

# Chapter 11

## Algorithms for Quantitative Pedology



D. E. Beaudette, J. Skovlin, A. G. Brown, P. Roudier, and S. M. Roecker

**Abstract** The Algorithms for Quantitative Pedology (AQP) project consists of a suite of packages for the **R** programming language that simplify quantitative analysis of soil profile data. The “*aqp*” package provides a vocabulary (functions and data structures) tailored to the complexity of soil profile information. The “*soilDB*” package provides interfaces to databases and web services, leveraging the “*aqp*” vocabulary. The “*sharpshootR*” package provides tools to assist with summary and visualization. Bridging the gap between pedometric theory and practice is central to the purpose of the AQP project. The AQP **R** packages have been extensively tested and documented, applied to projects involving hundreds of thousands of soil profiles, and integrated into widely used tools such as SoilWeb. These packages serve an important role in routine data analysis within the U.S. Department of Agriculture and in other soil survey programs worldwide.

**Keywords** Geomorphology · Pedology · Soil survey · Soil data analysis · Data visualization

---

**Supplementary information:** Annotated **R** code used to create the figures and table within this chapter are posted at <https://github.com/ncss-tech/geopedology-chapter>. The following package versions were used during the preparation of this content: *aqp* 1.42, *soilDB* 2.6.14, and *sharpshootR* 1.9.1.

---

D. E. Beaudette (✉) · A. G. Brown  
USDA-NRCS Soil and Plant Science Division, Sonora, CA, USA  
e-mail: [dylan.beaudette@usda.gov](mailto:dylan.beaudette@usda.gov)

J. Skovlin  
USDA-NRCS Soil and Plant Science Division, Missoula, MT, USA

P. Roudier  
Manaaki Whenua - Landcare Research, Palmerston North, New Zealand

S. M. Roecker  
USDA-NRCS Soil and Plant Science Division, Lincoln, NE, USA

## 11.1 Introduction

Soil survey data are a rich source of soil parent material information and observations relating to the distribution and extent of surficial deposits and geologic stratigraphy. The overlap of the domains of pedology and geomorphology, referred to as geopedology (Zinck 2016), emphasizes the need for clear relationships between soil parent materials and geomorphic concepts. Soil is the dynamic interface connecting the biosphere and the lithosphere (Wysocki et al. 2005). Soil provides a medium and conduit for water storage and nutrient flow. The development of soil properties is largely determined by the inherent nature of surficial deposits or soil parent material from which they originate. While soil survey has historically been grounded in a soil-landscape paradigm (Hudson 1992), geopedologic concepts aim to further integrate geomorphic concepts with soil survey information for broader application within the earth science community.

Data alone cannot support decisions, generate useful conclusions, or convey embedded relationships without thoughtful analysis and processing. The wide array of data sources, formats, and conventions (even within a single institution) can further complicate efforts to synthesize soil information from soil data sources. Growth in promising tools that link database APIs, spatial data, and enable the analysis of complex soil description data are expanding the potential for soil science data analysis within programming environments like **R** (R Core Team 2022).

The Algorithms for Quantitative Pedology (AQP) project are a suite of packages for the **R** programming language that simplify many facets of soil data analysis. The project began in 2006 as a loosely coordinated collection of **R** scripts used to support the management, analysis, and visualization of digital soil morphology records. By 2010, it became clear that an **R** package (code, manual pages, and example data following strict guidelines) hosted by CRAN would be the best route forward. The first version of the *aqp* package was submitted to CRAN in May of 2010; with a name and core functionality inspired by the concept of “quantitative pedology” (Jenny 1941), analysis by regular depth-intervals (Harradine 1963; Moore et al. 1972), and the characterization of depth-functions (Myers et al. 2011). A companion article by Beaudette et al. (2013b) contained a detailed description and simple demonstrations of the main package features.

As functionality evolved, the *aqp* package was split into three main categories which became **R** packages to increase modularity and divide administrative tasks: *aqp* (soil-specific data structures, profile sketches, color conversion, pedotransfer functions, etc.), *soilDB* (wrapper and convenience functions for accessing APIs and harmonization of results), and *sharpshootR* (specialized tasks and visualizations designed for use with soil database connections provided by *soilDB* and data structures provided by *aqp*). Like most **R** packages, the AQP suite of packages depends on other packages for optimized computation (Dowle and Srinivasan 2021), color conversion (Pedersen et al. 2021; Zeileis et al. 2020), numerical classification (Maechler et al. 2021), and methods for compositional data (e.g., sand, silt, and clay content) (Moeys 2018; van den Boogaart et al. 2021), to name a few. The authors

hope that other scientists will find a suitable foundation in *aqp*, *soilDB*, and *sharpshootR*, upon which more specialized tools can be built, documented, and delivered in the form of new **R** packages.

Since 2011, the AQP suite of **R** packages has been extensively updated and documented by U.S. Department of Agriculture – Natural Resources Conservation Service (USDA-NRCS) Staff to support routine operations within the Soil and Plant Science Division. Some examples include aggregation and synthesis of field data to support initial soil survey (new mapping), graphical comparisons and correlation analysis to support soil survey update projects (refinement of existing mapping), and visual presentation of soil survey data to the public via tools like SoilWeb (O’Geen et al. 2017). The *soilDB* package has become one of the most widely used interfaces to USDA-NRCS data sources, with support for queries that accept (and return) mixtures of spatial and tabular data from the Soil Survey Geographic Database (SSURGO) (Soil Survey Staff 2022b). Spatial formats defined by the *sf* (Pebesma 2018) and *raster* (Hijmans 2021) **R** packages are used extensively by the *soilDB* package to minimize data conversion or pre-processing steps.

### 11.1.1 Example Data: Clarksville Soil Series (Fig. 11.1)

A curated set of soil morphologic and laboratory characterization data correlated to the Clarksville soil series (Loamy-skeletal, siliceous, semiactive, mesic Typic Paleudults) is used to demonstrate key functionality and visualization possibilities provided by the AQP suite of **R** packages. These data represent a very deep (>150 cm), somewhat excessively drained soil of large extent in the Ozark Highlands of southern Missouri, USA. Clarksville soils are formed in residual and colluvial soil parent materials of cherty dolomite or cherty limestone (Kabrick et al. 2008).



**Fig. 11.1** Clarksville series soil profile (left) and associated representative landscape (right). (Soil profile photo: Satchel Gaddie, landscape photo: Jayme LeBrun)

These soils typically occur on ridges and steep side slopes, spanning summit, shoulder, and backslope positions of an idealized 2D hillslope. Mean annual air temperature ranges from 13–15 °C, mean annual precipitation ranges from 1150–1250 mm, with most precipitation falling as rain.

Clarksville soils are generally highly weathered, acidic with low to moderate base saturation, low cation exchange capacity and nutrient limiting for available phosphorus, calcium, and magnesium (Kabrick et al. 2011; Singh et al. 2015). Morphology of Clarksville soils commonly include soil textures high in silt with thick accumulations of translocated clay at depth. Although soils are generally high in rock fragments, silt-rich soil textures dominate surface soil horizons due the influence of wind-blown loess parent material. These landscapes support a mixed forest of black oak (*Q. velutina* Lam.), white oak (*Quercus alba* L.), blackjack oak (*Q. marilandica* Muench.), post oak (*Q. stellata* Wangenh.), shortleaf pine (*Pinus echinata* Mill.), black hickory (*Carya texana* Buckl.), red maple (*A. rubrum* L.), and dogwood (*Cornus florida* L.).

## 11.2 Representing Collections of Soil Profiles in R

Soil profile data are complex, and typically consists of site description, soil morphologic description, and optionally laboratory data. The SoilProfileCollection (SPC) is a data structure which attempts to capture this complexity and is designed to coordinate linkages between those elements. Functions operating on the SoilProfileCollection include special constraints to ensure linkages are not broken during routine operations such as editing, sub-setting, or combining collections.

The first level of abstraction involves two main tiers: “site” and “horizon” data. “Site” data refers to above-ground or those properties that are specific to a single soil profile description (e.g., surface slope). “Horizon” data refers to below-ground or those properties that are specific to a single genetic soil horizon or layer. An additional level of abstraction is used to store spatial data (coordinates and coordinate reference system) and depth-interval information such as diagnostic horizons. The SoilProfileCollection structure provides a means of storing user-defined metadata such as units of measure, horizon designation column name, data source, and data citation. The SoilProfileCollection object was designed with data analysis in mind; as compared to other (more complex) data structures used for archival purposes, such as those used within the USDA-NRCS National Soil Information System (NASIS).

Of primary importance are the horizon data, or the layers that comprise the profiles in the collection. The SoilProfileCollection is “horizon data forward,” in that a user starts with a table of horizon data. Each horizon record must have an upper and lower boundary, a unique ID linking to a single soil profile observation, and any other observed or measured properties. There are no set limits on the number of horizons per profile, or profiles per collection, but available memory will dictate practical limitations. Horizon depths should be specified as integers (typically centimeters) and should not overlap.

A `SoilProfileCollection` is created through “promotion” of an **R** `data.frame` with the `depths()` function. Other `data.frame`-like objects such as `tibble` (Müller and Wickham 2021) or `data.table` (Dowle and Srinivasan 2021) can be used as input. Promoting a `data.frame` to `SoilProfileCollection` requires the following parameters: `profile_id` (the name of a column containing unique profile IDs), `top_depth` (the name of a column containing horizon upper depths) and `bottom_depth` (the name of a column containing horizon lower depths) along with any additional horizon data associated with the horizons in the profile. For example, the promotion of a `data.frame` called `x` to `SoilProfileCollection` would follow `depths(x) <- profile_id ~ top_depth + bottom_depth`.

In the **R** language, the tilde symbol `~` separates the left and right-hand sides of a formula. Commonly `~` is used in formulas to mean “modeled as.” In a `SoilProfileCollection` the geometry and ordering of horizons within each unique profile using the upper and lower depths is “modeled.” Performing this operation automatically sorts horizon data first by profile ID and then by horizon top depth.

The `site()<-` method is used to move site-level data from horizon-level records (necessary when starting with a mixture of horizon and replicated site data in the same table), or to merge a new table of site-level data via common ID and left join (missing records in the new table are filled with NA). In a similar manner, the `horizons()<-` method is used to merge a new table of horizon-level data into the `SoilProfileCollection` object via common ID (unique to specific horizons) and left join. Additional site and horizon data can be created or extracted one by one using the `$` or `[[` methods. When creating a new variable, the `SoilProfileCollection` will check whether the length of the vector matches either the number of “sites” or the number of “horizons.” Extracting horizon or site-level data as plain `data.frame` objects is performed with the `horizons()` and `site()` functions. A detailed explanation of the `SoilProfileCollection` object and associated methods for manipulation of these objects is presented in the “Introduction to `SoilProfileCollection` Objects” tutorial (Beaudette 2022).

The `soilDB` package for **R** provides a common interface to many of the National Cooperative Soil Survey databases. Several functions from this package return data as a `SoilProfileCollection` object: `fetchKSSL()` (laboratory characterization data), `fetchOSD()` (basic soil morphology from the Official Series Description), `fetchSDA()` (SSURGO and STATSGO data from Soil Data Access), and `fetchNASIS` (National Soil Information System).

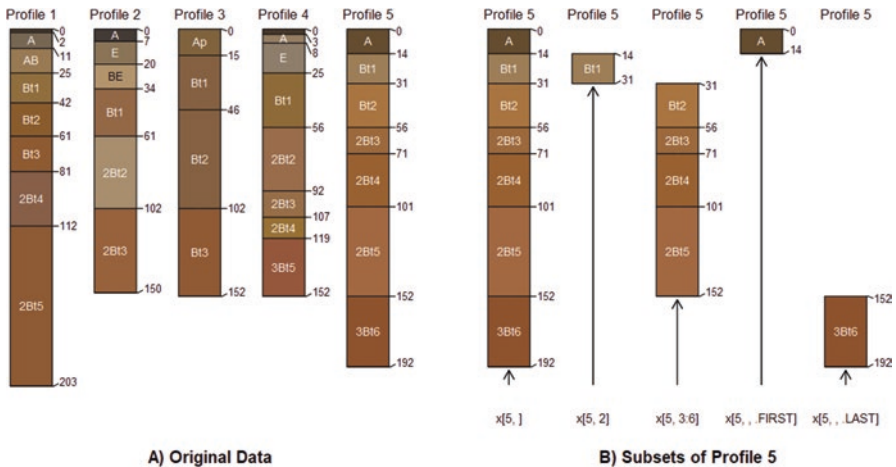
### 11.2.1 Subsetting

The “bracket” methods are one of the primary ways that objects in **R** can be subset by rows and columns (e.g., `data.frame`) or element (e.g., list, vector, etc.). The `SoilProfileCollection` builds on these patterns to extract specific profiles and/or horizon collections based on numeric or logical indices.

The syntax used by the SoilProfileCollection bracket method is  $x[i, j, k]$ ; where  $x$  is a SoilProfileCollection object,  $i$  is a profile index,  $j$  is a within-profile horizon index and  $k$  represent optional special functions that can operate on the horizon data in the collection to replace the profile-specific  $j$ -index (Fig. 11.2). To obtain the first profile in the collection use the syntax  $x[1, ]$ . For the first horizon in each profile use  $x[, 1]$ . To get the last horizon of each profile, use  $x[, .LAST]$  where `.LAST` is a special “keyword” that can identify the  $j$ -index of the deepest horizon in each profile. Subsets based on  $i, j$  and  $k$  indices of the SoilProfileCollection can be combined, for instance:  $x[1:2, 1:2]$  gives the first two horizons of the first two profiles. Also, the  $k$  index can be combined with the  $i$  index, for instance  $x[1:2, , .LAST]$  gives the last horizon of the first two profiles (Fig. 11.2).

The representation of horizon position with the  $j$ -index can be extended to develop other “horizon spatial predicates” such as `hzAbove()`, `hzBelow()` and `hzOffset()`. The former two take logical expressions to match horizons and return the part of the collection adjacent to the match (above or below respectively). The `hzOffset()` function allows arbitrary horizon indices and offsets to be calculated. This type of logic is further helpful for inspecting and fixing horizon geometry for errors or inconsistencies.

Common querying operations with criteria in the form of logical expressions can be used to subset profiles or horizons in a collection that meet specific criteria of interest. The *app* functions `subset()` and `subsetHz()` can be used with logical expressions in terms of the site or horizon variables to specify the constraints. These expressions make use of site or horizon-level variables in the collection. The `subset()` function returns whole profiles, if criteria were specified for horizon data, then only some of the horizons of those profiles may meet criteria. More specifically, `subsetHz()` requires horizon-level expressions and returns only the portion of horizons within profiles that meet criteria.



**Fig. 11.2** Five soil profiles correlated to the Clarksville soil series (a). Examples of bracket methods for subsetting profiles, sequences of horizons and top or bottom horizons (b)



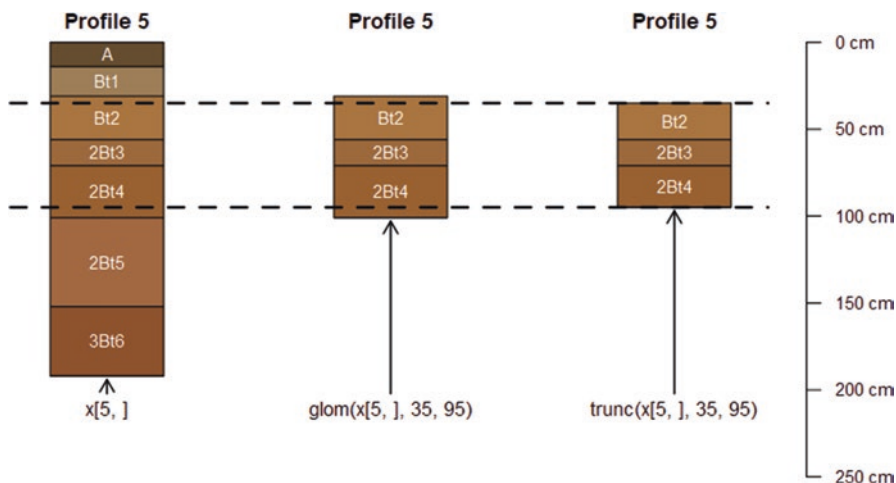
Partitioning soil profile collections on logical expressions of site and horizon level properties is powerful, but soil scientists often need to extract data within or overlapping specific depths. Two methods: `glom(x, z1, z2, ...)` and `trunc(x, z1, z2, ...)` facilitate this in *aqp*. The `glom()` function returns the subset of horizons in a collection that overlap with a specific depth interval  $[z1, z2]$ . The depth interval could be a point (only  $z1$  specified) or a range ( $z1$  and  $z2$  specified) (Fig. 11.3).

The interval  $[z1, z2]$  can be constant across the collection or unique to each profile. By default, the whole horizon is returned unmodified whether it falls fully within the range or not – creating a “ragged” `SoilProfileCollection` (Fig. 11.3). The upper and lower boundaries of the resulting horizons will be cleanly cut to the interval specified using `glom(truncate=TRUE)` or via `trunc()` and resulting profiles will have consistent upper and lower boundaries assuming there are no missing data in the specified interval.

### 11.2.2 Data Quality and Repairs

The *aqp* package provides several methods for identifying problematic profile geometry and attempts to “correct” it. Most soil databases and methods for storing soils information do not have front end validations that prevent entry of data with “illogical” content. Some analyses rely on having only one record of data per depth/profile combination such as those involving depth-weighted averages or those that rely on having a “complete” set of records in all profiles over a particular interval.

The *aqp* function `checkHzDepthLogic()` inspects a `SoilProfileCollection` object looking for four common errors in horizon depths: bottom depth shallower than top



**Fig. 11.3** A demonstration of selecting horizons that overlap with a depth interval via `glom()` and truncation to that interval via `trunc()`

depth, equal top and bottom depth, missing top or bottom depths, and gap or overlap between adjacent horizons. With `byhz = TRUE` it is possible to perform the first three of the above logic checks on individual horizons.

Assumptions concerning horizon order based on horizon top depth are tested by the `repairMissingHzDepths()` function. This can be used to fill in some missing (bottom) horizon depths. This function will set missing bottom depths of a horizon to the next deepest (adjacent) top depth. Also, it adds a constant vertical offset to the top depth of bottom-most horizons missing bottom depth.

The `fillHzGaps()` function attempts to find “gaps” in the horizon records of a `SoilProfileCollection` object and fill with placeholder horizons (profile ID, horizon ID, top/bottom depths, all else NA). This function is for filling profiles to a static top and bottom depth. For instance, a morphologic description containing horizons with omitted upper or undetermined lower depth as in the case of undescribed organic horizons or soil profiles that have variable bedrock depth below depth of excavation.

## 11.3 Soil Morphology

The field description of a soil profile (genetic horizon depths, boundaries, color, soil texture, rock fragment volume, structure, etc.) is typically the foundation upon which additional sampling, laboratory characterization, or soil survey are based. In aggregate, a complete collection of horizons, associated properties, and landscape context (e.g., catenary position or other geomorphic description) represent an atomic unit of pedologic inquiry: the pedon (Soil Science Division Staff 2017). The AQP family of **R** packages and the `SoilProfileCollection` data structure were designed specifically to elevate the pedon (and collections of pedons) to a convenient abstraction (an object), enabling a simpler interface to what would otherwise be a complex hierarchy of above and below-ground records. In *aqp*, the more generic term “profile,” is used instead of pedon to accommodate incomplete data (missing above-ground information) or otherwise truncated horizon observations. Central to this approach is the specification of profile IDs and horizon depths, above-ground (“site”) vs. below-ground (“horizon”) attributes, and ideally horizon designation with associated attributes such as soil color.

### 11.3.1 Soil Color

The color of soil material observed during field investigations is one of the most striking and useful properties recorded as part of a soil profile description. Typically recorded in the Munsell system (Munsell 1947; Simonson 1993; Soil Science Division Staff 2017) in the form of “hue, value/chroma,” the three components of this notation provide interpretive suggestions about iron oxides and oxidation state (hue and chroma) (Schwertmann 1993; Scheinost and Schwertmann 1999), soil



carbon (value) (Wills et al. 2007; Liles et al. 2013), as well as hints about the relative importance of catenary relationships (Brown et al. 2004). Several color-based metrics of soil development (Buntley and Westin 1965; Harden 1982), rubification (Barron and Torrent 1986; Hurst 1977), and melanization (Harden 1982; Thompson and Bell 1996) are implemented in the *aqp* package.

### 11.3.1.1 Color Conversion

The *aqp* package provides several interfaces for conversion between Munsell notation and sRGB or CIELAB color spaces, largely based on the 1943 Munsell renotation table (Centore 2012). Forward conversion from standard Munsell notation (e.g., 10YR 3/4) is performed via look-up table, derived from the renotation table, and interpolated to include odd chroma and 2.5 value. The function `munsell2rgb()` performs a direct transformation to sRGB-encoded colors in hexadecimal (#5C4222), sRGB coordinates scaled to the interval of 0–1 ([0.36187, 0.25989, 0.13375]), or CIELAB coordinates ([30.273, 7.2731, 23.753]) (Beaudette et al. 2013a, b). Inverse transformation from sRGB coordinates is performed by the `rgb2munsell()` function, approximated by nearest-neighbor search of the Munsell-sRGB look-up table using the CIE2000 color contrast metric (Pedersen et al. 2021). All color space coordinates are referenced to the CIE standard illuminant D65, which is a close approximation to average midday sunlight in the northern hemisphere (Marcus 1998). sRGB and CIELAB color spaces were selected to address two common applications: sRGB, for digital representation of color on computer screens or reproduction on printed media, and CIELAB for the convenient alignment of axes and common pigments in the soil environment (Viscarra Rossel et al. 2006; Liles et al. 2013).

Non-standard notation of Munsell colors (e.g., 10.6YR 3.3/5.5), as collected by digital colorimeter, can be converted to approximate sRGB coordinates using the `getClosestMunsellChip()` function. However, this approach uses rounding of value and chroma and snapping to the nearest standard hue (10YR). Exact conversion of non-standard Munsell notation can be performed using the *munsellinterpol* R package (Gama et al. 2021).

### 11.3.1.2 Color Contrast

Color contrast (perceptual difference between two colors) within a soil sample is an important component of field-described redoximorphic features, concentrations, and mottles (Schoeneberger et al. 2012). Soil Survey products and wetland delineation protocols adopted by the National Cooperative Soil Survey (NCSS) currently use *contrast classes* (faint, distinct, and prominent) to describe color contrast, based on differences in Munsell hue, value, and chroma (Soil Survey Staff 2022c). The `colorContrast()` function in *aqp* computes differences in Munsell (hue, value/chroma), soil color contrast class, and the CIE2000 color contrast metric (Sharma

**Table 11.1** Output from the `colorContrast()` function includes: change in hue (dH), change in value (dV), change in chroma (dC), CIE2000 color contrast (dE00), and NCSS soil color contrast class (CC)

Color 1	Color 2	dH	dV	dC	dE00	CC
10YR 3/3	10YR 3/4	0	0	1	3.13	Faint
7.5YR 6/6	5YR 4/6	1	2	0	21.1	Distinct
2.5Y 2/2	5G 4/8	9	2	6	30.5	Prominent

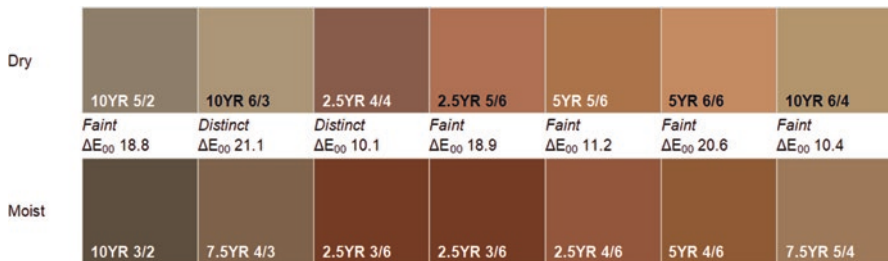
et al. 2005) for pairs of colors specified in Munsell notation. The function is fully vectorized meaning that multiple comparisons can be generated without explicit looping (Table 11.1).

Tabular color contrast output can be convenient when used as an intermediate step in a more complex workflow but can be difficult for non-specialists to interpret. A graphical representation of these data is created by the `colorContrastPlot()` function provided by *aqp*. For example, the differences between typical dry and moist soil colors for the Musick soil series (Fine-loamy, mixed, semiactive, mesic Ultic Haploxeralfs) are demonstrated in Fig. 11.4. While exact replication of Munsell colors is not possible on un-calibrated displays or printers, the sRGB approximation is sufficient to demonstrate relative differences in hue, value, and chroma.

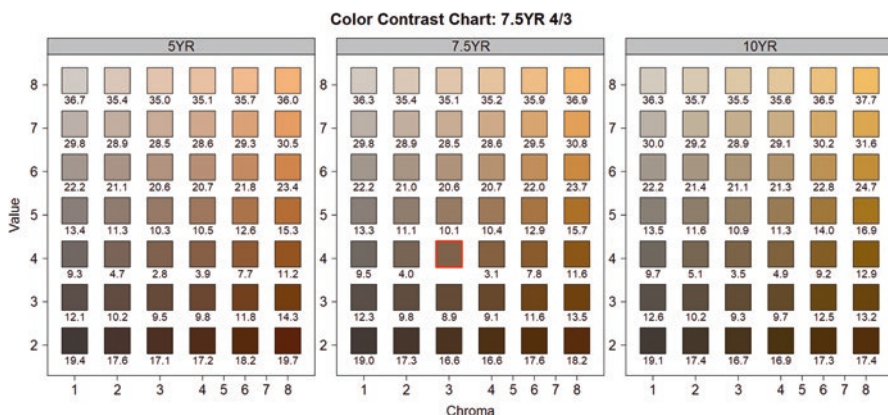
To further aid with the calculation and interpretation of color contrast, “color contrast charts” can be created with the `contrastChart()` function provided by *aqp*. These charts are based on a source color in Munsell notation (e.g., 7.5YR 4/3) and select pages of Munsell hue. Pair-wise metrics of color contrast are evaluated between all color “chips” and the source color (outlined in red). Hue is split across panels in a familiar format with Munsell chroma on the x-axis and value on the y-axis. Soil color contrast class and CIE2000 values are printed below each color “chip” (Fig. 11.5).

### 11.3.2 Soil Profile Sketches

Conceptual sketches of soil profiles that illustrate variation in morphology (e.g., horizon depths, horizon designations, color, texture, etc.) in relation to transect or catenary position are a pedologic staple. Either hand-drawn in field notes or carefully produced as part of a final soil survey manuscript, these sketches represent an important vehicle for communicating observation and context to technical and non-technical audiences. A data-driven approach to creating soil profile sketches was one of the original motivations for the *aqp* R package (Beaudette et al. 2013a, b). Since 2010, the profile sketch authoring tools in *aqp* have progressed from basic layout of filled rectangles (profiles and horizons) to thematic coloring of horizons based on properties or classes, encoding of horizon boundary information, and handling of label collision.

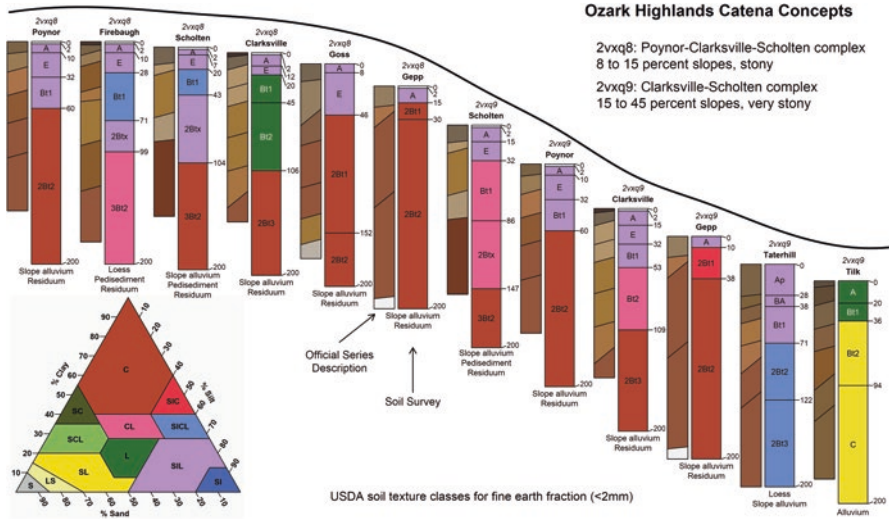


**Fig. 11.4** Color contrast plot, comparing the moist and dry soil colors of the Musick soil series. CIE2000 color contrast values are printed below soil color contrast classes. Smaller values describe smaller perceptual differences between colors



**Fig. 11.5** Color contrast chart for 7.5YR 4/3 (chip outlined in red), including reference 5YR, 7.5YR, and 10YR hue pages. Numbers below soil color chips represent CIE2000 color contrast values, as compared with the target color 7.5YR 4/3. Soil color contrast classes have been omitted for clarity

The `plotSPC()` function in the *aqp* package is the primary tool for creating soil profile sketches from `SoilProfileCollection` objects, using **R**'s “base graphics” system. Figure 11.6 demonstrates several possible data sources, processing steps, and output generated from `plotSPC()`. Soil components (retrieved from the detailed Soil Survey via `fetchSDA()` as a `SoilProfileCollection`) within map unit “2vxq8” occur on summit and shoulder hillslope positions while components in map unit “2vxq9” occur on backslope, footslope and toeslope positions. USDA soil texture classes (<2 mm fraction) of each horizon are symbolized with color to show the variation of textures within the catena. Labeling of horizon depths (vs. common depth axis), leader lines, and collision detection (common with thin horizons) are optional enhancements to the standard output, specified via function arguments (Soil Survey Staff 2022a). Narrower profiles to the left of each component sketch represent data from the Official Series Descriptions via `fetchOSD()`. These data represent the typical morphology (horizon depths, designations, colors, etc.) for all soil series used in the US Soil Survey. Munsell colors (moist conditions) have been converted to sRGB



**Fig. 11.6** Illustration of an idealized hillslope catena for soil components from two adjacent map units within the Ozark Highlands. SSURGO map unit soil components are placed within a geomorphic hillslope sequence to convey soil property to soil parent material relationships. Munsell soil colors from each official series description are displayed in tandem, companion profiles

coordinates using `munsell2rgb()`, and horizon boundary distinctness codes have been converted into vertical offsets using `hzDistinctnessCodeToOffset()`. The `plotSPC()` function can encode horizon distinctness offsets as diagonal horizon boundaries, where increasingly steeper angles represent the following sequence of boundary distinctness: “very abrupt,” “abrupt,” “clear,” “gradual,” “diffuse” (Schoeneberger et al. 2012). A visual explanation of the many arguments to `plotSPC()` is provided via `explainPlotSPC()` which shows the usage of ordering vectors, graphical offsets and scaling factors within the graphical space. Detailed examples of `plotSPC()` usage are available in the function documentation (Soil Survey Staff 2022a) and associated tutorials (Beaudette 2022). Future developments to `plotSPC()` will include conversion to the more advanced “grid” graphics system, pattern fills (e.g., geologic and stratigraphic symbols), and tighter integration with other plotting libraries such as *lattice* and *ggplot2*.

### 11.3.3 Functional Horizon Aggregation

Soil scientists use a common language of horizon designation nomenclature to describe and articulate the observed differences in soil horizons within a soil profile. These basic notations and the act of “naming” genetic horizons distills important information in the form of master horizons, characteristic subscripts, horizon and pedogenic sequences, and parent material discontinuities. Horizon designations

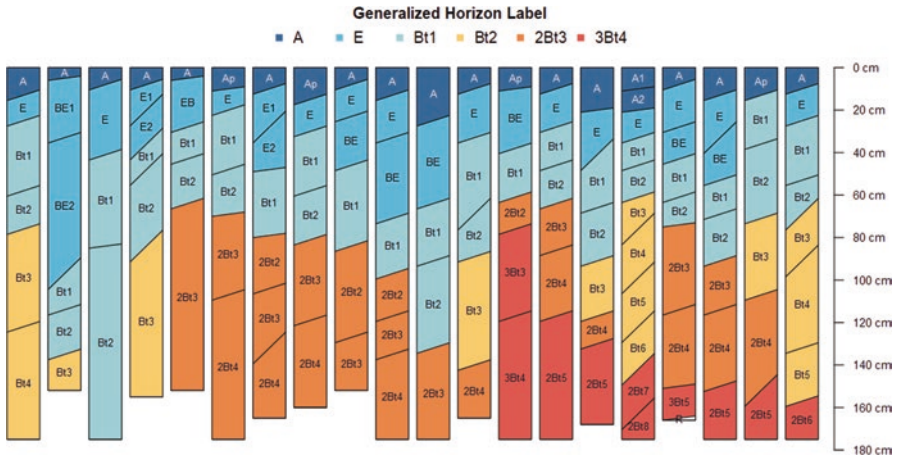
convey a qualitative description of soil properties and process while allowing flexibility in how horizon designations are applied. Experienced soil scientists will generally apply horizon nomenclature consistently from site to site and through time due to the rigid guidelines and definitions of their application. However, among a group of soil scientists there will be variability in the exact designations used—each having their own unique training, field experiences, and tendencies towards “lumping” (describing fewer and thicker horizons) or “splitting” (describing more and thinner horizons).

Building on the interpretation of horizon designations, the Generalized Horizon Label (GHL) concept seeks to unify functionally similar horizon designations for the purpose of aggregation, analysis, and summary operations (Beaudette et al. 2016; Roecker et al. 2016). The conceptual approach of functional aggregations of horizons within collections of soil profiles has been attempted as a framework for developing pedotransfer functions (Wagenet et al. 1991). The process of applying GHL to a collection of soil profiles involves a series of micro-correlations made by a soil scientist to determine which horizons have similar soil morphology and properties to be grouped together for aggregation across horizons within a `SoilProfileCollection`.

The application of GHL to a `SoilProfileCollection` is driven by Regular Expression (REGEX) pattern matching. REGEX provides a rich syntax for string-matching of horizon designation sets into unified horizon GHL groups. A user developed set of REGEX rules are matched to an identified vector of horization by the `generalize.hz()` function in the *aqp* **R** package.

Features unique to the Clarksville soil series are argillic diagnostic horizons and parent material discontinuities as horizons transition along boundaries between cherty dolomite slope alluvium and colluvium over dolomite residuum. GHL can be used to identify these common features in a set of soils once applied (Fig. 11.7).

The `depthOf()` family of functions also utilize REGEX pattern matching of horizon designations to determine the depth to horizons with matching designation. For example, the pattern “Bt” would match all horizon designations containing “Bt” (case sensitive). Within a `SoilProfileCollection`, the `minDepthOf()` and `maxDepthOf()` operations provide additional utility to find either the top (shallowest) or bottom (deepest) depth to a matching horizon pattern. These functions provide convenience handling for missing values or when target patterns are not found within a profile. Results are returned as a numeric vector for single profiles or a `data.frame` of results with profile ID, horizon ID, top or bottom depths, horizon designation and pattern provided. Profile sketches in Fig. 11.7 have been sorted by values returned by the `depthOf()` function, first according to depth of “3Bt4” GHL and second according to depth of “2Bt3” GHL.



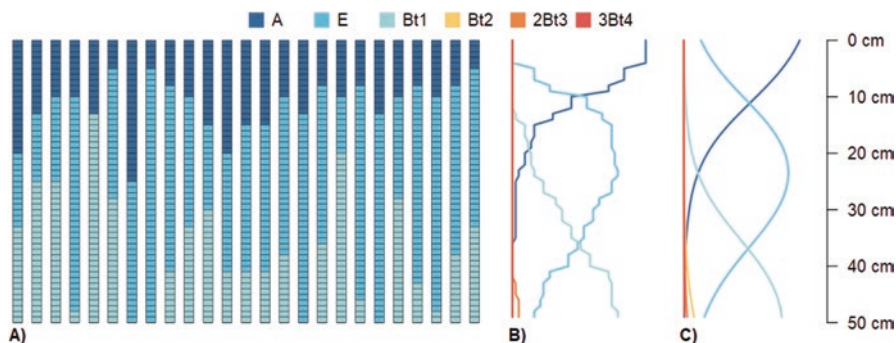
**Fig. 11.7** A series of generalized horizon labels (GHL) applied to a collection of soil profiles of the Clarksville soil series from the Ozark Highlands, Missouri. Horizon labels within each soil profile show the original horizon designations while colors indicate assigned GHL

### 11.3.4 Change of Depth Support

Soil data are typically collected either by genetic horizon with varying horizon depths, at regular depth intervals (every 10 cm), or from composite samples representing specified depth intervals (0–10 cm, 10–25 cm, etc.). The structure of these depth intervals will typically vary from one profile to the next. To facilitate analysis throughout the profile collection, profile horizon depths may need to be modified and/or harmonized comprising a change of depth support. A simple down-scaling of horizons (without interpolation) into a regular sequence of thinner depth slices, referred to as “slicing,” is implemented in the `dice()` function (Fig. 11.8a). The `segment()` function offers another approach to restructuring horizon depths, using horizon-thickness weighted mean values for conversion to fixed depth intervals (e.g., 0–25 cm). This is a common step in the thematic mapping of soil property data. A more complex change of horizon depths can be achieved using constrained interpolation. This method, popular for applications such as digital soil mapping (or other tasks requiring harmonized horizon depths) uses mass-preserving splines (Bishop et al. 1999). In *aqp*, this type of down-scaling is performed with the `sp2mpspline()` function which provides a convenient interface to the *mpspline2* package (O’Brien 2022) suitable for *SoilProfileCollection* objects.

A change of support operation performed over all profiles in a collection result in a statistical summary of that collection. The `slab()` function performs this kind of operation by “slicing” horizon data (continuous or categorical) into 1 cm-thick depth intervals (Fig. 11.8a), then aggregating “across” those depth intervals (Harradine 1963; Beaudette et al. 2013a, b). Continuous values are reduced to select percentiles (or any user-defined function) per slice, and categorical values are





**Fig. 11.8** Change of support demonstration using 25 pedons correlated to the Clarksville soil series. Generalized horizon labels (GHL) have been resampled to 1cm slices (a) via `dice()`, aggregated across slices to GHL proportions (b) via `slab()`, and modeled via proportional-odds logistic regression (c)

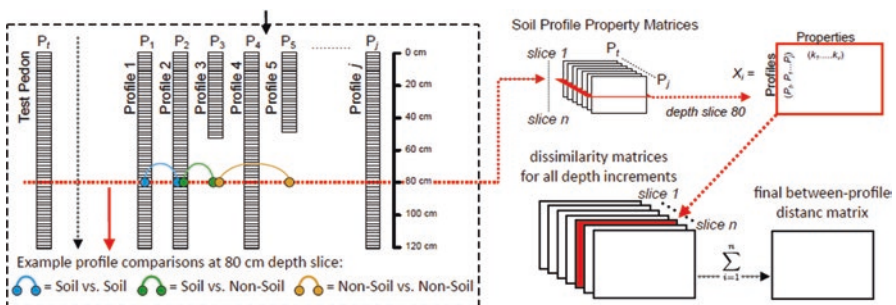
reduced to proportions (Fig. 11.8b) (Beaudette et al. 2013a, b). Alternatively, sliced categorical data such as GHL can be aggregated using statistical models for ordinal data, such as the proportional-odds logistic regression model (Fig. 11.8c) (Beaudette et al. 2016).

## 11.4 Numerical Classification of Soils

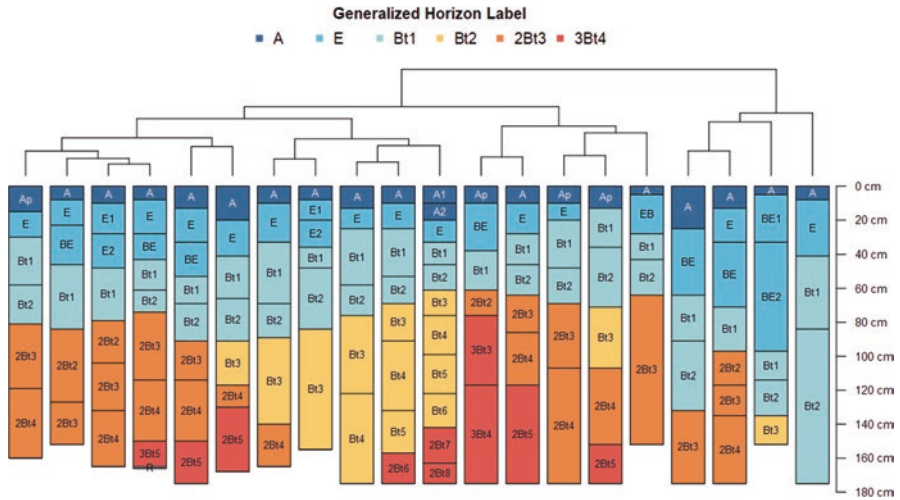
Since the 1960s (likely corresponding with increased availability of computing hardware) there has been considerable interest in the development of numerical alternatives to traditional soil classification systems such as Soil Taxonomy (Soil Survey Staff 1999) and World Reference Base (Chesworth et al. 2008). A “numerical taxonomy” (Sneath and Sokal 1973) of soil horizons or collections of horizons (i.e., soil profiles or aggregation thereof) relies on a deliberate selection of characteristics (soil properties), distance metric (e.g., Euclidean) and criteria used to identify clusters (e.g., hierarchical vs. partitioning methods) (Arkley 1976). Selection of characteristics is complex because a limited set of soil properties cannot universally describe differences between individuals, and the use of all measurable properties is unfeasible – requiring a selection to be made prior to analysis (Sarkar et al. 1966; Arkley 1971). Furthermore, a generalized approach to the numerical classification of soil profiles is complicated by the hierarchical nature of linked site and horizon-level properties, sampling style (depth-intervals vs. genetic horizons) and subtle differences in horizon designation (through time, regionally, etc.). Despite the many challenges, there have been many successful applications of numerical taxonomy to soil science and soil classification (Hole and Hironaka 1960; Rayner 1966; Moore et al. 1972; Dale et al. 1989; Carré and Jacobson 2009).

The NCSP (Numerical Comparison of Soil Profiles) algorithm, implemented in the *aqp* package for **R**, attempts to address many of the long-standing difficulties with a numerical classification of entire soil profiles (Beaudette et al. 2013a, b; Maynard et al. 2020). Building on methods suggested by Moore et al. (1972), pair-wise distances (between soil profiles) are evaluated along regular depth-slices by Gower's distance metric (Gower 1971), using any combination of continuous, categorical, or boolean attributes (Fig. 11.9). Total pair-wise dissimilarity is computed by taking the sum of slice-wise dissimilarities, to a user-defined depth. Variation in profile depth is accounted for by assigning maximum slice-wise dissimilarity to comparisons between soil and non-soil. Further customization of the NCSP algorithm is described in Beaudette et al. (2013b). The resulting dissimilarity matrix can be used to assist with topics ranging from initial mapping ("similar/dissimilar" soils), comparisons below family-level classification in Soil Taxonomy, soil series correlation, map unit harmonization, and correlation between different taxonomic systems.

Applied to the same set of soil profiles highlighted in Fig. 11.7, the NCSP algorithm was used to generate a distance matrix using only the GHL classes (ordinal values) to a depth of 175cm (Fig. 11.10). A dendrogram was created from the distance matrix via divisive hierarchical clustering (Kaufman and Rousseeuw 2005) and combined with profile sketches with the `plotProfileDendrogram()` function from the *sharpshootR* package. Profiles with similar GHL assignments, occurring at similar depths, are allocated to clusters defined by branching near the bottom of the dendrogram. When a combination of site (e.g., slope, drainage class, geoform, etc.) and horizon-level properties are requested, the final distance matrix is developed from a weighted average of the site and horizon-level distance matrices. Pair-wise distances between soil profiles can be difficult to interpret when a large number of properties are included in the calculation and may require a different approach to thematic coloring of profile sketches such as principal component analysis of the property matrix or principal coordinate analysis of the distance matrix scores. An alternative presentation of the data is possible by arranging profile sketches (with `plotSPC()`) according to the new axes created by 2-dimensional ordination, typically via non-Metric Multidimensional Scaling (nMDS) of the distance matrix.



**Fig. 11.9** Graphical outline of the Numerical Classification of Soil Profiles (NCSP) algorithm. (Figure c/o Jon Maynard, adapted from Maynard et al. (2020))



**Fig. 11.10** Soil profile sketches from Fig. 11.7, arranged according to divisive hierarchical clustering of the distance matrix generated by the Numerical Classification of Soil Profiles (NCSP) algorithm

### 11.5 Water Balance

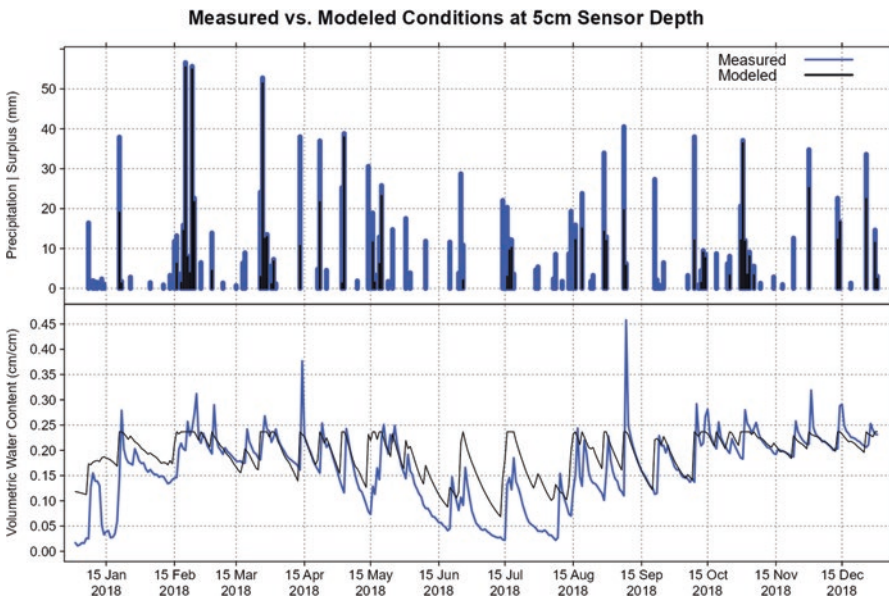
The soil forming state factor of climate and the interactions and timing of moisture and temperature are pivotal in understanding soil formation and describing site dynamics at local and regional scales. Water balance models accounting for inputs of precipitation and losses to evapotranspiration (Thorntwaite 1948) have evolved as a valuable tool for exploring the nuances of climate at a given point on the landscape. Water balance metrics relate to soil storage and the downward and upward flux of water through soil and associated soil property development as the soil acts as a sponge responding to atmospheric supply and demand (Arkley and Ulrich 1962). Correlation of water balance metrics to vegetation growth at sites has become an important tool for the study of site conditions related to existing vegetation distribution (Stephenson 1998; Lutz et al. 2010) and forecasting site climate trajectories. High quality, widely accessible gridded climate data has increased the use of soil water balance models.

The *sharpshootR* package provides methods (via dependencies on the *elevatr*, *daymetr*, *Evapotranspiration*, and *hyrdomad* packages) for calculating water balance variables of precipitation (PPT), potential evapotranspiration (PET), actual evapotranspiration (AET), deficit (D), soil moisture storage (S), surplus (U), volumetric water content (VWC) on monthly and daily time steps (monthlyWB() and dailyWB() functions). The prepareDailyClimateData() function assembles the available water-holding capacity (AWC) values derived for major components in soil map units of the US Soil Survey for specific point coordinates. Gridded DAYMET climate data (Thornton et al. 2020) is then downloaded for the specified location and daily water balance metrics are estimated via dailyWB\_SSURGO().

The *soilDB* package provides query access to the station data in the USDA-NRCS Soil Climate Analysis Network (SCAN). These stations provide above and below ground climate sensor networks that measure soil temperature, soil moisture, air temperature and precipitation. Many stations have associated soil characterization data in the Kellogg Soil Survey Laboratory (KSSL) records. The data in Fig. 11.11 were assembled through separate calls to `fetchSCAN()` for the SCAN climate station data and `dailyWB_SSURGO()` which derives AWC for the Scholten soil component sampled at SCAN station 2194, assembles DAYMET data for the SCAN station location and runs a daily water balance model. Comparisons of modeled and measured values allow for evaluation of water balance model utility and function.

### 11.6 Conclusions

Pedology and geomorphology are inherently visual, field-based sciences that share a common paradigm and fundamental units of description. Geomorphic description of landforms or geofoms and a merging of these disciplines functionally defined as geopedology is a progression that elevates the geofom as a primary landscape concept that can guide the operational inventory and integrated study of soil geomorphic relationships. Understanding the relevance and importance of contextual



**Fig. 11.11** Comparison of annual water balance metrics (volumetric water content and precipitation) for 2018 at Soil Climate Analysis Network (SCAN) station 2194 in the Ozark Highlands

linkages to geomorphology in soil survey products integrates soil information with other environmental data and is critical to informing the public and the wider scientific community. Embracing a visual, quantitative analytical approach to collections of soil profiles and varying formats of soil survey data allows for creative abstraction of geopedologic concepts.

Complex data structures and increasing volumes of available data demand progress in methods that allow iterative aggregation, summary, and graphical expression of soil data. The AQP suite of packages works to provide examples and routines that meet these challenges with an emphasis on generalized methods that can be applied to common data structures. This collection of tools leverages the open-source flexibility and extensibility of the **R** programming environment and the progress that can be gained from collaborative effort to solve common challenges in working with soil data. Providing visual alternatives for viewing soil morphologic data and improving the quality and accessibility of soil survey data will foster greater use and application of soil survey data to inform all users.

## References

- Arkley RJ (1971) Factor analysis and numerical taxonomy of soils. *Soil Sci Soc Am J* 35(2):312–315. <https://doi.org/10.2136/sssaj1971.03615995003500020038x>
- Arkley RJ (1976) Statistical methods in soil classification research. *Adv Agron* 28:37–70. [https://doi.org/10.1016/S0065-2113\(08\)60552-0](https://doi.org/10.1016/S0065-2113(08)60552-0)
- Arkley RJ, Ulrich R (1962) The use of calculated actual and potential evapotranspiration for estimating potential plant growth. *Hilgardia* 32(10):443–462
- Barron V, Torrent J (1986) Use of the Kubelka-Munk theory to study the influence of iron oxides on soil colour. *J Soil Sci* 37(4):499–510. <https://doi.org/10.1111/j.1365-2389.1986.tb00382.x>
- Beaudette DE (2022) Introduction to soil profile collection objects. <http://ncss-tech.github.io/AQP/aqp/aqp-intro.html>
- Beaudette DE, Dahlgren RA, O’Geen AT (2013a) Terrain-shape indices for modeling soil moisture dynamics. *Soil Sci Soc Am J* 77:1696–1710. <https://doi.org/10.2136/sssaj2013.02.0048>
- Beaudette DE, Roudier P, O’Geen AT (2013b) Algorithms for quantitative pedology: a toolkit for soil scientists. *Comput Geosci* 52:258–268
- Beaudette DE, Roudier P, Skovlin J (2016) Probabilistic representation of genetic soil horizons. In: Hartemink AE, Minasny B (eds) *Digital soil morphometrics*. Springer, pp 281–293. [https://doi.org/10.1007/978-3-319-28295-4\\_18](https://doi.org/10.1007/978-3-319-28295-4_18)
- Bishop TFA, McBratney AB, Laslett GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91(1-2):27–45
- Brown DJ, Clayton MK, McSweeney K (2004) Potential Terrain controls on soil color, texture contrast and grain-size deposition for the original Catena landscape in Uganda. *Geoderma* 122(1):51–72. <https://doi.org/10.1016/j.geoderma.2003.12.004>
- Buntley GJ, Westin FC (1965) A comparative study of developmental color in a Chestnut-Chernozem-Brunizem Soil Climosequence. *Soil Sci Soc Am J* 29(5):579–582. <https://doi.org/10.2136/sssaj1965.03615995002900050029x>
- Carré F, Jacobson M (2009) Numerical classification of soil profile data using distance metrics. *Geoderma* 148:336–345. <https://doi.org/10.1016/j.geoderma.2008.11.008>
- Centore P (2012) An open-source inversion algorithm for the munsell renotation. *Color Res Appl* 37:455–464



- Chesworth W, Arbestain MC, Macías F, Spaargaren O, Otto S, Mualem Y, Morel-Seytoux HJ et al (2008) Classification of soils: world reference base (WRB) for soil resources. In: Chesworth W (ed) *Encyclopedia of soil science*. Springer, Dordrecht, pp 120–122. [https://doi.org/10.1007/978-1-4020-3995-9\\_104](https://doi.org/10.1007/978-1-4020-3995-9_104)
- Dale MB, McBratney AB, Russell JS (1989) On the role of expert systems and numerical taxonomy in soil classification. *J Soil Sci* 40(2):223–234. <https://doi.org/10.1111/j.1365-2389.1989.tb01268.x>
- Dowle M, Srinivasan A (2021) data.table: extension of ‘data.frame’. <https://CRAN.R-project.org/package=data.table>
- Gama J, Centore P, Davis G (2021) Munsellinterpol: Interpolate Munsell Renotation Data from Hue/Chroma to CIE/RGB. <https://CRAN.R-project.org/package=munsellinterpol>
- Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27(4):857–871. <http://www.jstor.org/stable/2528823>
- Harden JW (1982) A quantitative index of soil development from field descriptions: examples from a Chronosequence in Central California. *Geoderma* 28(1):1–28. [https://doi.org/10.1016/0016-7061\(82\)90037-4](https://doi.org/10.1016/0016-7061(82)90037-4)
- Harradine F (1963) Morphology and genesis of noncalcareous brown soils in California. *Soil Sci* 96:277–287
- Hijmans RJ (2021) Raster: geographic data analysis and modeling. <https://CRAN.R-project.org/package=raster>
- Hole FD, Hironaka M (1960) An experiment in ordination of some soil profiles. *Proc Soil Sci Soc Am* 24:309–312
- Hudson BD (1992) The soil survey as paradigm-based science. *Soil Sci Soc Am J* 56:836–841
- Hurst VJ (1977) Visual estimation of iron in saprolite. *GSA Bull* 88(2):174–176. [https://doi.org/10.1130/0016-7606\(1977\)88<174:VEOIIIS>2.0.CO;2](https://doi.org/10.1130/0016-7606(1977)88<174:VEOIIIS>2.0.CO;2)
- Jenny H (1941) *Factors of soil formation: a system of quantitative pedology*. McGraw-Hill, New York
- Kabrick JM, Dey DC, Jensen RG, Wallendorf M (2008) The role of environmental factors in Oak decline and mortality in the Ozark Highlands. *For Ecol Manag* 255:1409–1417
- Kabrick JM, Dey DC, Jensen RG, Wallendorf M (2011) Landscape determinants of exchangeable calcium and magnesium in Ozark Highland Forest Soils. *Soil Sci Soc Am J* 75:164–180
- Kaufman L, Rousseeuw PJ (eds) (2005) *Finding groups in data*, Wiley series in probability and statistics. Wiley, Hoboken. <https://doi.org/10.1002/9780470316801>
- Liles GC, Beaudette DE, O’Geen AT, Horwath WR (2013) Developing predictive soil C models for soils using quantitative color measurements. *Soil Sci Soc Am J* 77:2173–2181. <https://doi.org/10.2136/sssaj2013.02.0057>
- Lutz JA, Franklin JF, Van Wagendonk JW (2010) Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. *J Biogeogr* 37(5):936–950
- Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2021) *Cluster: cluster analysis basics and extensions*. <https://CRAN.R-project.org/package=cluster>
- Marcus RT (1998) The measurement of color. In: Nassau K (ed) *Color for science, art and technology*. Elsevier, pp 31–96
- Maynard JJ, Salley SW, Beaudette DE, Herrick JE (2020) Numerical soil classification supports soil identification by citizen scientists using limited, simple soil observations. *Soil Sci Soc Am J* 84(5):1675–1692. <https://doi.org/10.1002/saj2.20119>
- Moeys J (2018) Soiltexture: functions for soil texture plot, classification and transformation. <https://CRAN.R-project.org/package=soiltexture>
- Moore AW, Russell JS, Ward WT (1972) Numerical analysis of soils: a comparison of three soil profile models with field classification. *J Soil Sci* 23:194–209
- Müller K, Wickham H (2021) *Tibble: simple data frames*. <https://CRAN.R-project.org/package=tibble>
- Munsell AH (1947) *A color notation*, 10th edn. Munsell Color Company, Inc



- Myers DB, Kitchen NR, Sudduth KA, Miles RJ, Sadler EJ, Grunwald S (2011) Peak functions for modeling high resolution soil profile data. *Geoderma* 166(1):74–83. <https://doi.org/10.1016/j.geoderma.2011.07.014>
- O’Brien L (2022) Mpspline2: mass-preserving spline functions for soil data. <https://CRAN.R-project.org/package=mpspline2>
- O’Geen A, Walkinshaw M, Beaudette DE (2017) SoilWeb: a multifaceted interface to soil survey information. *Soil Sci Soc Am J* 81. <https://doi.org/10.2136/sssaj2016.11.0386n>
- Pebesma E (2018) Simple features for R: standardized support for spatial vector data. *R J* 10(1):439–446. <https://doi.org/10.32614/RJ-2018-009>
- Pedersen TL, Nicolae B, François R (2021) Farver: high performance colour space manipulation. <https://CRAN.R-project.org/package=farver>
- R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>
- Rayner JH (1966) Classification of soils by numerical methods. *J Soil Sci* 17:79–92
- Roecker S, Skovlin J, Beaudette D, Wills S (2016) Digital summaries of pedon descriptions. In: Hartemink AE, Minasny B (eds) *Digital soil morphometrics*. Springer, pp 267–279. [https://doi.org/10.1007/978-3-319-28295-4\\_17](https://doi.org/10.1007/978-3-319-28295-4_17)
- Sarkar PK, Bidwell OW, Marcus LF (1966) Selection of characteristics for numerical classification of soils. *Soil Sci Soc Am Proc* 30:269–272
- Scheinost AC, Schwertmann U (1999) Color identification of iron oxides and hydroxysulfates: use and limitations. *Soil Sci Soc Am J* 63(5):1463–1471. <http://soil.scijournals.org/cgi/content/abstract/soilsci;63/5/1463>
- Schoeneberger PJ, Wysocki DA, Benham EC, Soil Science Division Staff (2012) *Field book for describing and sampling soils, Version 3.0*. Natural Resources Conservation Service, National Soil Survey Center
- Schwertmann U (1993) Relations between iron oxides, soil color, and soil formation. In: Bigham JM, Ciolkosz EJ (eds) *Soil color, SSSA special publication number 31*. Soil Science Society of America, Inc., Madison, pp 51–69
- Sharma G, Wencheng W, Dalal EN (2005) The CIE DE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations. *Color Res Appl* 30(1):21–30. <https://doi.org/10.1002/col.20070>
- Simonson RW (1993) Soil color standards and terms for field use—history of their development. In: Bigham JM, Ciolkosz EJ (eds) *Soil color, SSSA special publication number 31*. Soil Science Society of America, Inc., Madison, pp 1–20
- Singh G, Goyne KW, Kabrick JM (2015) Determinants of total and available phosphorus in forested alfisols and ultisols of the Ozark Highlands, USA. *Geoderma Reg* 5:117–126
- Sneath PHA, Sokal RR (1973) *Numerical taxonomy*. W.H. Freeman Company
- Soil Science Division Staff (2017) *Soil survey manual*. Ditzler C, Scheffe K, Monger HC (eds). USDA Handbook 18. Government Printing Office
- Soil Survey Staff (1999) *Soil Taxonomy: a basic system of soil classification for making and interpreting soil surveys*, U.S. Department of Agriculture Handbook 436, 2nd edn. Natural Resources Conservation Service, Washington, DC
- Soil Survey Staff (2022a) *Online Manual for the AQP Package*. <http://ncss-tech.github.io/aqp/>
- Soil Survey Staff (2022b) *Soil Survey Geographic (SSURGO) Database*. Edited by Natural Resources Conservation Service, United States Department of Agriculture. <https://sdmdataaccess.sc.gov.usda.gov/>
- Soil Survey Staff (2022c) *Soil Survey Technical Note 2*. Edited by Natural Resources Conservation Service, United States Department of Agriculture. [https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2\\_053569](https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_053569)
- Stephenson NL (1998) Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. *J Biogeogr* 25(5):855–870

- Thompson JA, Bell JC (1996) Color index for identifying hydric conditions for seasonally saturated Mollisols in Minnesota. *Soil Sci Soc Am J* 60(6):1979–1988. <https://doi.org/10.2136/sssaj1996.03615995006000060051x>
- Thornthwaite CW (1948) An approach toward a rational classification of climate. *Geogr Rev* 38(1):55–94. <http://www.jstor.org/stable/210739>
- Thornton MM, Shrestha R, Wei Y, Thornton PE, Kao S, Wilson BE (2020) DAYMET: daily surface weather data on a 1-Km grid for North America, Version 4. ORNL Distributed Active Archive Center. <https://doi.org/10.3334/ORNLDAAC/1840>
- van den Boogaart KG, Tolosana-Delgado R, Bren M (2021) Compositions: compositional data analysis. <https://CRAN.R-project.org/package=compositions>
- Viscarra Rossel RA, Minasny B, Roudier P, McBratney AB (2006) Colour space models for soil science. *Geoderma* 133(3):320–337. <https://doi.org/10.1016/j.geoderma.2005.07.017>
- Wagenet RJ, Bouma J, Grossman RB (1991) Minimum data sets for use of soil survey information in soil interpretive models. In: Mausbach MJ, Wilding LP (eds) *Spatial variabilities of soils and landforms*, SSSA Spec. Publ. 28. SSSA, Madison, pp 161–182
- Wills SA, Lee Burras C, Sandor JA (2007) Prediction of soil organic carbon content using field and laboratory measurements of soil color. *Soil Sci Soc Am J* 71(2):380–388. <https://doi.org/10.2136/sssaj2005.0384>
- Wysocki AD, Schoeneberger PJ, LaGarry HE (2005) Soil surveys: a window to the subsurface. *Geoderma* 126(1):167–180
- Zeileis A, Fisher JC, Hornik K, Ihaka R, McWhite CD, Murrel, I P., Stauffer, R., and Wilke, C.O. (2020) Colorspace: a toolbox for manipulating and assessing colors and palettes. *J Stat Softw* 96:1–49
- Zinck JA (2016) The geopedologic approach. In: Metternicht G, Zinck JA, Del Valle HF (eds) *Geopedology*. Springer, pp 27–59. [https://doi.org/10.1007/978-3-319-19159-1\\_4](https://doi.org/10.1007/978-3-319-19159-1_4)