From: Beaudette, Dylan - NRCS, Sonora, CA
To: Herrick, Jeff; Nauman, Travis; Fan, Zhaosheng - ARS
Cc: Colby Brungard (cbrung@nmsu.edu); Monger, Curtis - NRCS, Lincoln, NE; Levi, Matthew - ARS; Salley, Shawn - ARS; Wills, Skye - NRCS, Lincoln, NE; Toby O"Geen (atogeen@ucdavis.edu)
Subject: RE: LTAG-Soil Question (soil color depth - thoughts by Friday)
Date: Thursday, October 12, 2017 1:24:00 PM
Attachments: image001.png
image002.png
image003.png
image004.png
image005.png

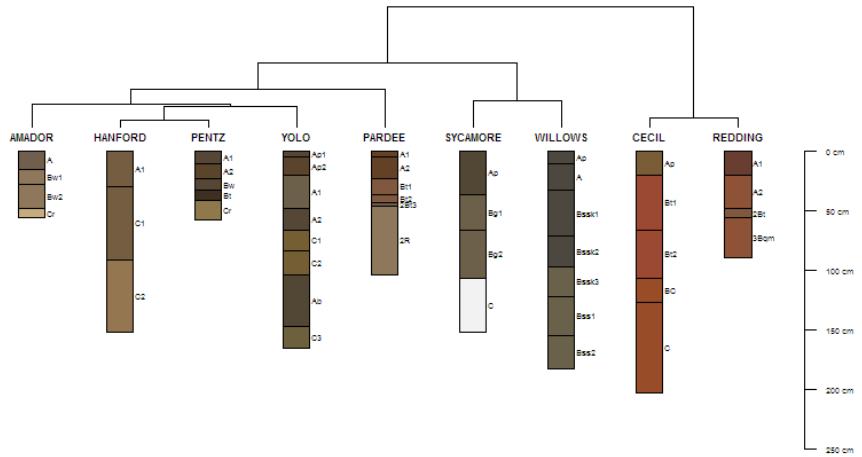
Correct me if I am wrong here: the plan is to use soil texture and rock fragments from multiple depths--why not use soil colors from the same depths? Using a single color from a single depth seems like an overly simplistic approach.

Another idea, inspired by a project that has been sitting on my back-burner for a while now: compare soil color "signatures" derived from reference pedon / components and user-supplied data.

What in the world is a color "signature"? Here are three possible answers, evaluated at the pedon / profile scale or over a depth interval

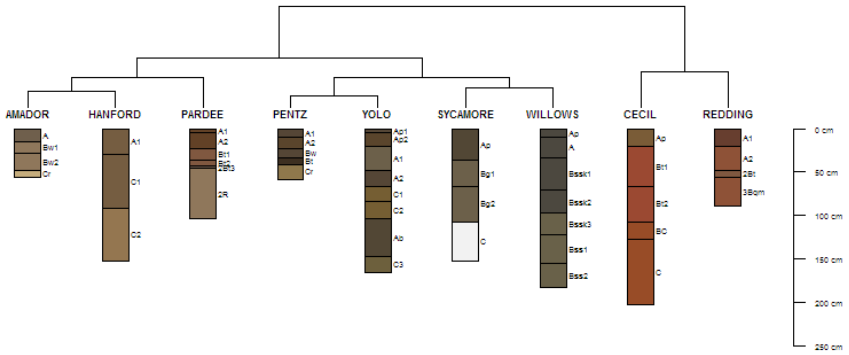
1. Compute proportions of white, red, green, yellow, and blue "pigments" based on the CIE LAB representation of soil colors distributed over horizons, weighted by horizon thickness. Pigment proportions are used so that shallow soils can be compared with deep soils. Each row is an observation, columns describe the multivariate soil color signature. Looks like this:

	id	.white.pigment	.red.pigment	.green.pigment	.yellow.pigment	.blue.pigment
1	AMADOR	0.3116786	0.12448336	0.000000000	0.5638381	0
2	CECIL	0.1199565	0.41292321	0.000000000	0.4671203	0
3	HANFORD	0.2441100	0.16184749	0.000000000	0.5940426	0
4	PARDEE	0.2446076	0.21447890	0.000000000	0.5409135	0
5	PENTZ	0.2446259	0.13943074	0.000000000	0.6159434	0
6	REDDING	0.1505148	0.39115739	0.000000000	0.4583278	0
7	SYCAMORE	0.3294305	0.07082285	0.000000000	0.5997467	0
8	WILLOWS	0.4115207	0.00000000	0.015388353	0.5730910	0
9	YOLO	0.2601527	0.09391880	0.001484137	0.6444444	0



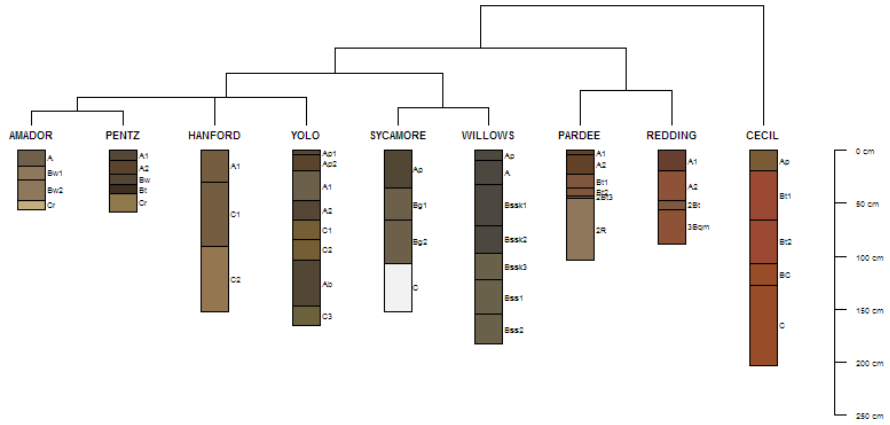
2. Extract CIE LAB coordinates at the 10th, 50th, and 90th percentiles of horizon mid-points. Each row is an observation, columns describe the multivariate soil color signature. Looks like this:

	id	A.0.1	A.0.5	A.0.9	B.0.1	B.0.5	B.0.9	L.0.1	L.0.5	L.0.9
1	AMADOR	3.68130123	4.87026202	1.9194537	13.297617	18.955628	27.82850	8.249710	10.324247	14.332776
2	CECIL	7.37778624	32.89502772	28.7224735	25.999460	29.496946	35.80688	8.254133	8.251177	8.253607
3	HANFORD	5.62934401	5.62934401	6.6108472	19.846688	19.846688	25.48043	8.252184	8.252184	10.326592
4	PARDEE	10.73024291	13.45285445	4.8702620	21.982404	20.845839	18.95563	6.160403	8.251421	10.324247
5	PENTZ	3.91234236	3.91234236	2.9939157	12.210272	12.210272	27.57974	6.157712	6.157712	10.327858
6	REDDING	16.32511816	22.38741697	22.3874170	16.933367	26.284949	26.28495	6.157777	8.251937	8.251937
7	SYCAMORE	1.75737124	1.55395330	1.5539533	12.868609	14.180105	14.18010	6.158297	8.250401	8.250401
8	WILLOWS	-0.03007361	-0.03007361	-0.5342186	5.832693	5.832693	14.57982	6.155210	6.155210	8.250861
9	YOLO	3.91234236	3.91234236	-0.3778239	12.210272	12.210272	21.99316	6.157712	6.157712	8.253823



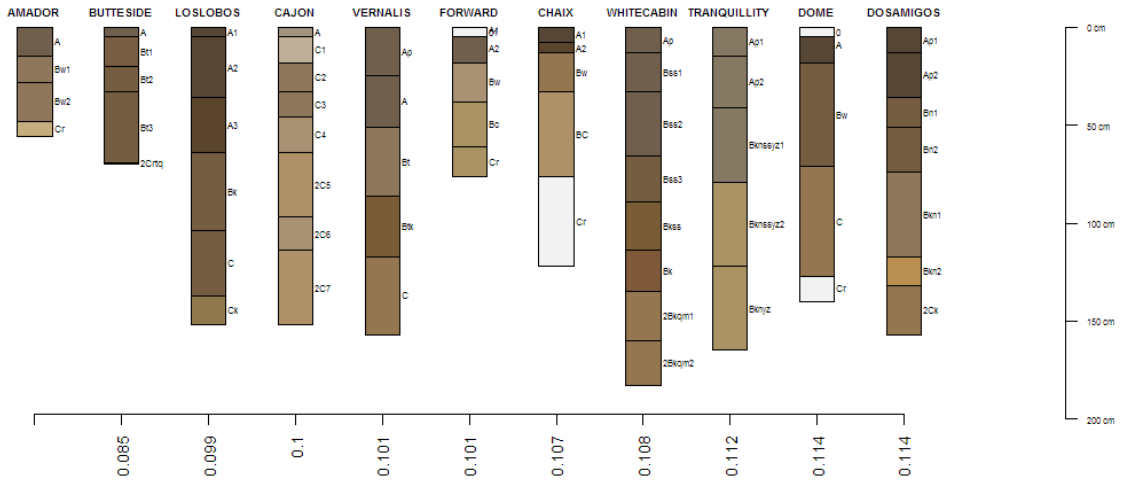
3. Select *k* medoids (PAM algorithm) from all possible colors within the profile or depth interval and use those CIE LAB coordinates as the signature. Each row is an observation, columns describe the multivariate soil color signature. Looks like this:

```
       id         A.1        A.2        A.3       B.1       B.2       B.3      L.1       L.2       L.3
1   AMADOR  3.68130123  4.8702620  1.9194537 13.297617 18.95563 27.82850 8.249710 10.324247 14.332776
2    CECIL  7.37778624 32.8950277 28.7224735 25.999460 29.49695 35.80688 8.254133  8.251177  8.253607
3 HANFORD  5.62934401  6.6108472  6.6108472 19.846688 25.48043 25.48043 8.252184 10.326592 10.326592
4  PARDEE 10.73024291 13.4528544  4.8702620 21.982404 20.84584 18.95563 6.160403  8.251421 10.324247
5   PENTZ  3.91234236  5.6389267  2.9939157 12.210272 18.06766 27.57974 6.157712  6.159809 10.327858
6 REDDING 16.32511816 22.3874170 13.4528544 16.933367 26.28495 20.84584 6.157777  8.251937  8.251421
7 SYCAMORE  1.75737124  1.5539533  1.5539533 12.868609 14.18010 14.18010 6.158297  8.250401  8.250401
8 WILLOWS -0.03007361 -0.5342186 -0.5342186  5.832693 14.57982 14.57982 6.155210  8.250861  8.250861
9     YOLO  1.75737124  3.5979795 -0.3778239 12.868609 27.76177 21.99316 6.158297  8.255234  8.253823
```



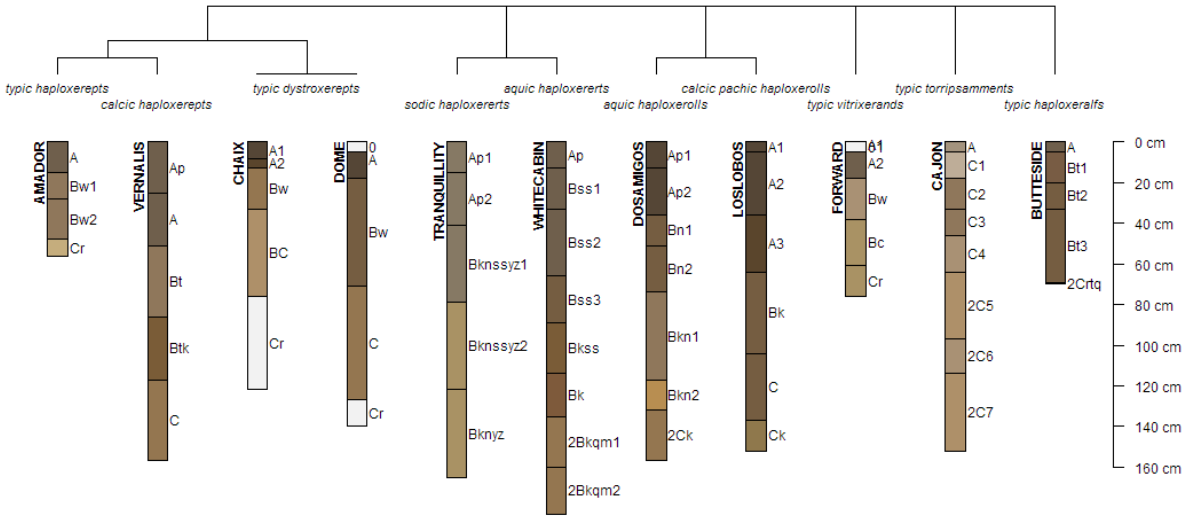
These methods require reasonable constraints (e.g. MLRA or the like) since pair-wise comparisons can quickly explode into a massive distance matrix. Computing pair-wise distances for the 24,000+ US soil series, constrained by MLRA, takes only 15 minutes. I am sure that it can be further optimized.

The results are interesting, but lack precision. For example, take the Amador soil series and the 10-closest matches (constrained to MLRA where Amador is observed) using soil color signatures (method #2 from above):



The lower axis is distance relative to Amador--note the tiny range in distances = all quite "similar" to the color signature of Amador.

When compared to subgroup level taxonomic labels, the color-based "similarities" cover a rather large range:



So what am I suggesting here? Well, *color-informed* matching is likely far more effective than *color-based* matching. A representative color signature coupled with physical / chemical properties might be an effective matching strategy--as long as meaningful constraints are in place.

The code behind all of this is in the aqp package for R (development version), available here:

http://ncss-tech.github.io/AQP/

I'll convert this into a proper document and post on the same website; been meaning to for over a year now.

A slightly different approach to describing soil color in aggregate can be found here:

http://ncss-tech.github.io/AQP/sharpshootR/aggregate-soil-color.html

Dylan

From: Herrick, Jeff
Sent: Thursday, October 12, 2017 8:22 AM
To: Nauman, Travis <tnauman@usgs.gov>; Fan, Zhaosheng - ARS <Zhaosheng.Fan@ARS.USDA.GOV>
Cc: Colby Brungard (cbrung@nmsu.edu) <cbrung@nmsu.edu>; Monger, Curtis - NRCS, Lincoln, NE <Curtis.Monger@lin.usda.gov>; Beaudette, Dylan - NRCS, Sonora, CA <Dylan.Beaudette@ca.usda.gov>; Levi, Matthew - ARS <Matthew.Levi@ARS.USDA.GOV>; Salley, Shawn - ARS <Shawn.Salley@ARS.USDA.GOV>; Wills, Skye - NRCS, Lincoln, NE <skye.wills@lin.usda.gov>; Toby O'Geen (atogeen@ucdavis.edu) <atogeen@ucdavis.edu>
Subject: LTAG-Soil Question (soil color depth - thoughts by Friday)

LTAG-Soil (LandPKS Technical Advisory Group – Soil),

Fan is working on developing a color signature for each HWSD Unit based on the Africa Soil Pedon Database and WISE. For now this will be global, for 1 depth, and not take into account change in soil depth, but these could be integrated in future.

Question: what depth? Fewer records available for deeper, but too close to surface and SOM becomes a problem.
We're thinking 35cm, which happens to be the midpoint of our 20-50cm LandPKS depth.

We would use any horizon/depth entry in the pedon database that includes the 35cm depth, excluding any where the horizon break (upper or lower) is between 32 and 38cm (i.e. upper break <32; lower break >38).

Thoughts on 35cm?
Thoughts on the horizon break rule?

Fan is ready to start this analysis, so any thoughts by Friday would be great.

Thanks,

Jeff and Fan

From: Nauman, Travis [mailto:tnauman@usgs.gov]
Sent: Wednesday, October 11, 2017 1:36 PM
To: Fan, Zhaosheng - ARS <Zhaosheng.Fan@ARS.USDA.GOV>
Cc: Colby Brungard (cbrung@nmsu.edu) <cbrung@nmsu.edu>; Monger, Curtis - NRCS, Lincoln, NE <Curtis.Monger@lin.usda.gov>; Beaudette, Dylan - NRCS, Sonora, CA <Dylan.Beaudette@ca.usda.gov>; Herrick, Jeff <Jeff.Herrick@ARS.USDA.GOV>; Levi, Matthew - ARS <Matthew.Levi@ARS.USDA.GOV>; Salley, Shawn - ARS <Shawn.Salley@ARS.USDA.GOV>; Wills, Skye - NRCS, Lincoln, NE <skye.wills@lin.usda.gov>; Toby O'Geen (atogeen@ucdavis.edu) <atogeen@ucdavis.edu>
Subject: Re: one-hour block for conference call

I like Wednesdays at 11 Mtn time. Maybe you could create a google drive folder or document that keeps track of the conversations we have and how that is changing the LandPKS direction. Then when we can't join, we could catch up quickly on that document and respond if compelled?

Best
Travis

On Wed, Oct 11, 2017 at 12:46 PM, Fan, Zhaosheng - ARS <Zhaosheng.Fan@ars.usda.gov> wrote:

> Hi All,
>
> Could we set up a regular one-hour block for our weekly conference call, so we do not have to do the doodles every week? Do you think that 11-12 on Wednesday will work (or we can do a doodle survey to determine the one-hour block)? We can always change the time slot if someone won't be able to make it, or skip a week or two if there is not much to discuss.
>
> Take care,
> Fan
>
>