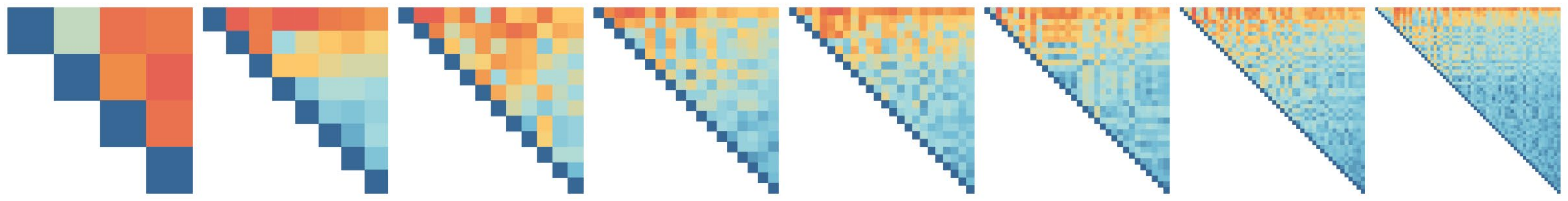
Numerical Classification of Soil Profiles

Dylan Beaudette, Shawn Salley, Jon Maynard | USDA-NRCS Soil and Plant Science Division



Rationale

Soil Taxonomy has served a critical role in the U.S. Soil Survey program for over 40 years: guiding the design of map unit and component concepts, shaping soil interpretations, and as a unified framework for communicating about the soil resource. Major changes to both Soil Survey and Soil Taxonomy over the last 40 years represent a coevolution driven by changing expectations of the land and understanding of the soil resource. However, there are many tasks in routine initial and update soil mapping that require classification *below* the family level. Numerical classification, or classification based on pair-wise distances, offers several possible solutions to organizing site and pedon data at several possible levels of generalization.

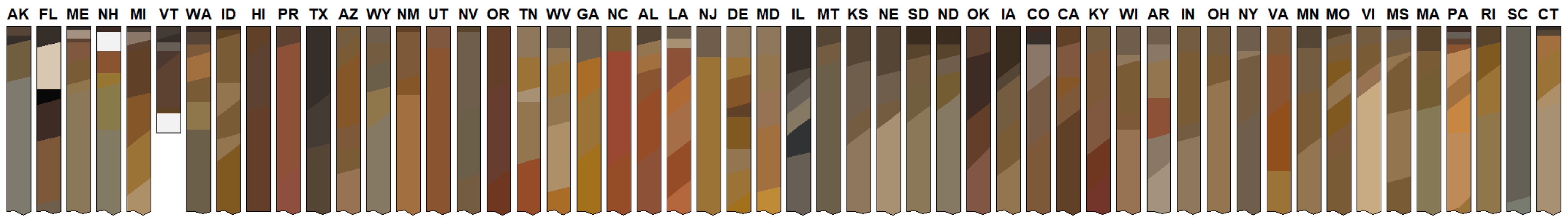


Figure 1: US state and territory soil series arranged according to subgroup level classification, as specified in the 12th edition of the "Keys to Soil Taxonomy". Regional trends in horizon thickness and moist colors reflect the impact of soil moisture regime and soil morphology on our classification system.

Numerical Classification

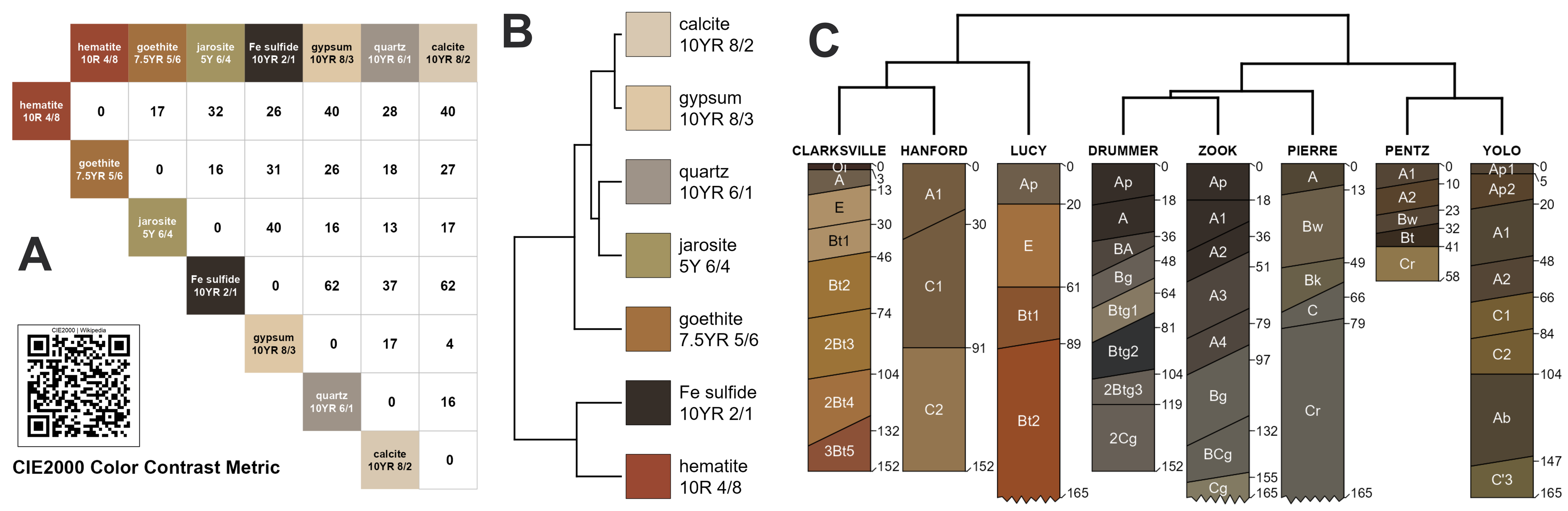


Figure 2: A demonstration of numerical classification applied to colors of common soil pigments: distance matrix (A), dendrogram representation after hierarchical cluster analysis (B), and "soil color signatures" of select soil series (C) via clustering of CIE2000 color contrast metric, evaluated over the 4 most distinct colors. Larger distances (A) represent greater perceptual contrast in color, signified by higher (left, top) branching in the dendrogram (B, C).

Numerical Classification of Soil Profiles (NCSP)

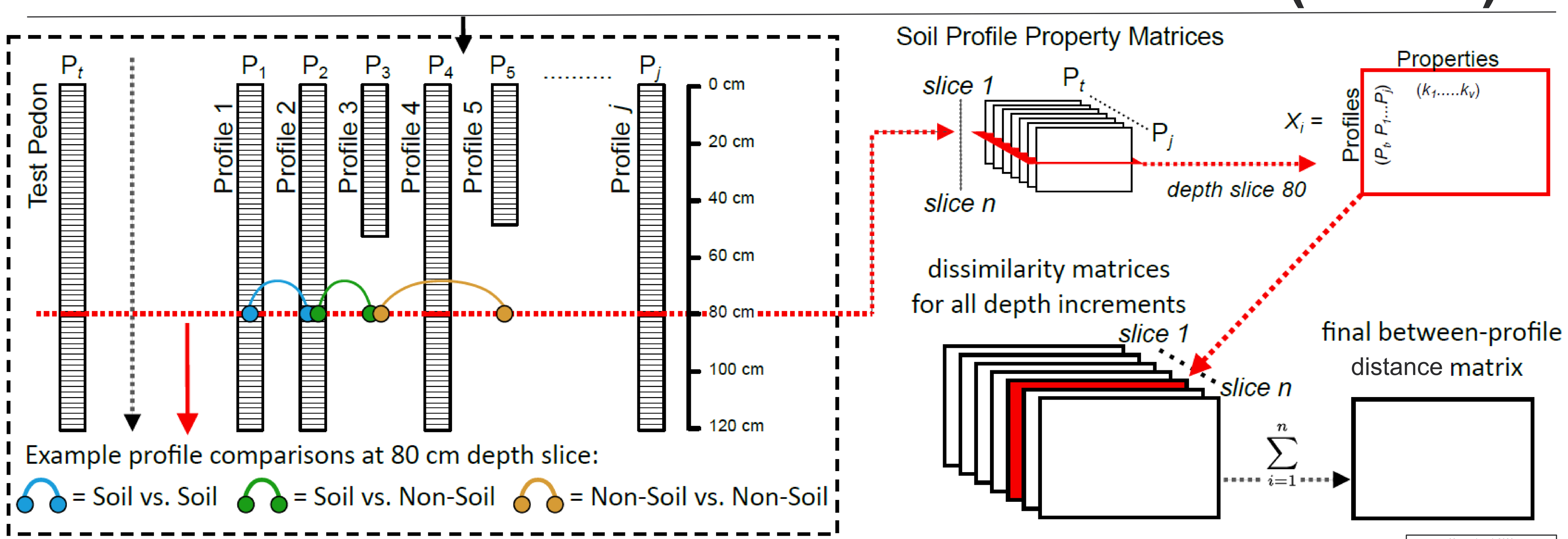
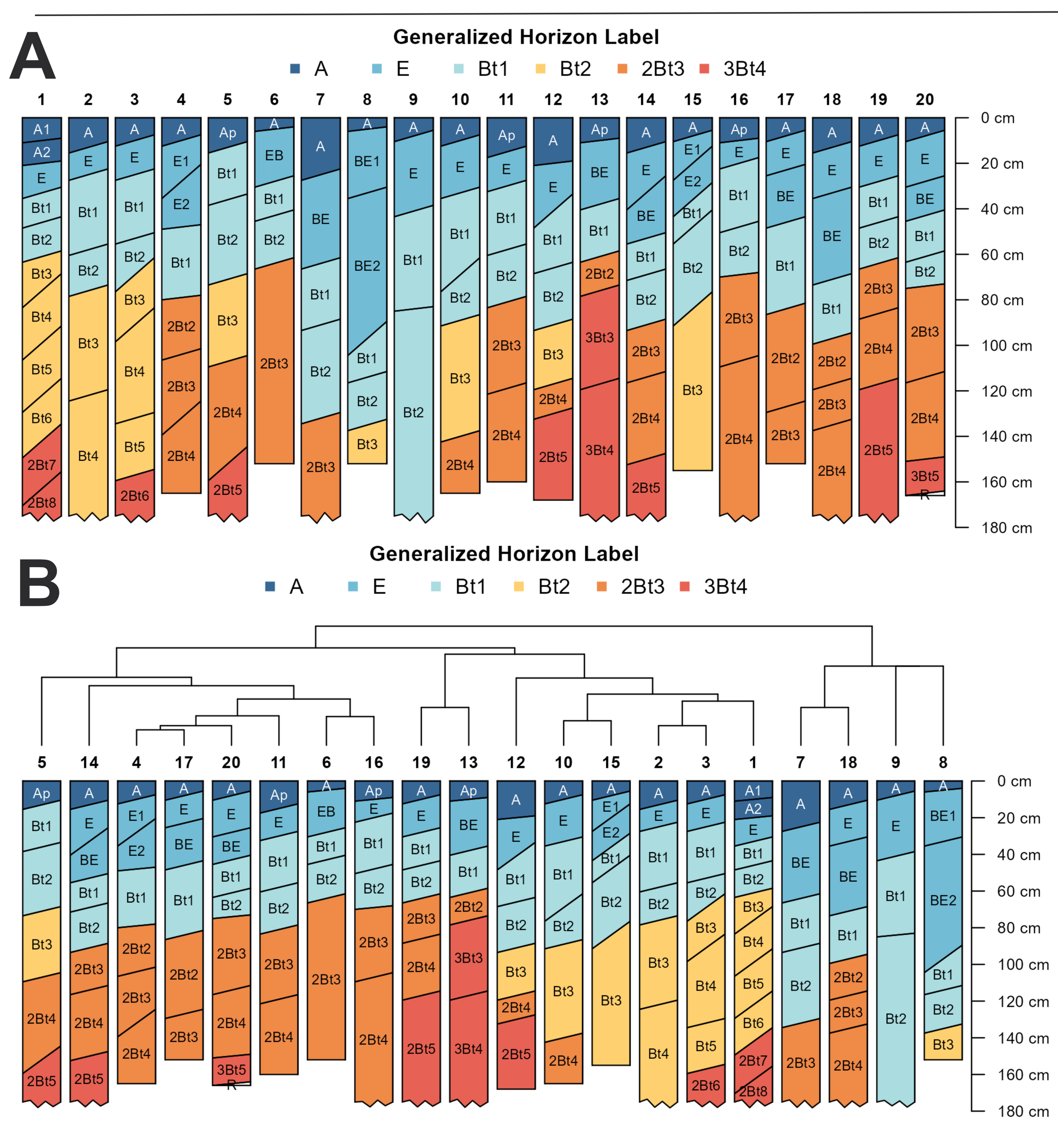
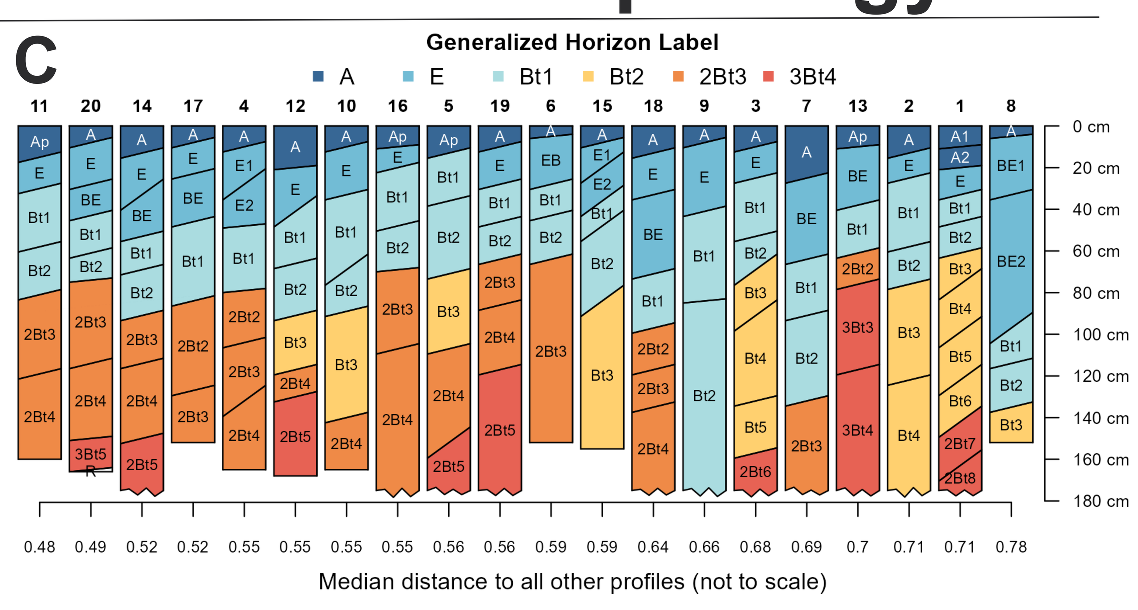


Figure 3: A graphical outline of the Numerical Classification of Soil Profiles (NCSP) algorithm implemented in the {aqp} package for R and described in Beaudette et al., 2013 and Maynard et al., 2020. This generalization of distance evaluation to soil profile depth-slices makes it possible to customize a classification system to specific soil properties (continuous and categorical) that explicitly weights comparisons according to depth-ordering, horizon thickness, and the possibility of varying depth to contact.

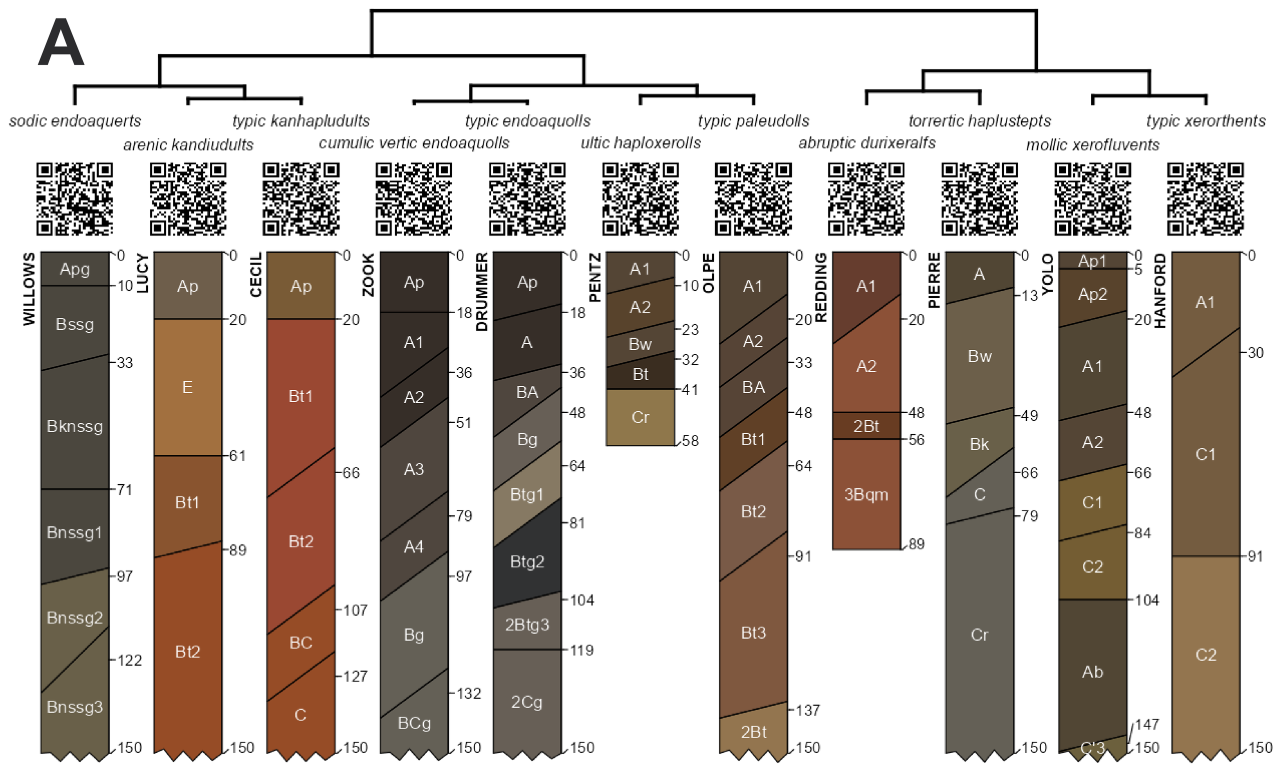
Numerical Classification of Soil Morphology

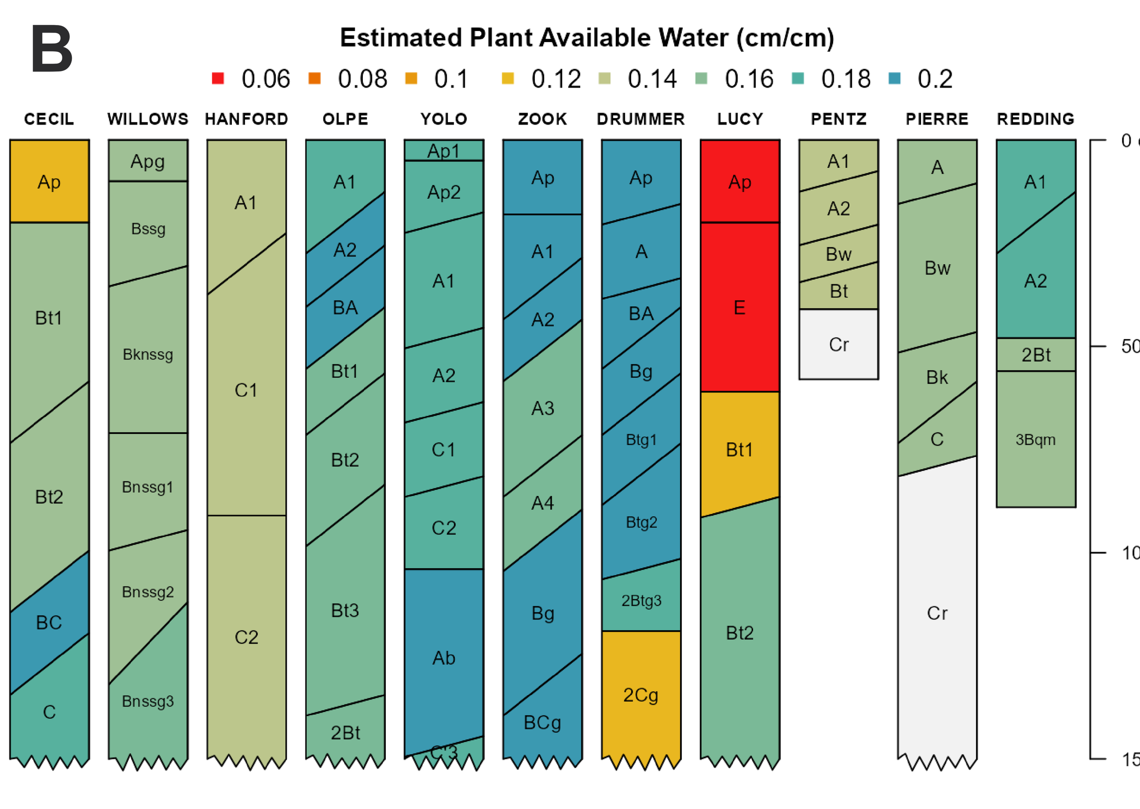


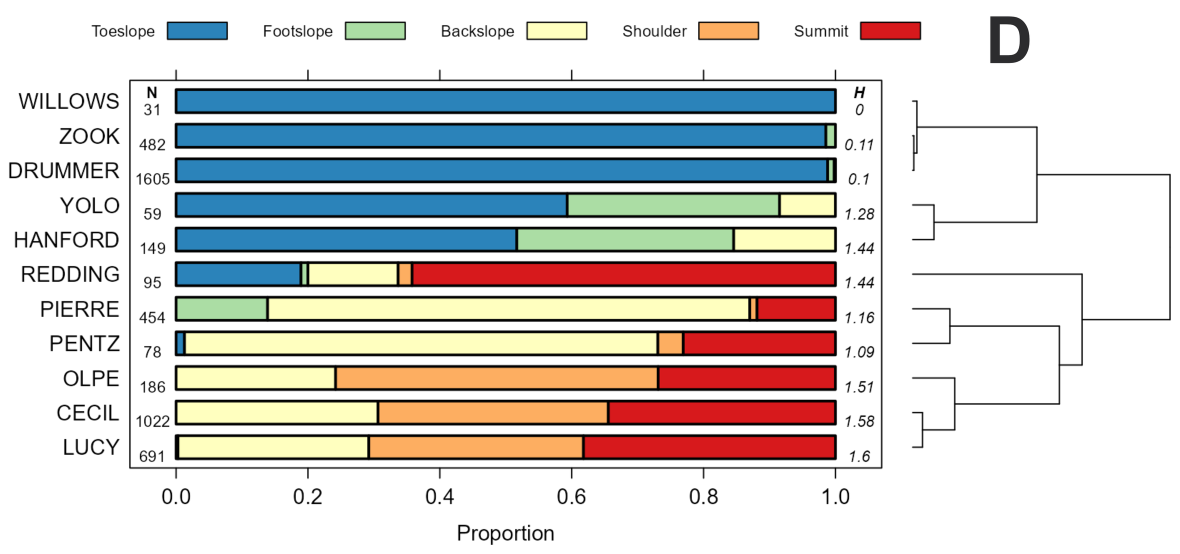


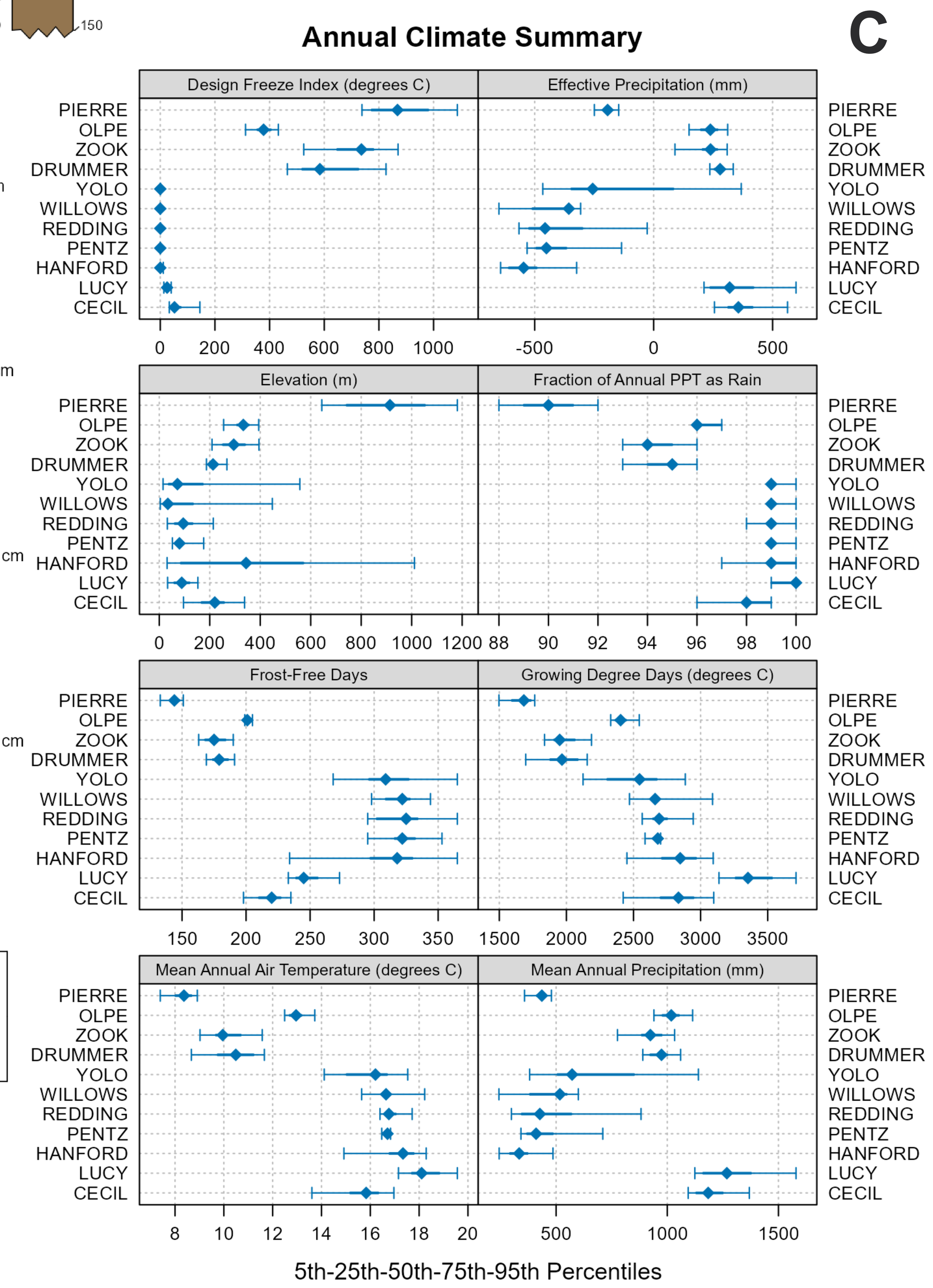
Figures 4A, 4B, 4C: A collection of 20 pedons (A) correlated to the Clarksville soil series (loamy-skeletal, siliceous, semiactive, mesic typic paleudults). Horizon colors indicate "generalized horizon labels" (GHL), a correlation of horizon designation to a simpler sequence of horizons (see QR codes for details). Dendrogram representation (B) of pair-wise distances computed by NCSP algorithm using only GHL. Clusters of similar pedons and outlier pedons (7, 18, 9, 8) are evident. Panel (C) demonstrates ranking of pedons according to median pair-wise distances in ascending order. Pedons 11 and 20 are most similar to the collection as a whole.

Official Series Descriptions

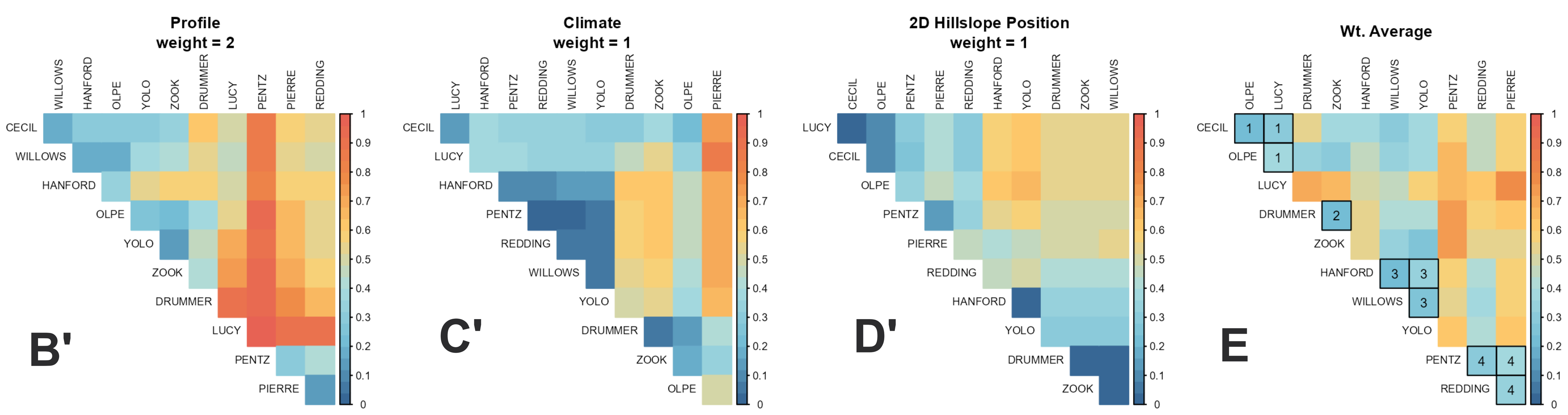








Figures 5A, 5B, 5C, 5D: A suite of soils with morphology and subgroup classification derived from Official Series Descriptions (A), organized according to clustering of distance matrix computed from ranked position of subgroups in the 12th edition of the "Keys to Soil Taxonomy". QR codes encode URLs to SoilWeb/Series Data Explorer. The same data with plant-available water holding capacity (PAW) estimated for each horizon, organized according to clustering of pair-wise distances computed by NCSP (B). Annual climate summaries sorted according to clustering of pair-wise distances between median values (C). Climate data from PRISM and summaries according to series extent. Panel D illustrates a similar clustering of pair-wise distances computed between 2D hillslope proportions, as derived from SSURGO. Each of these panels demonstrates an application of numerical classification to common sources of soil survey data. Comparisons similar to panels A, C, and D have been used in SoilWeb / Series Data Explorer since 2018.



Figures 6B', 6C', 6D', E: Distance matrices computed by NCSP using estimated PAW (B'), annual climate summaries (C'), 2D hillslope proportions (D'), and cell-wise weighted average (E). Row and column order are based on divisive hierarchical clustering (all panels), and match ordering of corresponding panels: 5B→6B', 5C→6C', 5D→6D'. Redder hues denote greater distance (i.e. less similar), bluer hues smaller distance (i.e. more similar). Four clusters are highlighted in panel E (outlined cells labeled1-4), suggesting groups of functionally-similar soils. Soils in cluster 1 have similar climates (e.g. udic soil moisture regime) and hillslope positions. Cluster 2 represents two *endoaquolls* with very similar PAW, climate, and hillslope position. Cluster 3 represents very deep soils from the San Joaquin Valley of California. Cluster 4 represents shallow and moderately deep soils from drier climates (xeric and ustic) which share very little in common with the other soils in this example. The selection and weighting of soil and site variables makes it possible to customize a classification system to specific soil function or interpretation, possibly spanning multiple levels of Soil Taxonomy.

Conclusions

- Concepts and methods from numerical classification (distance matrix, clustering thereof, dendrograms, ordination) can by applied to soil survey data.
- Distance metrics applied to ordinal data (ranked categories) support quantitative interpretation of soil morphology (horizon designation, structure).
- Development and interpretation of a distance matrix can lead to more objective selection of modal pedons for component and series concepts.
- Numerical classification will likely support semi-automated allocation of new and existing pedons to component and series concepts.
- Numerical classification may offer new linkages between specific soil function or interpretation as well as other classification systems such as Soil Taxonomy and World Reference Base.

References & Resources

Beaudette, D.E., P. Roudier, and A.T. O'Geen. 2013. Algorithms for quantitative pedology: A toolkit for soil scientists. Computers and Geoscience 52:258-268.

Maynard, J.J., S.W. Salley, D.E. Beaudette, and J.E. Herrick. 2020. Numerical soil classification supports soil identification by citizen scientists using limited, simple soil observations. Soil Science Society of America Journal 84:1675-1692.

Statistics for Soil Scientists
http://ncss-tech.github.io/aqp
https://ncss-tech.github.io/soilDB
https://ncss-tech.github.io/sharpshootR