



DOI: 10.22363/2312-797X-2018-13-4-317-335

DIGITAL SOIL MAPPING FOR SMART AGRICULTURE: THE SOLIM METHOD AND SOFTWARE PLATFORMS

A.X. Zhu^{1,2,3}, C.Z. Qin², P. Liang², F. Du³

¹Nanjing Normal University
Nanjing, 210023, China

²Institute of Geographic Sciences and Natural Resources Research
Beijing, 100101, China

³University of Wisconsin-Madison
Madison, Wisconsin, 53706, USA
azhu@wisc.edu

Abstract. The key challenges faced by many of the existing digital soil mapping (DSM) techniques are the rigid requirements on the size of soil samples to extract the relationships needed and on the stationarity of the extracted relationships. These requirements limit the application of these DSM techniques. This paper provides an overview of the SoLIM approach and an introduction to the operation of SoLIM through the software platforms available. SoLIM is based on the Third Law of Geography, which calls for the comparison of similarity in geographic (environmental) configuration of a prototype and an unsampled location and then use this similarity to predict the value of a soil property at a given location. DSM under SoLIM approach removes requirements on the sample size and the stationarity assumption. In addition, the uncertainty computed based on the similarities can be used to improve the efficiency of error reduction efforts. The SoLIM approach has been implemented in two platforms: SoLIM Solutions and CyberSoLIM. The theoretical foundation and the availability of software platforms under SoLIM make DSM possible and convenient over large and complex geographic regions.

Keywords: digital soil mapping, DSM, SoLIM, First Law of Geography, Second Law of Geography, Third Law of Geography, spatial prediction

1. INTRODUCTION

Smart agriculture, as a way to increase the productivity of agricultural lands while minimizing the negative impacts on the environment, must base its practices on detail information about the status of agricultural land and information on how this status varies over space. Data on the status of soils and its spatial variation across landscape is an essential part of the information about agriculture lands. Among the many methods for acquiring information on soil conditions, digital soil mapping (DSM), an emerging area in this field, is a major approach to soil spatial information gathering [1].

DSM techniques, an application of spatial prediction in soil mapping, are mostly based on three basic principles [2]: spatial autocorrelation principle (also referred to as the First Law of Geography [3]), the statistical principle, and the spatial heterogeneity (also referred to as the Second Law of Geography [4]), or the combination of these principles (such as the various versions of kriging [5] and geographically weighted regression [6]). The key challenges to the techniques based on these principles are: 1) the requirement of samples of sufficient size for the extraction of the spatial relation-

ships or covariate relationships needed for soil prediction; and 2) the stationarity assumption of the extracted relationships [1]. These requirements have limited the application of this type of techniques for DSM over large and complex geographic areas where collecting a sample set sufficient enough is prohibitively expensive and where geographic processes are so complex that the stationarity assumption required often does not hold.

In recognition of these limitations faced by the techniques based on above principles Zhu [7] and Zhu et al. [8] presented a similarity approach to DSM (referred to as the Soil Land Inference Model (SoLIM) approach). The basis behind this similarity approach is what now been referred to as the Third Law of Geography [1]. This paper provides an overview of the theoretical thinking, the implementation and operation of this similarity approach. The next section describes the theoretical grounding of the SoLIM approach which is then followed by a presentation on how this idea is implemented. Software realization and operation of the SoLIM idea are presented in Section 4. Future research issues related to SoLIM are discussed in Section 5.

2. SOLIM AS AN APPLICATION OF THE THIRD LAW OF GEOGRAPHY

The basic idea behind the SoLIM approach is another principle which has been commonly applied and now referred to as the Third Law of Geography [1]. This law states that “The more similar geographic configurations of two points (areas), the more similar the values (processes) of the target variable at these two points (areas)”. The SoLIM approach exploits the comparative nature of this law in predicting soil conditions at an unsampled location. With this comparative nature the soil property value at an unsampled location can be estimated by the similarities in the environmental configuration between the unsampled location and a known prototype available (Figure 1). A prototype in the sense of soil mapping can be perceived as the central concepts of a soil class, a representative case of the class, or a field sample [9]. Under this notion, the process of DSM under the SoLIM approach can be accomplished in three steps: First, the similarity in environmental configuration ($S_{i,j,E}^k$) between a prototype (k) and the unsampled location (i, j) is measured. The computed $S_{i,j,E}^k$ is used to approximate the similarity ($S_{i,j}^k$) in soil attribute value between prototype k and the unsampled point (i, j). Finally, the attribute value ($V_{i,j}$) at the unsampled point (i, j) is calculated through Equation 1 below [10]:

$$V_{i,j} = \sum_{k=1}^n w_{i,j}^k \cdot v_k, \quad \text{Equation 1}$$

where n is the number of prototypes involved, v_k is the soil attribute value for prototype k and $w_{i,j}^k$ is the weight assigned to prototype k for the unsampled point (i, j) and is calculated using Equation 2 [11].

$$w_{i,j}^k = \frac{S_{i,j}^k}{\sum_{l=1}^n S_{i,j}^l}. \quad \text{Equation 2}$$

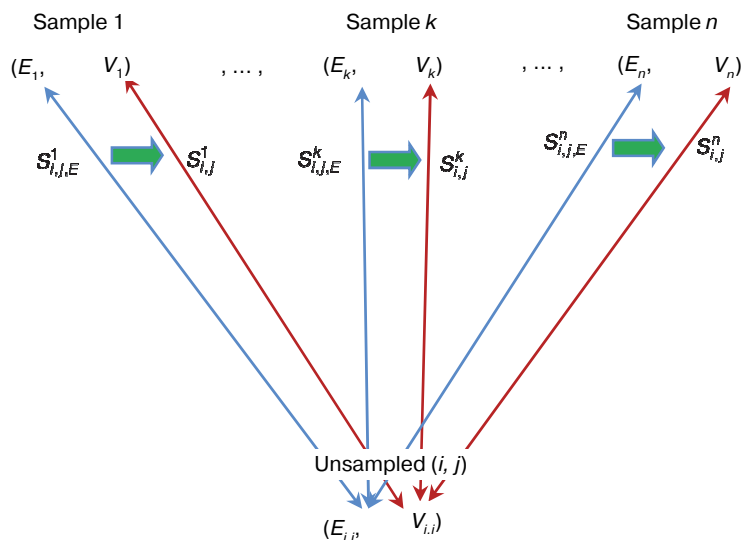


Fig. 1. The comparative nature of the Third Law of Geography used in SoLIM

E_k is the environmental configuration at prototype k and $E_{i,j}$ is the environmental configuration at unsampled point (i, j) . V_k is the soil attribute value at prototype k and $V_{i,j}$ is the soil attribute value at unsampled point (i, j) . $S^k_{i,j,E}$ is the similarity in environmental configuration between unsampled point (i, j) and prototype k . $S^k_{i,j}$ is the similarity in soil attribute value between unsampled point (i, j) and prototype k . $S^k_{i,j}$ is approximated by $S^k_{i,j,E}$

Figure 1 clearly shows that under the SoLIM approach the weight assigned to each prototype involved is based on the similarity of the unsampled point with each of individual prototypes, which is what the Third Law of Geography calls for. In this notion, no general relationships need to be extracted and quantified. Instead, the similarity in environmental configuration, as an individual representativeness of a prototype to an unsampled location, is captured and thus local variations in soil conditions can be expressed [12]. This similarity can also be used to compute the uncertainty for each prediction [13]. This uncertainty can be used to assess the quality of the so predicted results and be used to allocate future sampling efforts to further improve the quality of the prediction [12, 14, 15]. The use of individual prototypes and the uncertainty measure associated with the prediction using these prototypes under SoLIM removes the requirements on the specific number of prototypes (or samples) needed, the requirement on spatial distribution of these prototypes (or samples), and the stationarity assumption [1]. The impacts of the SoLIM approach on soil mapping have been documented in other studies [16, 17] and are not repeated here.

3. IMPLEMENTATION AND DEVELOPMENT OF SOLIM

The key issues to the success of the SoLIM idea as expressed above are the quantification of prototypes and the characterization of environmental configuration. Thus, the implementation and development of the SoLIM approach have focused on these two areas so far.

3.1. Quantification of prototypes

As stated earlier a prototype can be a field soil sample or the central concept of a soil class. The quantification of a prototype under the SoLIM approach consists of 1) the derivation of a set of covariates to characterize the geographic configuration for the prototypes and 2) the property values of the prototype.

Field soil samples as prototypes: Clearly, the field soil samples are good sources for prototypes under the SoLIM approach. Each sample has soil attribute values observed. The environmental configuration of a sample may be characterized in the field directly but more than often it is done after the fact using geographic information processing techniques and remote sensing methods based on the (x,y) coordinate values of the sample location. In other words, the two elements of prototype can be very conveniently defined with the use of field sample points as prototypes. Thus, samples naturally fit the use of the Third Law of Geography well with the SoLIM approach.

Central concepts of soil classes as prototypes: For a soil classes we often have information about the typical values and the ranges of soil properties. In this case the property values of the prototype representing this class is not difficult to obtain. The characterization of environmental configuration for the prototype would present challenges due to the fact that the environmental configuration is often not complete, sometimes even not available at all in soil survey reports. However, local soil scientists, particularly local soil surveyors, would normally understand under what kind of environmental conditions (or configurations) the soils belonging to a particularly soil class would exist or develop. This information is very useful in defining the environmental configurations for soil classes. To obtain environmental configurations of soil classes from local soil surveyor, Zhu [18] developed a personal construct based approach for obtaining knowledge from local experts on environmental configuration. The approach employs fuzzy logic to express the environmental conditions where a soil type will develop fully (assigned a fuzzy membership of 1) and where the soil type does not develop at all (assigned a fuzzy membership of 0) and where the soil type develop at half (assigned a fuzzy membership value of 0.5). This approach of defining a prototype is available through the SoLIMSolutions software described later in this manuscript.

The other source of knowledge on the environmental configuration for a soil class is existing soil maps where the spatial distribution of soil classes is portrayed. This type of maps would implicitly contain the information needed to define and characterize the environmental configuration for a soil class. Qi and Zhu [19] developed an inductive learning (decision tree) approach to extract environmental configurations for soil classes from soil maps. Cheng et al. [20] furthered this effort to make a use of the knowledge captured at the individual polygon level. This capability is also provided in the SoLIM Solutions software.

Due to the fact these environmental configurations are extracted from human experts or from existing soil maps in the form of knowledge rules, soil prediction under the SoLIM approach using environmental configurations extracted in this way is referred to as “rule-based”. It must be noted that this “rule-based” is not the same as the relation-

ships used in the statistical approaches for the following two reasons. The first is that these “rules” are not expressed in any fixed quantitative form as those in the statistical approaches. Second, these “rules” are used to describe the environmental configuration, not to relate environmental covariates to soil property values directly.

3.2. Characterization of environmental configuration

Characterization of environmental configuration calls for 1) a comprehensive list of covariates that can effectively describe the geographic environment relevant to a soil property; 2) the hierarchy of these variables; 3) the spatial foot prints of the target soil properties. The current efforts environmental configuration characterization for the SoLIM approach has been focus on the development of comprehensive list of covariates with initial research underway on the other two.

In addition to conventional soil covariates used to describe the soil formative environment (climatic conditions, topographic conditions, geological conditions, vegetation conditions), the SoLIM effort has added two new environmental variables into the list. The first one is the fuzzy landscape positions (fuzzy slope components) characterizing slope positions (such as ridge top, shoulder slope, backslope, footslope and valley bottom) in the form of fuzzy membership value [21]. This way of characterizing slope positions allows the gradation from one slope position to another to be represented in the covariate dataset and makes the characterization more realistic than the Boolean slope partitioning.

The other covariate developed is referred to as the surface dynamic feedback patterns [22]. This variable describes how the land surface reflectance at a location changes over time. It is done by constructing a 3D surface describing the change of reflectance across spectral bands over time at a location using remote sensing techniques (Figure 2). It has been shown that the difference in reflectance surface between two points is related to the difference in soil conditions given that other environmental factors are the same [22]. Therefore, it has been effectively used to map spatial variation of soil particle composition over flat areas [23]. Recent developments in this area were able to relate reflectance to accumulative evaporation over time [24] and to relate to rainfall magnitude in an effort to correct the pattern for large area applications [25].

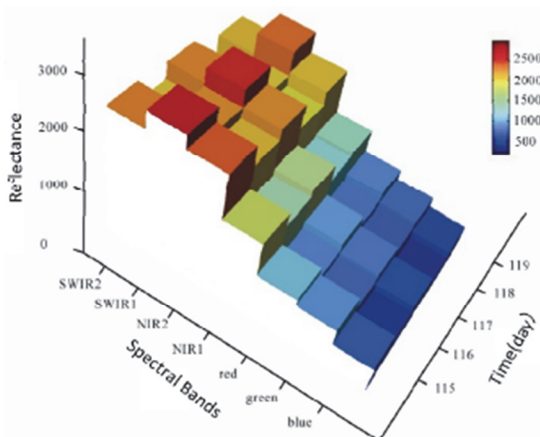


Fig. 2. Land surface dynamic feedback pattern of a location

4. SOFTWARE PLATFORMS AND OPERATIONS

The SoLIM approach has been made available to users through two distinct platforms. One is based on the desktop platform and the associated software is referred as SoLIMSolutions which is versioned by year. The other is based on the web platform which takes the advantages of the recent advancement in high performance computing and the cyber infrastructures. This platform is referred to as CyberSoLIM. Both of them are available through <https://solim.geography.wisc.edu/software>. This section provides an operational overview of these two platforms by first presenting the overall design of the platform and then by outlining the steps for conducting digital soil mapping using SoLIM under different circumstances.

4.1. SoLIMSolutions

SoLIMSolutions comes in a zip file. No special installation procedure is needed to install except unzipping the zip file into a directory where you want SoLIMSolutions to reside. The package also contains the tutorial data in the directory named Tutorial_Data as well as the online help file (SoLIMSolutions_Help.chm). There are two sets of documents to assist users to use SoLIMSolutions for soil mapping. The first, refer to it as the “Functionality manual”, is on the operation and functionality of the software which is contained in SoLIMSolutions_Help.chm and can be accessed through Help menu of the software. This manual is also available at the front webpage of SoLIMSolutions. The second document, referred to as the “Procedure manual”, is on the detail procedures of DSM using SoLIMSolutions which is only available at the front page of SoLIMSolutions and it came with its own tutorial data sets. This document and the associated tutorial data were compiled for various workshops given on SoLIM. The entrance to SoLIMSolutions is SoLIMSolutions.exe which will lead to the interfaces shown in Figure 3. A comprehensive description of the functionality through the menus system shown in Figure 3 is given in the Help system. The steps to conduct DSM using SoLIMSolutions under major scenarios are described below.

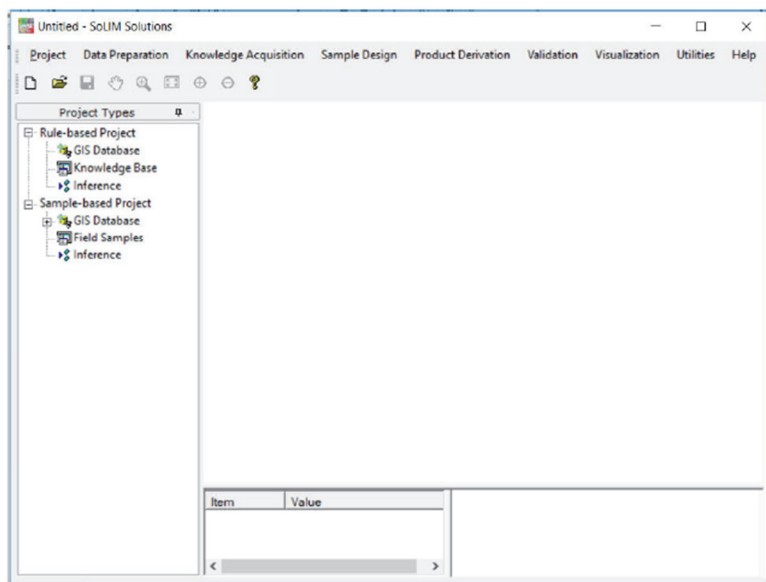


Fig. 3. Interface of SoLIMSolution 2015

4.1.1. Digital soil mapping based on field samples

Under this scenario users will use field soil samples as prototypes. These samples may not be well distributed over the area and may be limited in number, which normally cannot be used with the DSM techniques based on the First Law of Geography or the statistical principle, but these samples can be used for soil mapping under the SoLIM approach due to the fact that SoLIM is based on the Third Law of Geography which does not require samples to be of certain size nor specific spatial distribution.

Step 1: Create a sample-based project

On the main menu of SoLIMSolutions, select “Project->New” to create a new project. Specify the project to be “sample_based”.

Step 2: Add GIS data layers

Spatial data on environmental covariates are used to characterize the environmental configuration at each sample point. Therefore, spatial data on these set of covariates need to be loaded into SoLIMSolutions to characterize the configuration. In the left project panel, you will see five sub-nodes under the “GIS Database” node: “Climate Layers”, “Parent Material Layers”, “Topographic Layers”, “Vegetation Layers” and “Other Layers”. The environmental data layers can be loaded through these different sub-nodes. These sub-nodes are used to specify the hierarchy in the geographic configuration.

Step 3: Add the samples

Samples are the prototypes for digital soil mapping using the SoLIM approach. Each sample contains at least four pieces of information (Sample ID, X-coordinate, Y-Coordinate, Attribute). More than one attribute can be added for each point. The environmental configuration for each point does not need to be included in this sample point file because the environmental configuration can be easily defined once the location of the sample point is known and the spatial data on the covariates are loaded. The file containing the samples can be uploaded into SoLIMSolutions through the “Field Samples” node. It may be found that the panel on the right side will switch to a blank table correspondingly. Press the “Load Sample Point Table” button on the top to load the samples into this table.

Step 4: Run inference

Once both the spatial data on the covariates and the sample points are loaded, the environmental configuration for each sample as well as for any location in the study is constructed automatically in SoLIMSolutions. With the environmental configurations constructed, similarity in environmental configuration between each of the samples and any unsampled location can be computed. These similarities can then be used to predict the soil property value at the unsampled location by combining these similarities and the attribute values at the sample points involved using Equation 1. The “Inference” node will allow to perform the prediction.

Step 5: Result visualization

The results from the inference above can be viewed through the Visualization menu. The 2D tool is the in-house viewer in SoLIMSolutions but the 3D tool requires the installation of 3dMapper which is also available at the SoLIM software website.

Step 6: Validation

For validation, you need a set of independent validation samples which should be collected independently from the samples used as the prototypes. The validation samples are stored in a text file using the following format: SampleID, X-Coordinate, Y-Coordinate, SoilPropertyValue. The coordinate system used to define the locations of samples should be the same as that used for the spatial data as well as for the prototype samples. Validation is done through the “Property Validation” under the “Validation” menu.

4.1.2. Digital soil mapping based on knowledge from soil experts

The steps below describe the scenario when users only have local soil experts to provide the definition of the prototypes. The procedures given below are based on the assumption that users have obtained the knowledge from local experts on the prototypes. Figure 4 illustrates an example of such information. In this example, the knowledge on the prototypes of 4 soil classes is given. The environmental configuration for each of them was characterized by three environmental variables (Gradient, Elevation, and Profile curvature). The values for these environmental variables constitute the configuration. The soil A-horizon depth is the soil property.

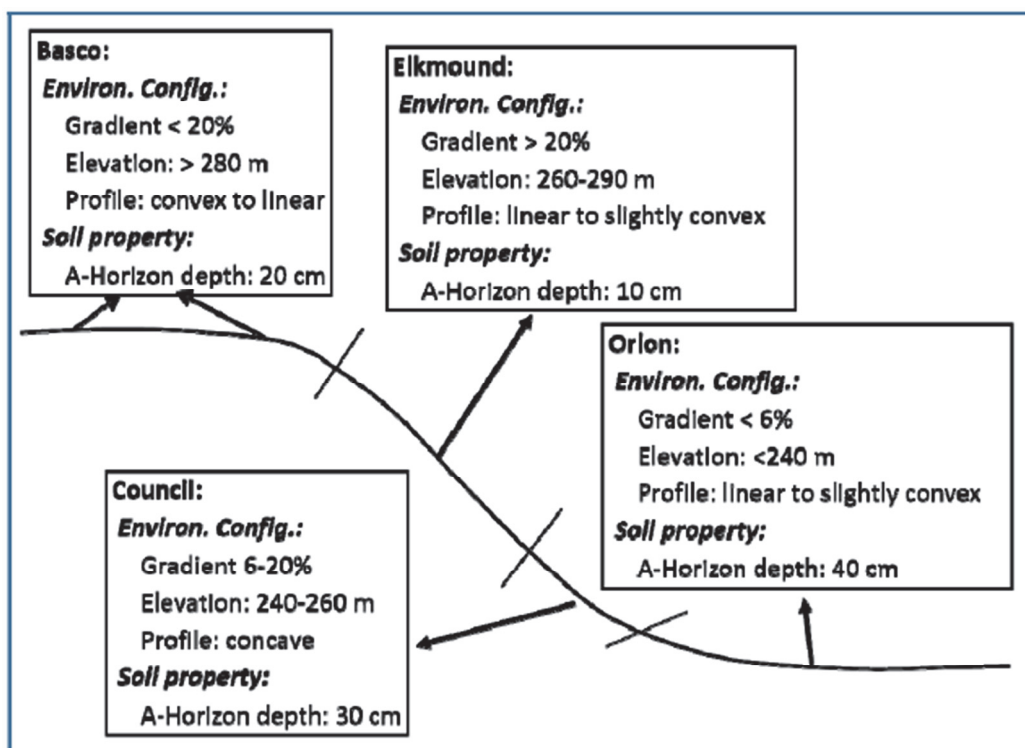


Fig. 4. Example of knowledge on the prototypes of soil classes

Below are the procedures for digital soil mapping based on this knowledge and the details of the steps are given in the Procedure manual.

Step 1: Create a rule-based project

To conduct DSM with SoLIMSolutions, you need to create a rule-based project due to the fact that the knowledge obtained on prototypes is described in the form of rules. Choose “Project -> new” on the main menu and specify the project type to be “rule-based”.

Step 2: Create a GIS database for environmental configuration

Users now add the GIS data layers which are used to describe the environmental configurations. In this example, users need to add the GIS data layers on slope gradient, elevation, and profile curvature into the project. This is done by right clicking “GIS Database” under “Rule-based project” and select “Add Layer”. Users can then add each of these GIS data layers into the project database.

Step 3: Define the prototypes for each soil class

For each class, users need to define the prototype representing this class using the knowledge extracted from local soil experts (such as these in Figure 4). This is done through **fuzzy membership curve** definition by right clicking on the “Knowledge Base” under “Rule based Project” and choose “Add Soil Type” and then for each soil type users define the typical environmental condition for this soil to develop under the given environmental variable, and the condition where the soil class will never develop and the value where the soil class will be halfway developed. Operational details of this task are provided in both the Functionality manual and the Procedure manual.

Step 4: Generate fuzzy membership maps of each soil class

Now users can compute the similarity of each location in the study area to the prototype of each soil class. The similarity is expressed as a fuzzy membership value. Fuzzy membership values to the prototype of a given soil class for all locations in the area make up a fuzzy membership map of that class. Generation of fuzzy membership maps is done through the “Inference” panel. Select the soil types to be inferred and specify where to save the result and the output format.

Step 5: Generate hardened soil map

If a soil class map is desired, users can achieve that through the “Hardened Map” hardening function under “Product Deviation” on the main menu. Add the fuzzy membership maps of the soil classes to be included in the soil class map and specify the output location. By hardening each location will be assigned a soil type to which the location has the maximum membership. Through this hardening process uncertainty maps associated with the creation of this hardened soil map are produced.

Step 6: Generate soil property map

Another product that can be derived from the fuzzy membership maps is soil property map. A look-up table that lists the typical soil property of each soil type should be prepared first. A weighted average approach as shown in Equation 1 is used to get the final soil property for each location. The “Property Map” function under the “Product Derivation” menu can be used to accomplish this task.

Step 7: Validation

Property map validation can be done using the step 6 in **4.1.1 Digital soil mapping based on field samples**. For validating the soil class map produced in Step 5 above, you also need a set of independent validation samples. The validation samples are stored in a text file in one of the predefined formats (see the Functionality Manual for details). Validation of soil class map is done through the “Type Validation” under the “Validation” menu.

4.1.3. Digital soil mapping based on knowledge from soil maps

Under this scenario, users are using knowledge from soil maps to define the prototypes for soil concepts (such as soil classes). The knowledge needed is characterized through a spatial data mining techniques [20]. Figure 5 illustrates the general process of mining knowledge for prototypes from existing soil maps. Due to the fact that the knowledge used to define prototypes are in the form of rules extracted from the soil maps, users need to set up the project as the “Rule-based Project” for this.

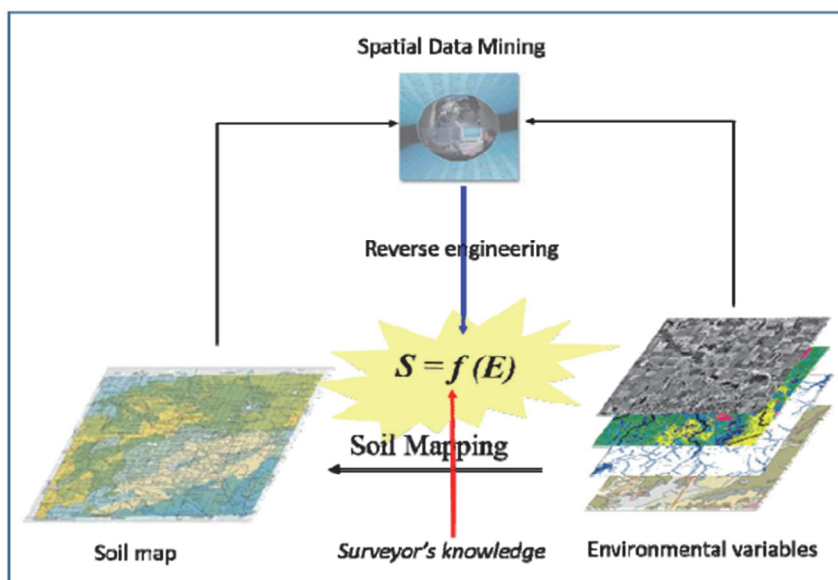


Fig. 5. Spatial data mining for knowledge defining the prototypes from existing soil maps

Step 1: Knowledge extraction from existing soil maps

The two elements needed for knowledge extraction from soil maps are: 1) a GIS database containing spatial data on environmental variables for defining the configuration; and 2) an existing soil map from which the knowledge on the prototypes of the mapped soil classes will be extracted. Once these are ready, select “Knowledge Acquisition->From Map” to start SoLIM-Knowledge Miner.

Step 2: Analysis of the extracted knowledge

The knowledge extracted from the existing maps may contain noise. The Knowledge-Miner in SoLIMsolutions allows the user(s) to increase the quality of knowledge by removing noises through knowledge analysis and editing. Knowledge analysis is

normally performed for every combination of map unit (soil class presented by polygon) and environmental data layer. Go to “Knowledge → Analyze ...” to start the knowledge analysis interface. When finished with editing, users can save the edits. Right click on the curve and choose “Save Knowledge Curve”, the curve will be saved in a .txt file.

Step 3: Knowledge import into SoLIMSolutions for soil mapping

The generated curves from step 2 can be imported into SoLIMSolutions for soil mapping. The import can be accomplished during the definition of a new rule to associate a prototype with an environmental variable (**Step 3: Define the prototypes for each soil class** described in 4.1.2 **Digital soil mapping based on knowledge from soil experts**). In order to use the knowledge extracted in step 2 above, the type of the new rules needs to be “Freehand Rule”. Click “Import From Data Mining Result” and specify the knowledge curve file (.txt file). The specified curve will be imported.

After all rules needed are added users can conduct inference soil type. The rest of steps are the same as in 4.1.2 (**Digital soil mapping based on knowledge from soil experts**).

4.2. CyberSoLIM

CyberSoLIM is another way to conduct DSM under the SoLIM framework and is a part of large framework, referred to as Easy Geographic Computing (EGC) contributed from the SoLIM group (Figure 6). It is a computing platform powered by intelligent geo-computing and high performance computing techniques. It provides a visual environment for easily constructing and executing DSM models for non-experts. The goal is to accomplish digital soil mapping tasks anywhere and anytime. CyberSoLIM provides a heuristically driven, visually assisted, high performance computing enabled cyber environment for digital soil mapping [26]. It exists in cyber space and can be accessed through the website stated above.

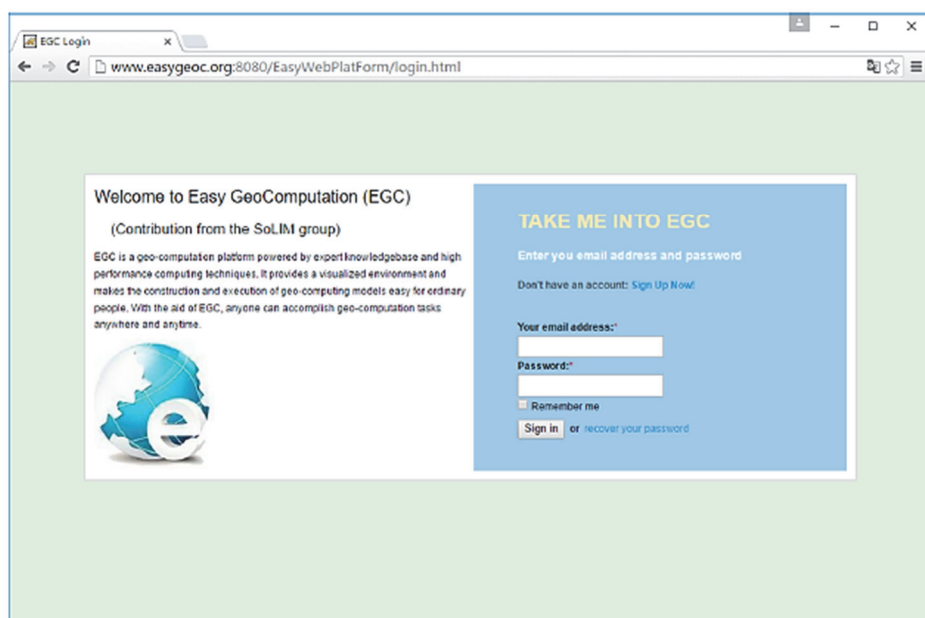


Fig. 6. CyberSoLIM through Easy Geographic Computing

At the time of this writing, CyberSoLIM is only capable of conducting DSM using the sample-based approach and is undergoing a major change in architecture and functionality. The description provided below is based on an earlier version of CyberSoLIM and includes data management, model construction, and model execution.

4.2.1. Data management

The current implementation of CyberSoLIM is for DSM under the sample-based approach using the SoLIM approach. Thus, we need both spatial data on environmental covariates and sample data. Due to the fact that it is based on cyber infrastructure, CyberSoLIM stores these data (environmental and samples) in cloud. The data management functionality of CyberSoLIM manages these data for users. Environmental data layers are required in “GeoTiff” format whose filename extension is .tif. The spatial reference (coordinate system) of all data should be consistent (the same). The samples locations should be in the same coordinate system as well and must be stored in .csv format. The easiest way to do this is to enter the field sample data into a spreadsheet and save it as a .csv file. In the table, there are at least three columns: X, Y and soil attributes and the file should contain a column heading so that it is clear which column is what.

The data a user uploaded to CyberSoLIM are under the control of the user through user account so the user can decide how the data are shared under CyberSoLIM. There are three basic modes for data sharing under CyberSoLIM. The most secure mode is that a user does not share any data with anyone. In this mode, the user is not able to access data shared by others except the data that are publically available. The next level is that a user shares the data with groups of the user’s choice. In this case the user will be able to access the data these groups share within the group. The third level is that a user share data with anyone under CyberSoLIM. In this case any data that are shared by others publically will be available to this user. The level of sharing can be assigned to individual data set.

4.2.2. Model construction

One of the key striking features of CyberSoLIM is the intelligent and automatic model building of DSM work flow. With CyberSoLIM users are presented with a map of the world. A user can navigate to a study area of interest. Right clicking on the area for DSM will bring the user an interface similar to what shown in Figure 7. Once “Digital Soil Mapping” is selected, the user will be taken to the model construction view (Figure 8) where the basic soil mapping structure is presented through the connection of three ellipses and one rectangle. The ellipse labeled “Property Map” is for the user to define the output file for the resultant soil property map and the one labeled “Sample Data” is for the user to specify the file which contains the sample data set to be used.

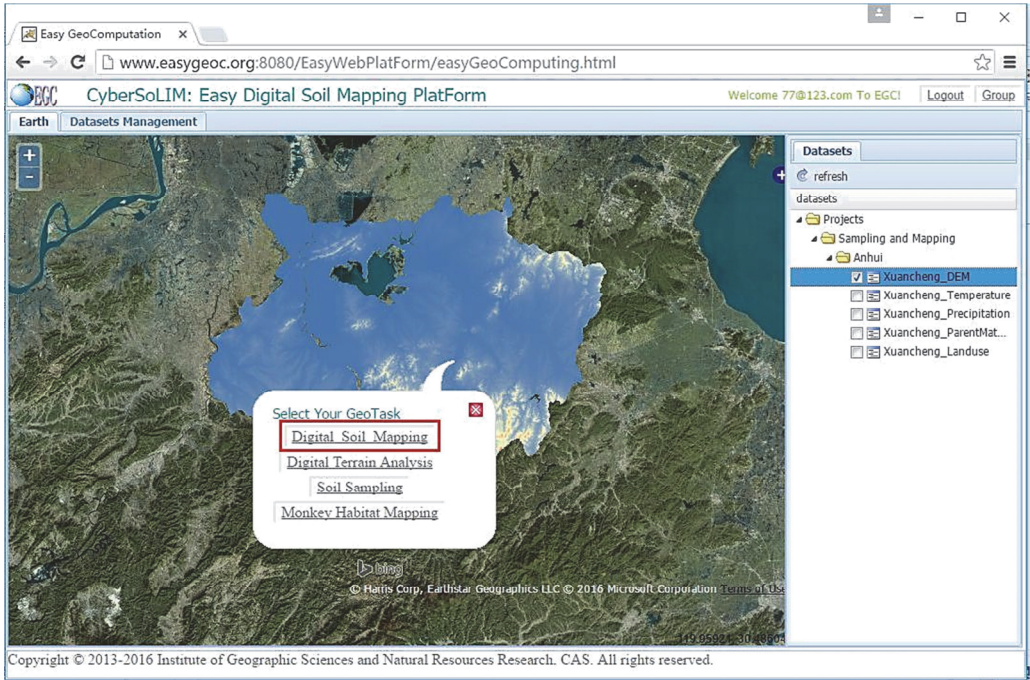


Fig. 7. Intelligent DSM model construction under CyberSolIM

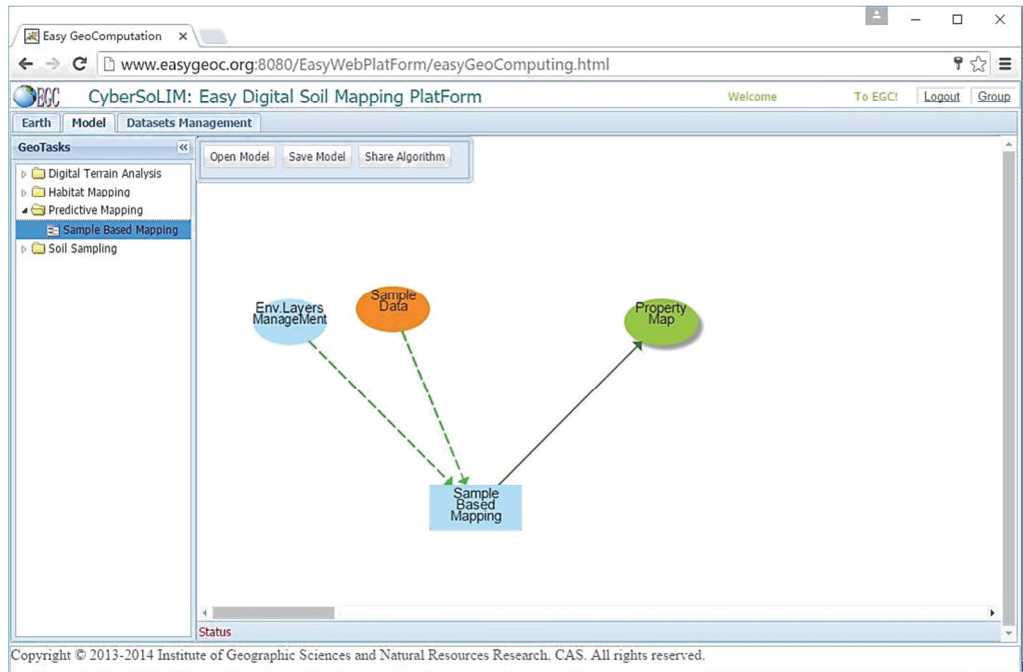


Fig. 8. The model construction view

The ellipse labeled with “Env. Layers Management” is for the user to define the environmental covariates to be used. Once the set of covariates is defined through this ellipse, the user will be presented with something as shown in Figure 9. Each of

the ellipses describes a covariate. The user can associate a data set to this covariate by right-clicking the ellipse, which will open a dialog box asking whether the user wants to provide a data set to it or the user wants CyberSoLIM to compute it. For variables such as TWI, slope gradient, profile curvature and planform curvature CyberSoLIM will be able to automatically compute them once the user specifies the gridded digital terrain model. For some of the covariates which use a common set of computing techniques, CyberSoLIM will automatically connect these techniques in a flow work (Figure 10). Once all of these covariates have been associated with a proper dataset, the model can be saved for later use.

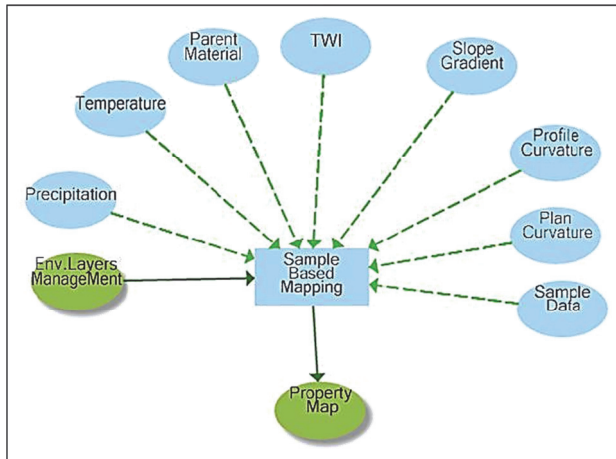


Fig. 9. Environmental covariates definition under CyberSoLIM (from [26])

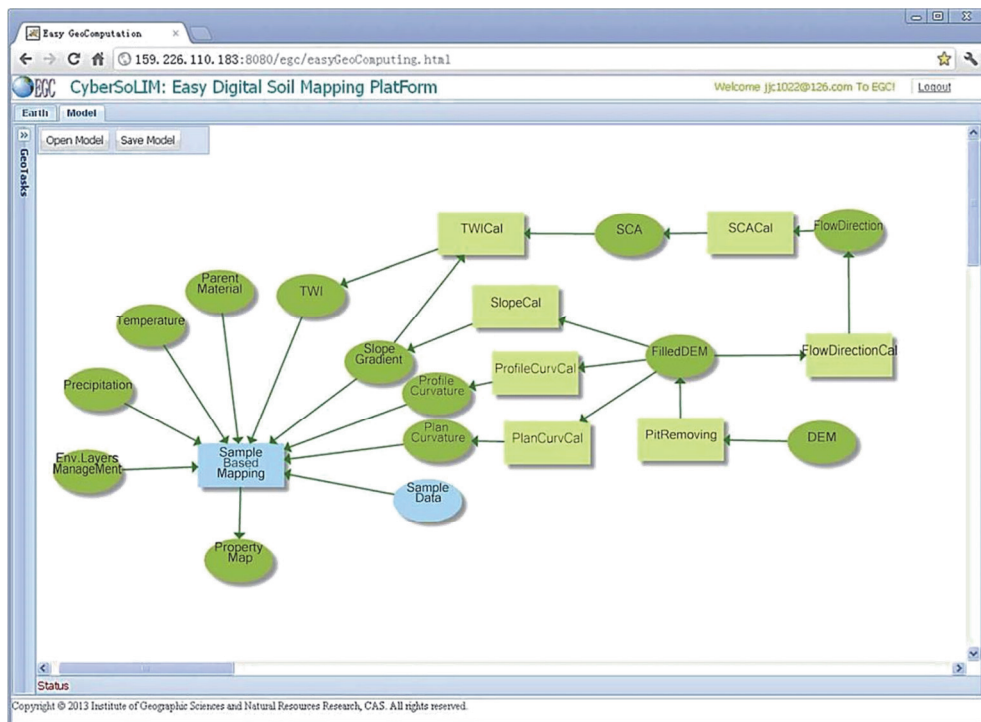


Fig. 10. A workflow model for DSM under CyberSoLIM

4.2.3. Model execution

Once all the environmental data and the soil sample data are set, the DSM model (workflow) has been constructed and is ready to be executed. The execution of the model is not done locally. In fact, the DSM model as captured in the work flow is sent to the high performance computer hosting CyberSoLIM and the work flow is then translated into executable web services and executed in the order specified in the work flow. This can be invoked by right-click on “Sample Based Mapping” and select run. The user can also click on “Operation Parameters” in “the Sample Based Mapping” box to adjust the parameters of the digital soil mapping model to customize this model. The result from the model will be presented through a web link once the execution is completed. It can be downloaded following this link and visualized in CyberSoLIM.

It is clear through this illustration, CyberSoLIM eliminates many tedious tasks for data preparation through intelligent model construction. GIS data preparation often becomes the bottle-neck for non-GIS specialists in their efforts in DSM. Under CyberSoLIM only a few source data layers (such as DEM, temperature, precipitation, remote sensing and geology) are needed, which dramatically reduce the burden on the users in data preparation. The other advantage is that computation is done using a high performance infrastructure which not only improve the speed of DSM but also removes the worry of maintaining computing hardware from users.

5. FUTURE RESEARCH ISSUES

The SoLIM approach is rather new, not only from the perspective of its methodological development but more importantly from the perspective of theoretical foundation (the application of the Third Law of Geography). Many research issues both in methodological and theoretical developments need further studying. We here highlight only a few which we think are key to the advancement of DSM under the Third Law of Geography and extend invitations to anyone who wants to collaborate to advance the research in these and other areas.

The first research issue is related to environmental configuration. As shown in Section 2.1, characterization of environmental configuration consists of three basic aspects. Aspect 1 is a comprehensive lists of environmental covariates which can effectively describe the environmental configuration for a given soil property. Although efforts are made in developing new variables [21, 24, 25], research efforts are needed in fuzzy landform characterization (such as plain, hills), dynamic vegetation growing conditions, creation of dynamic surface feedback patterns over large area.

Aspect 2 is the hierarchy of covariates for characterizing environmental configurations under the Third Law of Geography. There is little research on this. Research efforts are desired on questions such as following. What is the impact of hierarchy of covariates on environmental configuration characterization? How should the covariates be structured so that the characterization is more effective? Aspect 3 is the footprint for characterizing environmental configuration at a location for DSM [27]. Questions, such as: What is the footprint (neighborhood size) of environmental configuration for a given soil property? Is there a common footprint for all soil properties? — deserve more attention.

The second research issue is about sample verification. In SoLIM under the Third Law of Geography, the representativeness of a single sample is used. There is no doubt that this representativeness is extremely sensitive to the quality of this sample. The research question would be how to evaluate and increase the reliability of the soil samples under the Third Law of Geography and how this reliability impacts the quality of predicted soil map [28, 29].

The third research issue is knowledge extraction and integration for prototypes. Knowledge on environmental configurations about soil prototypes exist in various forms (paper maps, soil samples, survey reports etc.). Each of these forms has advantages and disadvantages as to the comprehensiveness and quality of the knowledge [30, 31]. Techniques are needed to extract knowledge from these various forms and integrate them into a holistic representation [30, 32].

The fourth research issue is the further development of CyberSoLIM. As it can be seen from above presentation, CyberSoLIM has two defining characteristics: the automatic construction of DSM work flow and the execution of the work flow using high performance computing. The automatic construction of work flow can drastically reduce the burden of users for knowledge on DSM work flow and the burden on conducting the analysis needed in the work flow. Sharing of analytical methods which can be used as web services are the bottleneck for systems like CyberSoLIM. The execution of DSM tasks on cyber platforms, particularly over platforms based on cloud infrastructure, demands new approaches to spatial data management and load management. Collaborative efforts in the deployment of cyber techniques in DSM are much needed.

6. CONCLUSIONS

This paper presents an overview of the SoLIM approach in light of the laws of geography and statistical principle used in DSM. The SoLIM approach was developed based on important geographic principle, now referred to as the Third Law of Geography. With SoLIM DSM does not require soil samples to be over certain size nor to be of specific spatial distribution. This dramatically reduces requirements for DSM.

The Third Law of Geography calls for the use of similarity in environmental (geographic) configuration between locations and uses this similarity for soil prediction. The SoLIM approach implemented the similarity of environmental configuration through the use of soil prototypes which can be defined by central concepts of soil classes or by field soil samples. This dramatically increases the sources available for defining soil prototypes, with the possibility of reducing the burden on collecting extensive new soil samples. The uncertainty derived based on similarities will also effectively target where the additional samples are needed to improve the quality of products efficiently.

The SoLIM approach are provided in two platforms: SoLIMsolutions and CyberSoLIM. SoLIMsolutions is a desktop deployment and contains more comprehensive set of functionality for DSM while CyberSoLIM is a new effort to increase the availability as well as computational efficiency by using cyber infrastructures and intelligent computing techniques. All of these make SoLIM more applicable over large and complex geographic areas, easily available to people less savvy in geospatial analysis, and more efficient in the use of resources.

Acknowledgements

This paper would not be possible to be included as part of the special issue of *RUDN Journal of Agronomy and animal industries* without the kind supports from many individuals. Two of them I want to mention here. Professor Igor Savin first made me aware of the Autumn School on Smart Agriculture to be held in Moscow of Russia between September 24 and October 26, 2018. He then forwarded my interest in serving as a faculty lecturer at this School to the organization committee. Professor Timur Shiyapov, head of the office, Department for science and innovation policy of RUDN University, on behalf of steering committee of Smart Agriculture school invited me to participate. Due to my teaching commitments as well as the delay in getting a new passport I was unfortunately unable to participate to share and to learn at this Autumn School. Professor Timur Shiyapov was kind enough to invite me to write up the materials to be included in the special issue.

The work reported here was supported by grants from National Natural Science Foundation of China (Project No.: 41431177), National Basic Research Program of China (Project No.: 2015CB954102), PAPD, and Outstanding Innovation Team in Colleges and Universities in Jiangsu Province. Supports to A-Xing Zhu through the Vilas Associate Award, the Hammel Faculty Fellow Award, and the Manasse Chair Professorship from the University of Wisconsin-Madison are greatly appreciated. The materials presented here also greatly benefitted from the projects my colleagues and I worked on with the Natural Resources Conservation Services (NRCS), US Department of Agriculture. The early supports from NRCS are much appreciated.

REFERENCES

1. McBratney AB, Santos MM, Minasny B. On digital soil mapping. *Geoderma*. 2003; 117(1—2): 3—52. Available from: doi: 10.1016/S0016-7061(03)00223-4.
2. Zhu AX, Lu G, Liu J, Qin CZ, Zhou C. Spatial prediction based on Third Law of Geography *Annals of GIS*. 2018. Available from: doi: 10.1080/19475683.2018.1534890.
3. Tobler WR. A computer movie simulating urban growth in the Detroit region, *Economic geography*. 1970; 46(sup1):234—240.
4. Goodchild MF. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*. 2004; 94(2):300—303. Available from: doi: 10.1111/j.1467-8306.2004.09402008.x.
5. Goovaerts P. *Geostatistics for natural resources evaluation*. New York: Oxford University Press; 1997.
6. Fotheringham AS, Brunson C, Charlton ME. *Geographically Weighted Regression — the Analysis of Spatially Varying Relationships*. Chichester: John Wiley & Sons; 2002.
7. Zhu AX. A similarity model for representing soil spatial information. *Geoderma*. 1997; 77(2—4): 217—242. Available from: doi: 10.1016/S0016-7061(97)00023-2
8. Zhu AX, Hudson B, Burt JE, Lubich K, Simonson D. Soil mapping using GIS, expert knowledge and fuzzy logic. *Soil Science Society of America Journal*. 2001;65(5):1463—1472. Available from: doi: 10.2136/sssaj2001.6551463x.
9. Qi F, Zhu AX, Harrower M, Burt JE. Fuzzy soil mapping based on prototype category theory. *Geoderma*. 2006; 136(3—4):774—787. Available from: doi: 10.1016/j.geoderma.2006.06.001.
10. Zhu AX, Band LE, Vertessy R, Dutton B. Derivation of soil properties using a Soil-Land Inference Model (SoLIM). *Soil Science Society of America Journal*. 1997; 61(2):523—533. Available from: doi: 10.2136/sssaj1997.03615995006100020022x
11. Zhu AX, Qi F, Moore A, Burt JE. Prediction of soil properties using fuzzy membership. *Geoderma*. 2010; 158(3—4):199—206. Available from: doi: 10.1016/j.geoderma.2010.05.001.
12. Zhu AX, Liu J, Du F, Zhang SJ, Qin CZ, Burt JE, Scholten T. Predictive soil mapping with limited sample data. *European Journal of Soil Science*. 2015; 66(3):535—547. Available from: doi: 10.1111/ejss.12244.
13. Zhu AX. Measuring uncertainty in class assignment for natural resource maps using a similarity model. *Photogrammetric Engineering and Remote Sensing*. 1997; 63(10):1195—1202.

14. Zhang SJ, Zhu AX, Liu J, Yang L, Qin CZ, An YM. An heuristic uncertainty directed field sampling design for digital soil mapping. *Geoderma*. 2016; 267:123—136. Available from: doi: 10.1016/j.geoderma.2015.12.009.
15. Li Y, Zhu AX, Shi Z, Liu J, Du F. Supplemental sampling for digital soil mapping based on prediction uncertainty from both the feature domain and the spatial domain. *Geoderma*. 2016; 284:73—84. Available from: doi: 10.1016/j.geoderma.2016.08.013
16. Zhu AX, Mackay DS. Effects of spatial detail of soil information on watershed modeling. *Journal of Hydrology*. 2001; 248(1—4):54—77. Available from: doi: 10.1016/S0022-1694(01)00390-0.
17. Quinn T, Zhu AX, Burt JE. Effects of detail soil spatial information on watershed modeling across different model scales. *International Journal of Applied Earth Observation and Geoinformation*. 2005; 7(4):324—338. Available from: doi: 10.1016/j.jag.2005.06.009.
18. Zhu AX. A personal construct-based knowledge acquisition process for natural resource mapping. *International Journal of Geographic Information Science*. 1999; 13(2):119—141. Available from: doi: 10.1080/136588199241382.
19. Qi F, Zhu AX. Knowledge discovery from soil maps using inductive learning. *International Journal of Geographical information Science*. 2003; 17(8):771—795. Available from: doi: 10.1080/13658810310001596049.
20. Cheng W, Zhu AX, Qin CZ, Qi F. Updating conventional soil maps by mining soil-environment relationships from individual soil polygons. To be published in *Journal of Integrative Agriculture*. [Preprint] 2018.
21. Qin CZ, Zhu AX, Shi X, Li BL, Pei T, Zhou CH. The quantification of spatial gradation of slope positions. *Geomorphology*. 2009; 110(3—4):152—161. Available from: doi: 10.1016/j.geomorph.2009.04.003.
22. Zhu AX, Liu F, Li BL, Pei T, Qin CZ, Liu GH, Wang YJ, Chen YN, Ma XW, Qi F, Zhou CH. Differentiation of soil conditions over low relief areas using feedback dynamic patterns. *Soil Science Society of America Journal*. 2010; 74(3):861—869. Available from: doi: 10.2136/sssaj2008.0411.
23. Liu F, Geng X, Zhu AX, Fraser W, Waddell A. Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS. *Geoderma*. 2012; 171:44—52. Available from: doi: 10.1016/j.geoderma.2011.05.007
24. Guo S, Zhu AX, Meng L, Burt JE, Du F, Liu J, Zhang G. Unification of soil feedback patterns under different evaporation conditions to improve soil differentiation over flat area. *International Journal of Applied Earth Observation and Geoinformation*. 2016; 49:126—137. Available from: doi: 10.1016/j.jag.2016.02.002.
25. Zeng C, Zhu AX, Liu F, Yang L, Rossiter DG, Liu J, Wang D. 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*. 2017; 72:297—309. Available from: doi: 10.1016/j.ecolind.2016.08.023.
26. Jiang J, Zhu AX, Qin CZ, Zhu T, Liu J, Du F, Liu J, Zhang G, An Y. CyberSoLIM: a cyber platform for digital soil mapping. *Geoderma*. 2016; 263:234—243. Available from: doi: 10.1016/j.geoderma.2015.04.018.
27. Zhu AX, An YM, Qin CZ, Yang L. Spatial footprint for digital soil mapping based on the Third Law of Geography. To be published in *Annals of GIS*. [Preprint] 2018.
28. Liu J. *Integration of samples from multiple sources for predictive mapping over large areas*. [Dissertation] University of Wisconsin — Madison, Madison, WI, USA; 2017.
29. Liu J, Zhu AX, Rossiter D, Du F, Burt JE. Reliability estimation of individual sample points in individual predictive soil mapping [Preprint] 2018.
30. Du F. *Knowledge Integration in Geospatial Predictive Mapping*. [Dissertation] University of Wisconsin — Madison, Madison, WI, USA; 2017.
31. Du F, Zhu AX, Liu J, Yang L. Predictive mapping with small field sample data using semi-supervised machine learning. To be published in *Transactions in GIS*. [Preprint] 2018.

32. Wang DS, Liu JZ, Zhu AX, Wang S, Zeng C, Ma T. Automatic extraction and structuration of soil-environment relationship information from soil survey reports. To be published in *Journal of Integrative Agriculture*. [Preprint] 2018.

INFORMATION ABOUT AUTHORS

Zhu A-Xing — Department of Geography, University of Wisconsin-Madison. E-mail: azhu@wisc.edu.

Qin Cheng-Zhi — State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences.

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

Fei Du — Department of Geography, University of Wisconsin-Madison.

For citation:

Zhu AX, Qin CZ, Liang P, Du F. Digital soil mapping for smart agriculture: the SOLIM method and software platforms. *RUDN Journal of Agronomy and Animal Industries*, 2018, 13 (4), 317—335. doi: 10.22363/2312-797X-2018-13-4-317-335.

DOI: 10.22363/2312-797X-2018-13-4-317-335

ЦИФРОВОЕ КАРТИРОВАНИЕ ПОЧВ ДЛЯ ИННОВАЦИОННОГО СЕЛЬСКОГО ХОЗЯЙСТВА: SOLIM МЕТОД И ПЛАТФОРМЫ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

A.X. Zhu^{1,2,3}, C.Z. Qin², P. Liang², F. Du³

¹Nanjing Normal University
Nanjing, 210023, China

²Institute of Geographic Sciences and Natural Resources Research
Beijing, 100101, China

³University of Wisconsin-Madison
Madison, Wisconsin, 53706, USA

azhu@wisc.edu

Ключевыми проблемами, с которыми сталкиваются многие из существующих подходов цифрового картографирования почв (DSM), являются жесткие требования к размеру проб почвы, необходимой для построения соответствующих взаимосвязей и обеспечения их устойчивости. Эти требования ограничивают применение методов DSM. В статье представлен обзор подхода SoLIM и специфика его использования через доступные программные платформы. SoLIM базируется на Третьем законе географии, который требует сравнения сходства в географическом (экологическом) положении прототипа и места без выборки, а использования этого сходства для прогнозирования значения свойства почвы в заданном месте. Реализация цифрового картографирования почв в рамках подхода SoLIM устраняет требование к размеру выборки и устойчивости взаимосвязей. Кроме того, неопределенность, рассчитанная на основе сходства, может быть использована для повышения эффективности усилий по уменьшению размера ошибок. Подход SoLIM реализован на двух платформах: SoLIM Solutions и CyberSoLIM. Теоретическое обоснование и наличие программных платформ для SoLIM открывают возможности для использования DSM в больших и сложных географических регионах.

Ключевые слова: цифровое почвенное картирование, DSM, SoLIM, Первый закон географии, Второй закон географии, Третий закон географии, пространственное прогнозирование