A concept document for a digital soil mapping (DSM) comparison project for *GlobalSoilMap.net*.

Compiled by Bob MacMillan
on behalf of the *GlobalSoilMap.net* Consortium
22/02/2010

Background

In order to move forward towards operational mapping, the *GlobalSoilmap.net* project needs to identify, apply, illustrate and document a number of potential prediction methods that can be applied under differing conditions of availability and type of data, human and computing resources and soil landscape patterns. These methods may differ in terms of how they use available evidence (soil point and map data), the type, spatial scale and number of environmental covariates they make use of and the prediction methods (statistical or otherwise) that they use to predict soil properties given available evidence and covariates.

A Proposal for a Digital Soil Mapping Comparison Project

A convenient way to rapidly identify, apply and compare several different approaches for producing predictive maps on an operational basis, given available data and resources, is to have a digital soil mapping (DSM) comparison project.

We are proposing to set up and run such a digital soil mapping comparison project using data supplied for four "proof-of-concept" study locations. Think of it as a sort of competition or "DSM Olympics".

Four responses were received to a request for test data sets to use in a GlobalSoilMap.net DSM comparison project. The responses came from the following individuals, for the following locations:

No	Location	Contact Person	Length (km)	Width (km)	Described Profiles	Analyzed Profiles	Scale of Maps
1	Africa (Milawi)	Markus Walsh	725	250	3029	1004	1:250 k
2	USA/Canada	Sharon Waltman	100	80	2860	3029	various
3	Queensland Au	David Jacquier	340	200	2425	1081	1:250 k
4	Hungary, EU	Endre Dobos	100	100			1:100 k

Proposed Sequence of Activities for a Digital Soil Mapping Comparison Project

1. Identify, gather and collate all relevant spatial data pertaining to soil legacy data (evidence) for each of the 4 proof of concept locations in both point and map form.
    a. Massage the raw soils data to convert it into formats that are easy to assess and use (e.g. Shape files, DBF tables, Excel tables)
    b. Re-project all data into an agreed-upon equal area projection (Albers, UTM?).
    c. Apply the spline function of McBratney et al., (2009) to all horizon data to standardize soil profile data to the 6 depth intervals specified for the project.
    d. Store all point profile data in simple to read text, DBF or Excel tables.
    e. Store polygonal map data in shape file format with attached attribute tables.
2. Identify, gather and collate all relevant spatial data for all main environmental covariates (predictors).
    a. Obtain SRTM DEM data at 3 arc-seconds (90 m) and also at 1 arc-second (30 m) if available.
    b. Process the SRTM data to reduce artefacts and remove noise and vegetation?
    c. Process SRTM data (90 & 30 m) to compute complete list of terrain derivatives.
        i. To permit investigation of scale influences, compute terrain derivatives using grid resolutions and search windows of different size and extent.
    d. Obtain as many other layers of covariate map data as feasible (climate, vegetation, land use, geology, parent material, landforms, wetlands, time).
    e. Obtain satellite image data and process to extract information on land use, parent material texture and type, wetlands and organic areas, surface salts or crusts, bare rock, erosion and so on. Investigate novel ideas for using imagery.
3. Set up and populate a web site to support interactive collaboration and controlled access to the data and programs needed to contribute to the pilot activities.
    a. Set up a web platform that requires interested individuals to register and be approved in order to participate and to access the data.
    b. Post all assembled data on a web site available for download by all interested and registered participants or contributors.
4. Invite any and all interested and capable individuals or organizations to apply to participate in the proof of concept project by:
    a. Distributing a description of the project and an invitation to contribute to it to as wide a list of potential participants as possible using mailing lists compiled by digitalsoilmapping.org and pedometrics.org.
    b. Include these rules and expectations for participation in the project with the distribution and also post them on the project web site.
    c. Request potential participants to register to participate in the project and approve participation of all volunteers with demonstrated capabilities in DSM.
    d. Set a hard final date for registering to participate in the project.
5. Request that approved participants contribute to the project by:

    a. Confirming their intention to participate and indicating which study areas, soil properties and depth intervals they are willing to prepare predictions for.
    b. Agreeing to submit any additional useful data sets of environmental covariates for use by others that they may identify or prepare.
    c. Agreeing to apply one or more prediction methods of their choosing to the assembled covariate data to predict at least two of the soil properties that the project is proposing to deliver (e.g. soil carbon, pH, clay) for all 6 depths.
    d. Agreeing to release their results and the associated description of methods for use and distribution by the project.
6. Develop a method for objective comparison of the relative strengths and weaknesses of each method and of the results produced by application of each method.
    a. This assessment methodology ought to be based on using an independent data set not used in the construction or application of any of the models being evaluated.
    b. It ought to focus on assessing the accuracy of the predictions of soil properties made by the models at locations for which training data sets were not available.
7. Apply the agreed upon objective comparison methods to each of the output products produced by each of the contributors.
    a. Complete an objective evaluation and comparison of the relative strengths and limitations of each of the applied methods.
8. Have each contributor provide full documentation of the methods used to produce any given set of predictions.
    a. This documentation will include the type and amount of evidence used (legacy data), the number, type, resolution and scale of the environmental covariates used to make the predictions and the statistical or empirical models applied to the data to produce the predictions.
9. Present documentation of as many of the methods as it is feasible and reasonable to report on in a special report on side by side comparison of different methods of digital soil mapping all applied to the same area.
    a. The report, book or special journal issue will reward contributors who have been most successful in prediction of soil properties.
    b. It will include a range of approaches that make use of differing amounts and types of soil evidence data and environmental covariate data, as well as significantly different prediction methods.
    c. A concluding chapter for this document will compare and contrast the various reported methods in terms of relative accuracy, data and processing requirements, time and effort to apply and overall cost.
10. Use the data and the documentation assembled for the comparison project to produce tutorials and manuals that document and illustrate application of each of the main methods of potential interest.

   a. The assembled documentation will be the initial "cookbook" of prediction methods available for use in production mapping in any given region.

11. Prepare training courses in application of any of the methods documented by the comparison project and offer these to any node that elects to use a particular method to produce predictions for any or all of its area of responsibility.
    a. These courses would best be designed and led by the individuals who developed and demonstrated the method selected for use in any given area.

Rules for participation in the Digital Soil Mapping Comparison Project

1. The project cannot offer any funding or payment for work done to apply methods and make predictions. There is unfortunately no budget for this at the moment.
2. Individuals approved for participation in the project must have demonstrated proficiency in digital soil mapping through completion of previous DSM projects or publication of refereed papers describing DSM methods and applications.
3. Participants must agree to deliver outputs that conform to the agreed upon specifications for GlobalSoilMap.net outputs (90 m grid, 6 depths, 2 of 6 properties).
4. Participants must be prepared to demonstrate significant progress in applying their methods and producing interim results by Aug 1, 2010.
5. Participants must be prepared to aim for final completion of their work and submission of their predicted maps by Dec 1, 2010.
6. There will be four main categories of approaches that are encouraged to be applied and evaluated. These categories are classified as:
    a. Methods that use only point data sets as evidence to make predictions.
    b. Methods that use only soil maps as evidence to make predictions.
    c. Methods that use both soil maps and point data to make predictions.
    d. Methods that require use of very limited or no map or point data to make predictions.
7. The compilation document will aim to select and include examples of the 2 most successful prediction methods in each of the above 4 categories.
    a. Other examples of successful methods may also be included, depending upon the total number of completed projects and the success of other methods.
8. The project will award a certificate of merit for the method that achieves the highest overall prediction accuracy in each of the four pilot areas.
    a. Accuracy will be determined by an independent sample of data points not used in the construction of any map whose locations and values will not be known in advance to any participant.
    b. A cash prize of Euro 100 will be awarded to the individual who receives the certificate in each of the 4 pilot areas.
9. All active project work must be completed by Dec 1, 2010 and all predictive maps submitted by then.