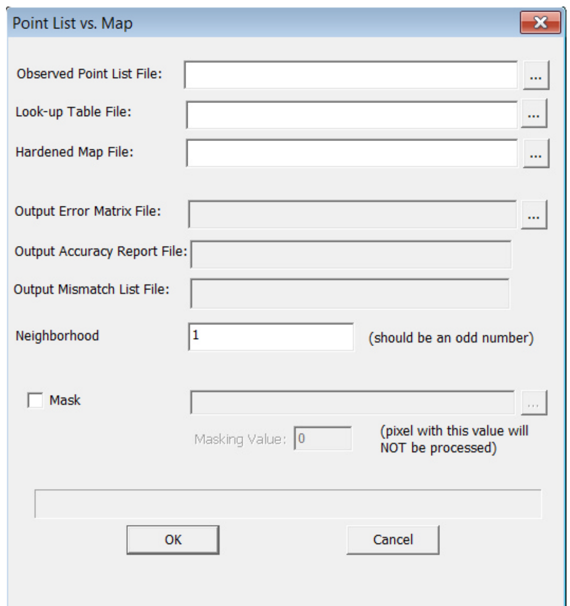
The output is an accuracy report which contains following four pieces of statistical information and the point list:

- RMSE (Root Mean Squared Error)
- Agreement Coefficient
- Mean Absolute Error
- Standard Deviation of Observed Values.

3.7.2 Type Validation

Point List *vs* Map

Select "Validation->Type Validation->Point List vs. Map", you can validate a hardened soil map with a point list file containing field observations.



Observed point list file

The information about the sample locations. The observed point list file has the following format:

PointID	***X***	***Y***	***SoilName***
index1	*X1*	*Y1*	*soilName1*
index2	*X2*	*Y2*	*soilName2*
……	……	……	……

The first row contains the column headings and *PointID* contains the IDs assigned to the sample points. *X* and *Y* are the samples coordinates. *SoilName* contains the observed soil names at sample locations.

Look-up table file

The soil IDs and the corresponding soil names. This file links the soil name in the field observation file with the ID in the hardened soil type map so that comparison can be made. A look-up table file has the following format:

SoilID	***SoilName***
soilID1	*soilName1*
soilID2	*soilName2*
......	......

The first row contains the column headings. *SoilID* contains the IDs assigned to the soils in the hardening process. These IDs are the IDs in the file name of the fuzzy membership file for the soil types. *SoilName* contains the soil type names used to describe the soil at the field points.

Hardened map file

Hardened soil type map in .3dr format. Different soil types in this map are differentiated based on the *SoilID*.

Neighborhood size

The window size over which the most frequently occurred soil types (*soilID*) will be used as the inferred soil type (*soilID*).

Mask (optional)

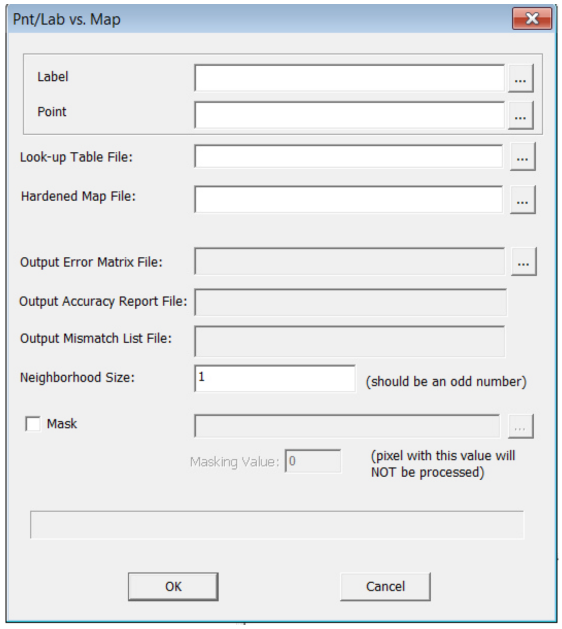
If you want to use mask, check the box next to "Mask" and specify a mask file.

Outputs

The outputs of this function include an error matrix file, an accuracy report file, and a mismatch list file. The names of three files share the same prefix which is chosen by users, but they have different suffixes. For example, if you choose "c:\validation" as the validation result file. SoLIM Solutions will create "c:\validation_Matrix.csv", "c:\validation_Report.txt" and "c:\validation_Mismatch.csv" for you.

Pnt/lab vs Map

Select "Validation->Type Validation->pnt/lab vs. Map", you can validate a hardened soil map with field observations in a .pnt and a .lab file pair.



Label

The names of the soils observed at the field sites. A label file has the following format:

InternalID	***SoilName***
index1	*soilName1*
index2	*soilName2*
……	……

The first column contains the column headings. *InternalID* contains the unimportant ID linking the soil names to the location information in the point file below. *SoilName* are the names of soils.

Point

The location information about the sample points. A point file has the following format:

InternalID	***X***	***Y***
index1	*X1*	*Y1*
index2	*X2*	*Y2*
……	……	……

The first column contains the column headings. *InternalID* is the unimportant ID as before. *X* and *Y* are the coordinates of sample points. SoLIM Solutions will use the index in label file to attach a soil name to a sample point. Therefore, every sample point will get a soil name.

Look-up table file

It records the soil IDs and the corresponding soil names. This file links the soil name from field observation with the ID in the hardened soil type map so that comparison can be made. A look-up table file has the following format:

SoilID	***SoilName***
soilID1	*soilName1*
soilID2	*soilName2*
......	......

The first row contains the column headings. *SoilID* contains the IDs assigned to the soils in the hardening process. These IDs are the IDs in the file name of the fuzzy membership file for the soil types. *SoilName* contains the soil type names used to describe the soil at the field points.

Hardened map file

Hardened soil type map in .3dr format. Different soil types in this map are differentiated based on the *SoilID*.

Neighborhood size

The window size over which the most frequently occurred soil types (*soilID*) will be used as the inferred soil type(*soilID*).

Mask (optional)

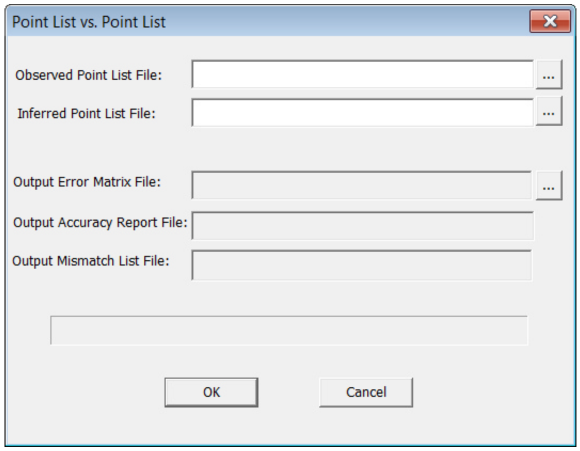
If you want to use mask, check the box next to "Mask" and specify a mask file.

Outputs

The outputs of this function include an error matrix file, an accuracy report file and a mismatch list file. The names of three files share the same prefix which is chosen by users, but they have different suffix. For example, if you choose "c:\validation" as the validation result file. SoLIM Solutions will create "c:\validation_Matrix.csv", "c:\validation_Report.txt" and "c:\validation_Mismatch.csv" for you.

Point List vs Point List

Select "Validation->Type Validation->Point List vs. Point List", you can validate inference by using two points list files, one of which recorded observed values and another of which records inferred results for field locations visited.



Observed point list file

The contains information about the sample locations. The observed point list file has the following format:

PointID	***X***	***Y***	***SoilName***
index1	*X1*	*Y1*	*soilName1*
index2	*X2*	*Y2*	*soilName2*
……	……	……	……

The first row contains the column headings and *PointID* contains the IDs assigned to the sample points. *X* and *Y* are the samples coordinates. *SoilName* contains the observed soil names at sample locations.

Inferred point list file

The information about the inferred results at samples locations. An inferred point list file has the following format:

PointID	***X***	***Y***	***InferredSoilName***
index1	*X1*	*Y1*	*InferredsoilName1*
index2	*X2*	*Y2*	*InferredsoilName2*
……	……	……	……

The first row contains the column headings and *PointID* contains the IDs assigned to the sample points. *X* and *Y* are the samples coordinates. *InferredSoilName* contains the inferred soil names at sample locations.

Note: The numbers and orders of points in observed point list file and inferred point list file should be identical.

Outputs

The outputs of this function include an error matrix file, an accuracy report file, and a mismatch list file. The name of three files share the same prefix which is chosen by users, but with different suffix. For example, if you choose "c:\validation" as the validation result file. SoLIM Solutions will create "c:\validation_Matrix.csv", "c:\validation_Report.txt" and "c:\validation_Mismatch.csv" for you.

3.8 Visualization

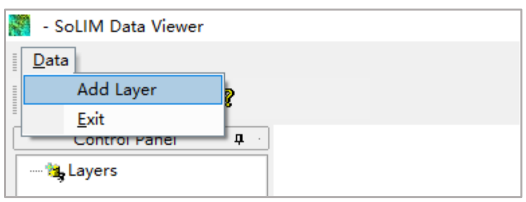
- 2D Visualization: visualize .3dr file in 2D
- 3D Visualization: visualize a 3dr file with 3dMapper

3.8.1 2D Visualization

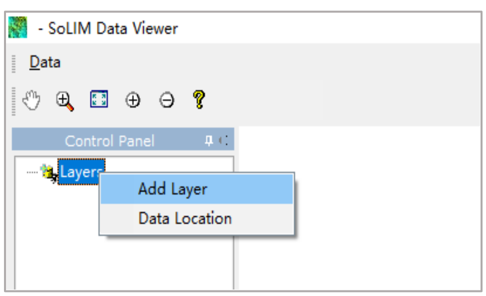
Choose "Visualization->2D" and open the Data Viewer.

There are two ways to add layers

1. Choose "Data->Add Layer".



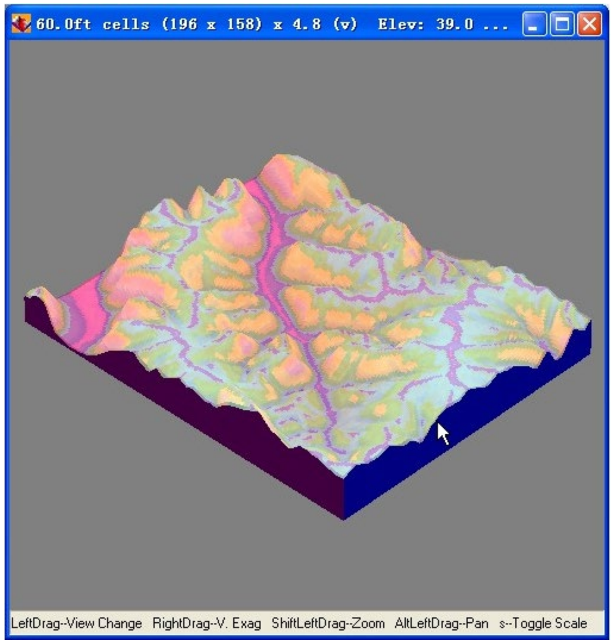
2. Right click "Layers" in the layer list panel. In the popup menu, choose "Add Layer".



You can open one layer or multiple layers at one time. You should now see those data layers are added to the left layer list panel. You can select one layer to view it the graphic area

3.8.2 3D Visualization

If the user chooses "Utilities->3D Visualization", the program will open 3dMapper (available at http://solim.geography.wisc.edu/software). You can use 3dMapper to view the environmental data layers and soil maps in the 3D view.



The program will read information in the registry of the operating system to locate 3dMapper. If it cannot find 3dMapper, it will prompt a dialog to let users find 3dMapper by themselves.

3.9 Utilities

Utilities Overview

- Data Format Conversion: convert the file between different file formats
- Values at Points: obtain the values at sample points
- Clip: clip 3dr file using a clip file
- Reclassify: reclassify the 3dr file
- Stretch: stretch the value of pixels in the 3dr file into a new user-defined interval
- Overlay: perform a set of overlay operations on two 3dr files
- Filter: pass a user-defined filter over the 3dr file
- Frequency: produce a list of the unique code occurrences and their frequency in a 3dr file
- Color: assign a color palette to a 3dr file
- FCM Clustering: conduct FCM clustering analysis of environmental data
- 3 Bands to Color File: generate color file form 3 bands

3.9.1 Data Format Conversion

In "Utilities->Data Format Conversion" there are a set of tools to convert files between different file formats:

GDAL - Supported Raster Formats → 3dr

SoLIM Solutions integrates GDAL (http://www.gdal.org), so it can convert other raster file of different format to a 3dr file as long as the format is supported by GDAL. It must be noted that if the raster file contains more than one bands, only the first band will be converted.

3dr → Grid Ascii

Convert a 3dr file to a ArcGIS Grid Ascii file.

GDAL - Supported Raster Formats → 3dm

SoLIM Solutions integrates GDAL (http://www.gdal.org), so it can convert other raster files of different format to a 3dm file as long as the format is supported by GDAL. It must be noted that if the raster file contains more than one bands, only the first band will be converted.

3dr → 3dm

Convert a 3dr file to a 3dm file.

Calc Statistics

Calculate the statistics of a 3dr file.

Shapefile → 3dr(rasterization)

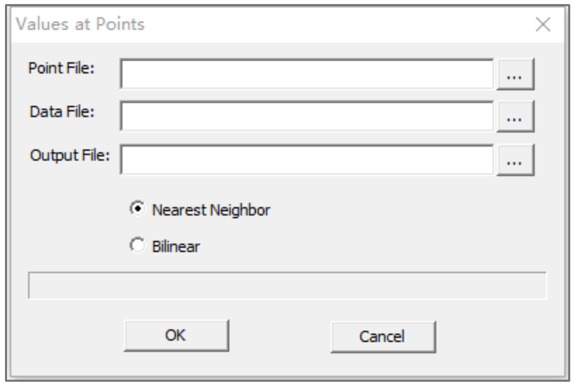
Rasterize shapefile to raster and store in .3dr format.

3dr → Saga

Convert a 3dr file to a saga file.

3.9.2 Values at Points

Select "Utilities->Values at Points". Specify a point file and a data layer (.3dr file), the program will find the data values at the points defined in the point file:



Output File

The list of values at given points.

Point File

The coordinates of points. The format of point file is:

label1	*x1*	*y1*
label2	*x2*	*y2*
	....	
labeli	*xi*	*yi*
	....	
labeln	*xn*	*yn*

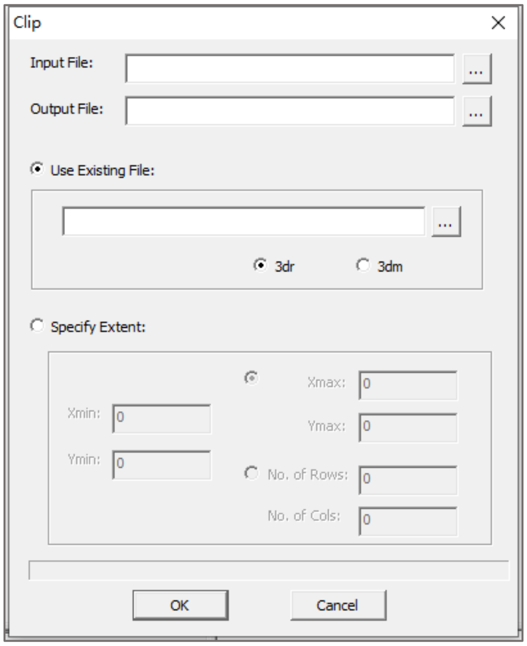
Labels are the IDs of the points; *xi* and *yi* are the coordinates of the points.

Data File

The map to extract values from. The data file is in the format of .3dr.

3.9.3 Clip

"Utilities->Clip" is a tool to extracts those features from an input 3dr file that overlap with an existing 3dr/3dm file or to clip the input 3dr file with specified extent.

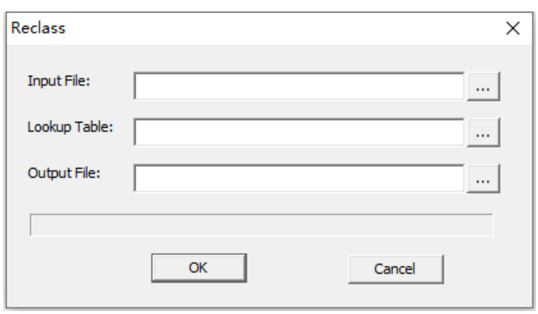


There are two way to clip a file:

- You can clip the *input file* using an *existing file*, which could be 3dr file or 3dm file.
- You can also specify an extent to clip the file.: the *Xmin*, *Ymin* need to be set firstly, then you can select setting *Xmax*, *Ymax* or the *number of rows* and *the number of columns*.

3.9.4 Reclassify

"Utilities->Reclassify" is a tool the reclassify the data in a .3dr file and save result into a new file.



Output File

The reclassified map.

Input File

The file to be reclassified, in the format of .3dr.

Lookup table

The information based on which the value of the input file is reclassified. The lookup table is like:

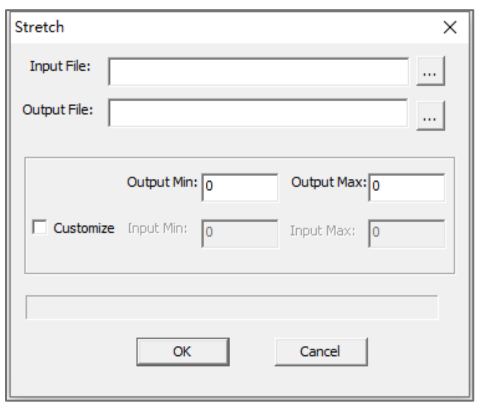
v1	*w1*
v2	*w2*
v3	*w3*

....

In the input file, pixels with value [v1, v2) are classified as w1, and so on.

3.9.5 Stretch

"Utilities->Stretch" is a tool to stretch the pixel values of the input file to a user-defined interval.



Output Min:

The minimum value of the output file.

Output Max:

The maximum value of the output file.

Customize:

Enable/Disable the "Input Min" and "Input Max".

Input Min:

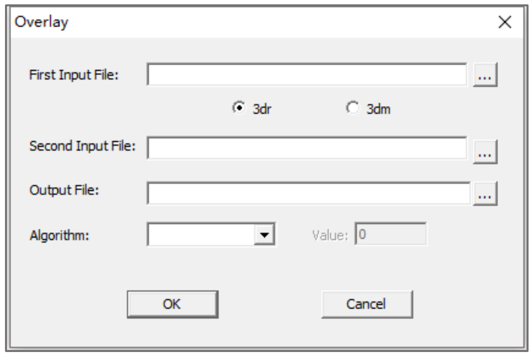
If the value of a pixel is smaller than the threshold, the value of the pixel will be set to "Output Min"

Input Max:

If the value of a pixel is larger than the threshold, the value of the pixel will be set to "Output Max"

3.9.6 Overlay

"Utilities->Overlay" is a tool to perform some overlay operations between two files.

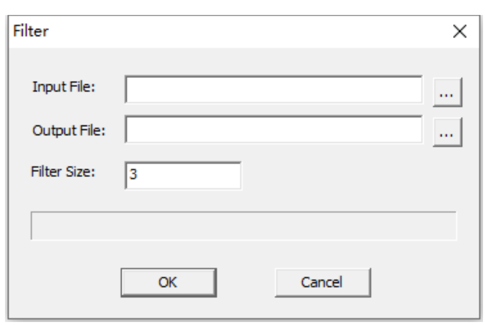


The algorithm has the following options: *Max, Min, Supplement, Select, Mask, +, -, *, /*

- *Max:* create a new 3dr file whose pixel value is the larger value of the respective pixels in the two input files.
- *Min:* create a new 3dr file whose pixel value is the smaller value of the respective pixels in the two input files.
- *Supplement:* use the pixel values in the second input file to supplement those in the first input file.
- *Select:* select pixels in the first input file on a pixel-by-pixel basis so that the respective pixels in the second input file satisfy the selection condition. You need to set the selection condition by setting the Value field.
- *+, -, *, /:* performs the algebra calculation between the two input files.

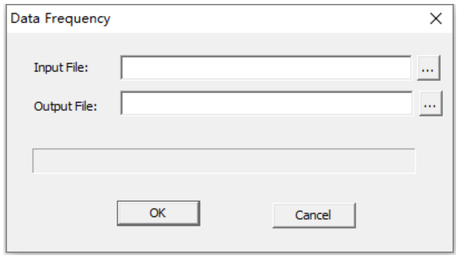
3.9.7 Filter

"Utilities->Filter" is a tool to filter the 3dr file by passing a mean-filter over the file. The default filter size is 3 x 3.



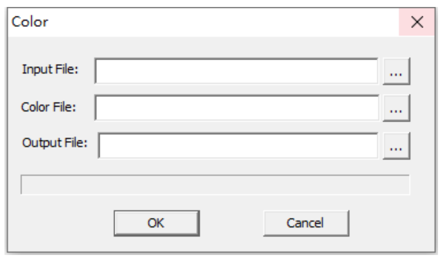
3.9.8 Frequency

"Utilities ->Frequency" is a tool to produce a list of the unique code occurrences and their frequency in a 3dr file.



3.9.9 Color

"Utilities->Color" is a tool to set a color palette for the displayed 3dr file.



You need to edit a Color Palette File, which has the format:

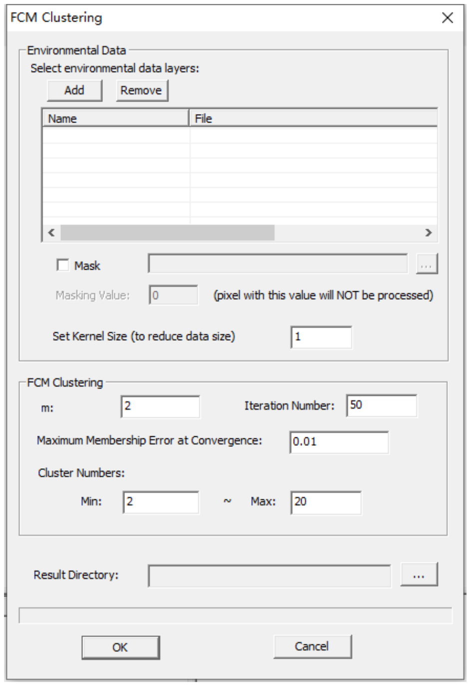
index1	*r1*	*g1*	*b1*
index2	*r2*	*g2*	*b2*
index3	*r3*	*g3*	*b3*
index4	*r4*	*g4*	*b4*

......

Each index is a floating-point number pointing to the pixel values in the data file. r, g, b are the RGB indices to define the color in RGB model.

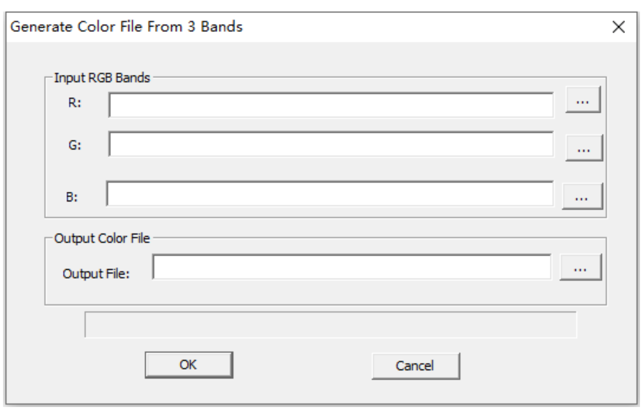
3.9.10 FCM Clustering

"Utilities->FCM Clustering" is a tool that divides the environmental data determined in advance to be classified into several categories, classifies them according to the optimal principle, and iterates for several times until the classification is reasonable.



3.9.11 3 Bands to Color File

"Utilities->3 Bands to Color File" is a tool to generate color file from 3 Bands.



4. Known problems and updates

Problem 1: Out of memory

Causes: This problem occurs when your study area is very large, or you are using very high-resolution datasets. The reason for this to occur is very simple, that is your computer does not have sufficient RAM to store the GIS data layers in computer memory. We are working on a solution to fix it.

Quick fix: The simplest solution is to divide the study area not smaller chunks and run the individual chunks respectively and then merge these chunks together once you are done with all of the chunks. One must be aware that there should be at least few pixels overlap among the neighboring chunks so that there are no edge effects between the chunks when the chunks are merged.

Problem 2: software crush when input character in the dialog

Causes: The edit box is not compatible with Chinese input.

Quick fix: change your input method to English.

Problem 3: software crush after right-click project item under windows 7 platform

Causes: The MFC view control might be edited by security software.

Quick fix: Change the visual effects to "Adjust for best appearance".

Update 1:

SoLIM Solutions now can read GIS data layers in GDAL supported formats, such as GeoTiff (*.tif), Erdas Imagine (*.img), SAGA GIS Binary format (*.sdat). Currently, only the Knowledge Acquisition from map functions do not support other formats.

Update 2:

Purposive sampling has also been added to Knowledge Acquisition and Sampling Design.

Appendices

In the appendices, some useful information is provided:

- Appendix A: the tutorial data set for quick start tutorial.
- Appendix B: the file suffix used in SoLIM Solutions.
- Appendix C: Glossary.
- Appendix D: the SoLIM publications.

Appendix A: Tutorial Data Set

Tutorial dataset for rule-based project

The tutorial dataset for rule-based project is stored in:

"<SoLIM Solutions installed directory>\Tutorial_Data\Rule_Based" and can be divided into 3 parts:

1. 3dMapper Base File

PleasantValley.3dm: the 3dMapper base file that will be used in 3D visualization

2. Environmental Data Layers

Aspect.3dr: the aspect file

Elevation.3dr: the elevation file

Geology.3dr: the geology data layer.

PlanformCurve.3dr: the planform curvature file

ProfileCurve.3dr: the profile curvature file

Slope.3dr: the slope file

3. Sample Word Library File

TestWordLib.wdl: It can be used to create word rule for the knowledge base

Tutorial dataset for sample-based project

The tutorial dataset for sample-based project is stored in:

"<SoLIM Solutions installed directory>\Tutorial_Data\Sample_Based" and can be divided into 2 parts:

1. Environmental Data Layers

Geology.3dr: the geology data layer.

Aspect.3dr: the aspect file

Elevation.3dr: the elevation file

PlanformCurve.3dr: the planform curvature file

ProfileCurve.3dr: the profile curvature file

Gradient.3dr: the slope file

Wetness.3dr: the wetness index file

2. Field Sample File

field_samples.csv

Appendix B: File Suffixes

.sip: SoLIM Solutions project file. Used to record project configurations and knowledge base.

.3dm: 3dMapper's data format for orthophoto-DEM pair. Used by 3dMapper for visualization and terrain attributes calculations.

.3dr: 3dMapper's raster data format. Used by SoLIM Solutions as the native raster data format. All the environmental data layers input to SoLIM Solutions must be in this data format. All the maps produced by SoLIM Solutions are in this format.

.csv: Comma-separated values file format. Can be edited by MS Excel or any text editor. Used by SoLIM Solutions as the filed sample base format in sample-based project.

.asc: The ASCII format for Arc/Info Grid. Used to convert raster data between 3dMapper/SoLIM Solutions and ArcGIS.

_chk.3dr: Check file created by SoLIM Solutions.

.pal: Color Palette Files, used Color.

.ply: Polygon File, a text file generated from an ArcGIS polygon coverage, used in Value at Sample Sites

.lkt: Look-up Table File, the look up table used in Type Validation, Property Map and Reclass.

.lab: label file for soils, used in Type Validation->pnt/lab vs. Map.

.pnt: point file for soils,used in Type Validation->pnt/lab vs. Map.

Appendix C: Glossary

Rule: A rule is a function defining the relationship between the values of an environmental feature and the optimality values for a soil type.

Instance: An instance is a representation of the soil scientist's knowledge of the relationship between a soil type and its environmental conditions characterized by topography, geology, climate, vegetation, and other environmental variables. It is not explicitly location-specific and is applicable to the whole mapping area, thus is usually considered to be the global knowledge. Note that a given soil type might appear in several instances, for example, a soil type may occur on both south facing and north facing slopes with different elevation and slope gradient conditions. These two environmental configurations form two instances of the soil. Technically, an instance refers to an environmental configuration defined by one or more rules, each of which corresponds to one environmental variable and characterizes the relationship between the optimality of the soil type and the environmental variable.

Occurrence: An occurrence is a positive exception, which means a particular soil will occur in places where the global knowledge does not cover. It is a representation of the soil scientist's knowledge of the relationship between a soil type and its environmental conditions characterized by topography, geology, climate, vegetation, and other environmental variables within limited geographic area. The area is called "influenced area" in SoLIM Solutions. In the influenced area, the relationship between soil type and the environment is different from other places. The influence area is defined by a central location and an impact radius. Technically, in SoLIM Solutions, an occurrence consists of one or more rules (rule part) and a spatial setting. When creating occurrences, the soil scientist needs to define rules that characterizes the relationship between the optimality of the soil type and the environmental variable. Besides, they need to pinpoint the coordinates of central location and provide the impact radius and a distance decay factor which determines how the optimality value will change as the distance to central location changes.

Exclusion: An exclusion is a negative exception, which means a particular soil will be very unlikely to occur in some places where the global knowledge does cover. It is a representation of the soil scientist's knowledge of the relationship between a soil type and its environmental conditions characterized by topography, geology, climate, vegetation, and other environmental variables within limited geographic area. The area is called "influenced area" in SoLIM Solutions. In the influenced area, the relationship between soil type and the environment is different from other places. The influenced area is defined by a central location and an impact radius. Technically, in SoLIM Solutions, an exclusion consists of one or more rules (rule part) and a spatial setting. When creating exclusion, the soil scientist needs to define rules that characterizes the relationship between the optimality of the soil type and the environmental variable. Besides, they need to pinpoint the coordinates of central location and provide the impact radius and a distance decay factor which determines how the optimality value will change as the distance to the central location changes.

Soil type: A soil type corresponds to a named soil and contains instance(s), occurrence(s) and exclusion(s). In many cases there will be only one instance for a soil type, but in some cases, there might be multiple instances or even occurrences and exclusions for a single soil type.

Knowledge base: A knowledge base is a collection of soil types. It is supposed to be created for a certain mapping area and contains all the soil types existing in the mapping area.

Global knowledge: Global knowledge refers to the knowledge that is effective in the whole mapping area. In SoLIM, global knowledge is expressed as instances. Each instance is a representation of the soil scientist's knowledge on the relationship between a soil type and its environmental conditions. An instance contains one or more rules.

Local knowledge: Local knowledge refers to the knowledge that takes effect within limited area. In SoLIM, local knowledge is expressed as occurrences and exclusions. An occurrence is a positive exception, which means a particular soil type will occur in places where the global knowledge does not cover. An exclusion is a negative exception, which means a particular soil type will be very unlikely to occur in some places where the global knowledge does cover. An occurrence/exclusion contains one or more rules plus spatial setting which defines the area being influenced.

Appendix D: SoLIM Publications

Below is a list of the representative publications on SoLIM. For more recent publications, please visit http://solim.geography.wisc.edu.

1) Overview of SoLIM

- Overview of SoLIM (February, 2007). http://solim.geography.wisc.edu/pubs/Overview2007-02-16.pdf

- Zhu, A.X., B. Hudson, J. E. Burt, and K. Lubich, 2001. "Soil mapping using GIS, expert knowledge, and fuzzy logic", Soil Science Society of America Journal, Vol. 65, pp. 1463-1472.

- Zhu, A.X., L.E. Band, B. Dutton, and T. Nimlos, 1996. "Automated soil inference under fuzzy logic", Ecological Modelling, Vol. 90, No. 2, pp. 123-145.

2) Representation Scheme

For soil data representation:

- Zhu, A.X., 1997. "A similarity model for representing soil spatial information", Geoderma, Vol. 77, pp. 217-242.

For soil-environment knowledge representation:

- Zhu, A.X., L. Yang, B.L. Li, C.Z. Qin, T. Pei, B.Y. Liu, 2013. "Construction of quantitative relationships between soil and environment using fuzzy c-means clustering", Geoderma, Vol. 155, No. 3-4, pp. 166-174.

3) Inference methods

- Zhu, A.X. and L.E. Band, 1994. "A knowledge-based approach to data integration for soil mapping", Canadian Journal of Remote Sensing, Vol. 20, No. 4, pp. 408-418.

- Shi, X., A.X. Zhu, J.E. Burt, F. Qi, and D. Simonson, 2004. "A case-based reasoning approach to fuzzy soil mapping", Soil Science Society of America Journal, Vol. 68, pp.885-894.

- Zhu, A.X., Liu, J., C. Qin, S. Zhang, Y. Chen, X. Ma, H. Zhang, R. Liu, T. Behrens, T. Scholten, 2015. “Predictive soil mapping with limited sample data”, European Journal of Soil Sciences, Vol. 66, 535–547.

- Zhu, A.X., 2008. “Rule based mapping”. In: J.P. Wilson and A.S. Fotheringham (eds.) The Handbook of Geographic Information Science, Blackwell, Malden, MA, pp. 273-291.

- Zhu, A.X., 2006. “Fuzzy logic models in soil science”. In Sabine Grunwald and Mary E. Collins (eds.) Environmental soil-landscape modeling: geographic information technologies and pedometrics, CRC/Taylor & Francis, New York, pp. 215-239.

- 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, 155(3-4), 164-174.

- Behrens, T., K. Schmidt, A.X. Zhu, T. Scholten, 2010. “The ConMap approach for terrain-based digital soil mapping”, European Journal of Soil Science, 61, 133-143.

- Zhu, A-Xing, Yamin Miao, Rongxun Wang, Tongxin Zhu, Yongcui Deng, Junzhi Liu*, Lin Ynag, Chengzhi Qing, Haoyuan Hong. 2018, "A comparative study of an expert knowledge-based model and two data-driven models for landslide susceptibility mapping.", Catena, 166, 317-327.

- Zhang, Guiming, A-Xing Zhu*, Steve K. Windels, ChengZhi Qin. 2018, “Modelling species habitat suitability from presence-only data using kernel density estimation.”, Ecological Indicators, 93, 387-396.

- Zhu, A-Xing, Yamin Miao, Junzhi Liu, Shibiao Bai, Canying Zeng, Tianwu Ma, Haoyuan Hong. 2019, "A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods.", Catena, 183, 104188.

- Zhu A-X, Lu G, Liu J, Qin C-Z, Zhou C. "Spatial prediction based on Third Law of Geography. " Annals of GIS, 2018, 24(4): 225-240.

- Qi, F., A.X. Zhu, M. Harrower, and J.E. Burt. 2006. "Fuzzy soil mapping based on prototype category theory", Geoderma, Vol. 136, pp. 774-787.

4) Knowledge Acquisition

From Experts

- Zhu, A.X. 1999. "A personal constructed-based knowledge acquisition process for natural resource mapping using GIS", International Journal of Geographic Information Systems., Vol. 13, No. 2, pp. 119-141.

- Liu, J., Zhu, A.X., 2009. “Mapping with words: a new approach to automated digital soil survey”, International Journal of Intelligent Systems, Vol. 24, No. 3, pp. 293-311.

- Zhu, A.X., R.X. Wang, J.P. Qiao, C.Z. Qin, Y.B. Chen, J. Liu, F. Du, Y. Lin, T.X. Zhu, 2014. “An expert knowledge-based approach to landslide susceptibility mapping using GIS and fuzzy logic”, Geomorphology, 214, pp.128–138.

From Maps

- Qi, F., A.X. Zhu, 2003. "Knowledge discovery from soil maps using inductive learning", International Journal of Geographical 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.

- Yang, L., F. Sherif, Y. Jiao, H. Sheldon, A.X. Zhu, C.Z. Qin, Z.G. Xu, 2011. “Updating conventional soil maps through digital soil mapping”, Soil Science Society of America Journal. Vol. 75, NO. 3, pp. 1044-1053.

- Du, Fei, A-Xing Zhu, Jing Liu, Lawrence Band, 2015. “Soil Property Variation Mapping through Data Mining of Soil Category Maps for Hydrological Modelling”, Hydrological Processes, 29, 2491-2503.

- Liu Xueqi, A-Xing Zhu, Lin Yang, Tao Pei, Junzhi Liu, Canying Zeng, Desheng Wang. 2020. “A graded proportion method of training sample selection for updating conventional soil maps.”, Geoderma, 357: 113939.

From field points

- Zhu, A.X. 2000. "Mapping soil landscape as spatial continua: the neural network approach", Water Resources Research, Vol. 36, No. 3, pp. 663-677.

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

- 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.

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

- Zhang, Shu-Jie, A-Xing Zhu*, Jing Liu, Cheng-Zhi Qin, Yi-Ming An, 2016. "An heuristic uncertainty directed field sampling design for digital soil mapping", Geoderma, 267, 123-136.

- Yang, Lin, Feng Qi, A-Xing Zhu*, Jingjing Shi, Yiming An, 2016, "Evaluation of Integrative Hierarchical Stepwise Sampling for Digital Soil Mapping", Soil Science Society of America Journal. 80, 637-651.

- Li, Yan, A-Xing Zhu, Zhou Shi, Jing Liu, Fei Du, 2016. "Supplemental sampling for digital soil mapping based on prediction uncertainty from both the feature domain and the spatial domain", Geoderma, 284, 73-84.

- Yang, Lin, Dick J. Brus*, A-Xing Zhu, Xinming Li, Jingjing Shi, 2018, "Accounting for access costs in validation of soil maps: A comparison of design-based sampling strategies." Geoderma, 315, 160-169.

5) Environmental Covariates Characterization

- Smith, M.P., A.X. Zhu*, J.E. Burt, C. Stiles, 2006. "Effects of DEM resolution and neighborhood size on digital soil survey", Geoderma,Vol. 137, pp.58-69.

- 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. Vol. 74, NO. 3, pp. 861-869.

- Liu, F., X. Geng, A.X. Zhu, W. Fraser, 2012. "Soil texturemapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS", Geoderma, 171–172 (2012) 44–52.

- Qin C, Zhu A-X, Pei T, Li B, Zhou C, Yang L. An adaptive approach to selecting a flow-partition exponent for a multiple-flow-direction algorithm. International Journal of Geographical Information Science, 2007, 21(4): 443-458.

- Qin C-Z, A-X Zhu, T Pei, B-L Li, T Scholten, T Behrens, C-H Zhou. An approach to computing topographic wetness index based on maximum downslope gradient. Precision Agriculture, 2011, 12(1): 32-43.

- Qin C-Z, Zhu A-X, Shi X, Li B-L, Pei T, Zhou C-H. Quantification of spatial gradation of slope positions. Geomorphology, 2009, 110: 152-161.

- Qin, C.Z., A.X. Zhu, W.L. Qiu, Y.J. Lu, B.L. Bao, T. Pei, 2012. "Mapping soil organic matter in small low-relief catchments using fuzzy slope position information", Geoderma, 171-172(2012) 64-74.

- Guo, Shanxin, Lingkui Meng, A-Xing Zhu*, James E. Burt, Fei Du, Jing Liu and Guiming Zhang, 2015. "Data-Gap Filling to Understand the Dynamic Feedback Pattern of Soil", Remote Sensing, 7(9), pp.11801-11820.

- Guo, Shanxin, A-Xing Zhu*, Lingkui Meng, James E. Burt, Fei Du, Jing Liu, Guiming Zhang, 2016. "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.

- 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 Low Relief Areas Using Feedback Dynamic Patterns”, Soil Science Society of America Journal. 74(3), 861-869.

- Zeng, Canying, A-Xing Zhu*, Feng Liu, Lin Yang, David G. Rossiter, Junzhi Liu, Desheng Wang, 2017. "The impact of rainfall magnitude on the performance of digital soil mapping over low-relief areas using a land surface dynamic feedback method". Ecological Indicators.72: 297-309.

- Yang, Lin, Ming Song, A-Xing Zhu, Cheng-Zhi Qin, Chenghu Zhou, Feng Qi, Xinming Li, Ziyue Chen, 2019. “Predicting soil organic carbon content in croplands using crop rotation and Fourier transform decomposed variables”, Geoderma, 340:289-302.

6) Soil Property Derivation

- Zhu, A.X., L.E. Band, R. Vertessy, B. Dutton, 1997. "Derivation of soil properties using a Soil-Land Inference Model (SoLIM)", Soil Science Society of America Journal, Vol. 61, No. 2, pp. 523-533.

- Zhu, A.X., F. Qi, A. Moore, J.E. Burt, 2010. “Prediction of soil properties using fuzzy membership”, Geoderma, Vol. 158, pp. 199-206.

- Zhang, S.J., A.X. Zhu, W.L. Liu, J. Liu, L. Yang. 2013. “Mapping detailed soil property using small scale soil type maps and sparse typical samples”, Chinese Geographical Science, 23(6), 680-691.

- Du, Fei, A-Xing Zhu, Lawrence Band, Jing Liu, 2015. “Soil Property Variation Mapping through Data Mining of Soil Category Maps”, Hydrological Processes, 29, 2491-2503.

- Shi Jingjing, Lin Yang, A Zhu, Chengzhi Qin, Peng Liang, Canying Zeng, Tao Pei, 2018. “Machine-learning variables at different scales vs. Knowledge-based variables for mapping multiple soil properties”, Soil Science Society of America Journal, 82(3): 645-656.

- Fan N-Q, Zhu A-X, Qin C-Z, Liang P. Digital soil mapping over large areas with invalid environmental covariate data. ISPRS International Journal of Geo-Information, 2020, 9, 102.

7) Uncertainty Quantification and Visualization

- Zhu, A.X., 1997. "Measuring uncertainty in class assignment for natural resource maps under fuzzy logic", Photogrammetric Engineering & Remote Sensing, Vol. 63, pp. 1195-1202.

- Qi, F., A.X. Zhu, 2011. "Comparing three methods for modeling the uncertainty in knowledge discovery from area-class soil maps", Computers & Geosciences 37, 1425-1437.

- Burt, J.E., A.X. Zhu, M. Harrower, 2011. "Depicting classification uncertainty using perception-based color models", Annals of GIS, 17(3): 147-153.

- Qin, C.Z., L.L. Bao, A.X. Zhu*, R.X. Wang, X.M. Hu, 2013. "Uncertainty due to DEM error in landslide susceptibility mapping", International Journal of Geographical Information Science, 27(7), 1364-1380.

8) Detailed Soil Mapping and Environmental Modeling

- A.X. Zhu, 1998. "Fuzzy inference of soil patterns: implications for watershed modeling", in D.L. Corwin, K. Loague, and T.R. Ellsworth (eds.) Application of GIS, Remote Sensing, Geostatistical and Solute Transport Modeling to the Assessment of Nonpoint Source Pollution in the Vadose Zone, American Geophysical Union, Washington, DC.

- Zhu, A.X., D. Scott Mackay, 2001. "Effects of spatial detail of soil information on watershed modeling", Journal of Hydrology, Vol. 248, pp. 54-77.

- Quinn, T., A.X. Zhu*, and J.E. Burt, 2005. "Effects of detailed soil spatial information on watershed modeling across different model scales", International Journal of Applied Earth Observation and Geoinformation, Vol. 7, pp.324-338.

- Sun, Xiaolin, Yuguo Zhao, Huili Wang, Lin Yang, Chengzhi Qin, A-Xing Zhu*, Ganlin Zhang, Tao Pei, Baolin Li, 2012. "Sensitivity of digital soil maps based on FCM to the fuzzy exponent and the number of clusters", Geoderma, 171-172, 24 –34.

- Li, Runkui, A-Xing Zhu, Xianfeng Song*, Baolin Li, Tao Pei, Chengzhi Qin, 2012. "Effects of Spatial Aggregation of Soil Spatial Information on Watershed Hydrological Modeling", Hydrological Processes, 26(9), 1390-1404.

- Zhu, A.X., P. Wang, T.X. Zhu, L. Chen*, Q. Cai, H. Liu, 2013. "Modeling runoff and soil erosion in the three-gorge reservoir drainage area of China using limited plot data", Journal of Hydrology, 492, 163-175.

9) Software

- Zhu, A-X., Qin, C.-Z., Liang, P., Du, F., 2018. "Digital Soil Mapping for Smart Agriculture: the SoLIM method and Software Platforms", RUDN Journal of Agronomy and Animal Industries, 13(4), 317-335.