

Re: Some Canadian Soil Profile Data

From: Robert MacMillan <bobmacm@gmail.com>
To: Tomislav Hengl <tom.hengl@isric.org>
Cc: "Shaw, Cindy (NRCan/RNCan)" <cindy.shaw@canada.ca>, Tomislav Hengl <tom.hengl@gmail.com>

Priority: Normal
Date: 02/01/2016 19:56

Tom and Cindy,

Tom, here is a copy of the Forest Ecosystem Carbon Database (FECD) that Cindy Shaw assembled and has made available to us.

I linked her 3 tables together to put all location and profile data in the same table rather than related tables. Don't know if this is a help or a hindrance.

I will send you a copy of Cindy's FECD Report as a PDF later when I have better WiFi access. It will provide documentation you may need.

Again, I do not know if this data set duplicates profiles that you already got from Ottawa. But I send them to you just in case these are not profiles you already have.

These are the easiest of the data sets that Cindy shared with me to get to you right away. I will keep in touch with Cindy in the new year as she moves ahead with compiling, cleaning and assembling a larger collection of profile data that she has already scanned in from many provincial soil survey reports and other publications but still needs to work on to clean up. No point in me doing anything more here till we see what she does. If I do anything, it will be to help Cindy in some way that she suggests.

I just wanted to make a start at familiarizing myself with the work that Cindy has done and the soil profile data that she has assembled. From what I can see, this is very good work that could provide us with a lot of soil profiles that you are not likely to have for Canada. I hope so. That would represent a start at improving the profile data that you have to use for predicting within Canada.

Let me know if this is a help or just a distraction.

Bob

On Tue, Dec 29, 2015 at 11:54 PM, Tomislav Hengl <tom.hengl@isric.org> wrote:

Hi Cindy, Bob,

If I may join this discussion, you might never have a perfectly 'clean' and 'harmonized' and complete data set. However, what helps people using your data is that you:

1. provide the right disclaimer / warning,
2. start versioning i.e. report what has changed from version to version (I version things on Github and R-forge for years),

If you have <5% of point with unharmonized/unchecked values than this will not make much difference for the analysis. What makes difference for the analysis are serious omissions and 'blunders', for example I usually go mad when somebody maintaining the data decides to use 0 for missing values (if it is undocumented it can completely mess up the results of analysis) or similar.

Google spreadsheets or Google fusion tables allow for both versioning and crowd-sourcing the data collection. Here is an example:

https://docs.google.com/spreadsheets/d/1GaNpiH65yiuHusNVkUrKog2FCiVUO6kz_wNdIzHbdfg/edit#gid=1593210234

there are 10+ people on this doc continuously entering new info and commenting 'suspicious' things.

T. (Tom) Hengl
Researcher @ ISRIC - World Soil Information
Team member Africa Soil Information Services <http://africasoils.net>
Url: <http://www.wageningenur.nl/en/Persons/dr.-T-Tom-Hengl.htm>
Network: <http://profiles.google.com/tom.hengl>
Publications: <http://scholar.google.com/citations?user=2oYU7S8AAAAJ>

On 29-12-2015 17:19, Robert MacMillan wrote:

Hi Cindy,

Thanks for that caution.

Today I started reading your FEDC report. Good work ahead of its time.
Gives a good description of problems and solutions in assembling legacy data.

For all the millions that were spent we did a terrible job of consistently reporting any kind of minimum data set.

There is a real argument in this past data in devising and implementing some kind of consistent framework for future field sampling.

That is one of the things AfSIS is trying to promote. But everyone always wants to be unique and different.

I have some spare time on my hands here if you want to point me towards particular files in your collection that it would be worth my time to try to finish compiling.

I don't go out after dark here and it gets dark at 7 precisely.

Bob

On 29 Dec 2015 18:25, "Shaw, Cindy (NRCan/RNCan)" <cindy.shaw@canada.ca<mailto:cindy.shaw@canada.ca>>>
wrote:

Hi Bob,

I try, repeatedly, to warn people about the utility of the Siltanen dataset. When I compiled data for our use I did a bit of QAQC on Siltanen but it really needs more. For, example, if you quickly calculate C:N you can pick off some of the weird ones. There are records where mineral N instead of total N, was entered for total N. There are profiles that only include the top 10 cm, and in going back to the original data sources we have found data for the LFH that were not included in Siltanen. The biggest issue is that for a large number of the Siltanen records the lat and long are guesses, because, back in 1997 no one was thinking of using geolocated data for the purposes that it is used for today. The only goal was to produce a map to show the distribution of the plots. So, I would use the data with caution. Some of it is good and some of it is not.

Cheers.

Cindy

From: Robert MacMillan [mailto:bobmacm@gmail.com<mailto:bobmacm@gmail.com>]
Sent: December-28-15 2:52 AM
To: Tomislav Hengl; Tomislav Hengl; Shaw, Cindy (NRCan/RNCan)
Subject: Some Canadian Soil Profile Data

Tom,

Today I finally started looking at some of the Canadian soil profile data that Cindy Shaw sent my way last year.

One of the things I noticed right away is that she had collected a couple of data sets that had been collated by other people. These data sets were already pretty clean and pretty complete. I did a little work to join and organize one of the collated data sets this morning. It comes from a report by Siltanen et al, 1997 that I attach.

I am sending you a simplified and joined version of this database that has Site_No, Lat, Long and all of the reported soil properties.

My question is about whether you already have this profile data from the data you got from CanSIS or do these represent profiles for Canada that you do not already have?

You would need to know that the soil profile data in this set specifically exclude ALL organic soils in Canada's forested area. This is therefore a VERY skewed data set that will NOT represent the properties of the extensive organic soils in Canada. It could adversely affect your predictions by being non-representative of organics. Just keep that in mind.

There is 1 more collated data set in the data Cindy sent me. Then there is her work to scan and collate data from all the unique, individual soil survey reports she could find. Some of the data in the collations may have come from these same reports, but not all.

I only brought an old Laptop with an old version of Excel here to Africa so I can't open a lot of Cindy's Excel tables that are in xlsx format. But I could open this Siltanen data set as it was in old XLS format.

Anyway, I just want to know if you already have these data and, if not, if you think they would be useful to have for global mapping.

Bob

--

R. A. (Bob) MacMillan Ph.D.
LandMapper Environmental Solutions Inc.
7415 118 A Street NW,
Edmonton, AB, Canada T6G 1V4

--

R. A. (Bob) MacMillan Ph.D.
LandMapper Environmental Solutions Inc.
7415 118 A Street NW,
Edmonton, AB, Canada T6G 1V4

FECD_JOIN_ALL.DBF	Content-Type: application/octet-stream Size: 2.15 MB
--------------------------	---