**PEDOLOGY**

Soil Science Society of America Journal

# Numerical soil classification supports soil identification by citizen scientists using limited, simple soil observations

Jonathan J. Maynard[1] ⓘ | Shawn W. Salley[2] ⓘ | Dylan E. Beaudette[3] |
Jeffery E. Herrick[2]

[1] University of Colorado Boulder, Boulder, CO 80309, USA

[2] USDA-Agricultural Resource Service, Jornada Experimental Range, Las Cruces, NM 88003, USA

[3] USDA-NRCS, Soil Science Division Sonora, Sonora, CA 95370, USA

**Correspondence**
Jonathan Maynard, University of Colorado Boulder, Boulder, CO 8030,9 USA.
email: jonathan.maynard@colorado.edu

**Abstract**

Accurately identifying the soil map unit component at a specific point-location within a landscape is critical for implementing sustainable soil management. Recent developments in smartphone-based technologies for characterizing soil profiles, coupled with improved numerical soil classification algorithms, have made it more accessible for non-soil scientists to sample, characterize, and classify soil profiles. The main objective of this study was to evaluate an operational soil classification framework for identifying the soil component at a sampling-location based on the numerical similarity of soil property values between the sampled soil profile and the soil components mapped in that area. To evaluate this soil identification framework, we used a subset of the U.S. National Cooperative Soil Survey Soil Characterization Database (NCSS–SCD) as our soil profile test dataset and the U.S. Soil Survey Geographic (SSURGO) database as our reference dataset using profile data of soil components in the area surrounding each test profile. Numerical similarity was tested using soil property data representing different degrees of generalization, both in terms of generalizing depth-wise variability (i.e., depth-support) and generalizing across feature space (i.e., soil properties). Three soil property groups (i.e., Novice, Expert, Expert-Plus) representing different levels of detail and three types of depth-support (i.e., genetic horizon, depth intervals, and depth functions) were evaluated. Using a simple set of soil property inputs (i.e., Novice: soil texture class, rock fragment volume class, and soil color) resulted in nearly as high identification accuracy (46–53%) as that achieved with an Expert (48–57%) dataset that included more precise determinations (percent sand, silt, clay, and rock fragment volume), and virtually no further improvement with the addition of pH and organic matter in the Expert-Plus dataset (53–60%). This study also showed minimal effect from the type of depth-support used to represent depth-wise variability. Furthermore, we evaluated several measures of soil functional similarity (i.e., ecological sites, land capability, taxonomic distance) which resulted in management relevant accuracies ranging from 65–89%. These findings support the utility of simple soil observations sampled at fixed depths for soil identification.

# 1 | INTRODUCTION

Soil taxa (e.g., soil series) are information carriers that allow land managers to infer a general range of soil behavior in response to management actions and disturbance effects. Accurate identification of the soil taxa at specific locations within a landscape is therefore critical for implementing sustainable soil management. Soils are inherently complex systems that exhibit a high degree of variability, both in geographic and feature space (i.e., soil property values), making the identification of a soil taxa at an unsampled location extremely difficult. This is exemplified by the current suite of class-based soil maps available to end-users (e.g., SSURGO/STATSGO, HWSD) where individual soil map units often contain multiple soil taxa (i.e., soil series/series-phases), commonly referred to as soil map unit components. However, even at locations where a soil's properties are known, its correct identification to an existing soil taxa or series can be challenging. Within a soil map unit, the management requirements for co-occurring soil components can be drastically different, resulting in damaging impacts to the soil resource when incorrectly identified and improperly managed as well as missed opportunities to increase production or switch to more profitable production systems (Duniway, Bestelmeyer, & Tugel, 2010).

Soil classification systems, such as Soil Taxonomy (Soil Survey Staff, 1999) and the World Reference Base (IUSS - WRB, 2015), provide a hierarchical framework for grouping soils based on our current knowledge of soil genesis and its effects on land-use potential. While these systems are highly effective at classifying soils based on a set range of depth-dependent soil property values and conditions, they do not provide a means to quantitatively describe differences between soils (Beaudette, Roudier, & O'Geen, 2013). The ability to quantitatively describe differences between soil components may provide a means of correctly identifying the soil component at a sampling location based on a numerical comparison of soil property values between the sampled soil profile and the soil components mapped in that area. Furthermore, the development of a numerical soil classification system could provide a framework for non-soil scientists (e.g., citizen scientists) to identify the soil component at unknown sampling locations, contingent on their ability to accurately measure the soil properties being compared.

Numerical soil classification encompasses two types of activities: arranging individual soil profiles in classes ('classification'), and assigning an individual profile to an existing class ('identification') (de Gruijter, 1977). Both types of activities employ numerical methods which typically involve the calculation of a pair-wise dissimilarity between soil profiles using a feature-space distance metric (e.g., Euclidean distance). Soil profiles that are

**Core Ideas**

- Evaluates operational modeling framework for location-based soil identification
- Generalized soil inputs produced similar identification accuracy as detailed inputs
- Type of depth-wise soil sampling had minimal effect on simulated soil identification results
- Supports utility of simple soil observations sampled at fixed depths by citizen scientists

similar will be closer in feature-space distance than those that are less similar (Carré & Jacobson, 2009). Several algorithms for numerical soil classification have been proposed over the last several decades based on a comparison of: (1) genetic horizons (Carré & Jacobson, 2009; Rayner, 1966), (2) standardized depth intervals or depth slices (Beaudette et al., 2013; Fan et al., 2018; Moore, Russell, & Ward, 1972), (3) soil property depth functions (e.g., polynomial coefficients) (Moore et al., 1972), and (4) transition matrices based on the global clustering of horizon samples into a small number of horizon classes (Little & Ross, 1985; Moore et al., 1972; Norris & Dale, 1971).

Recently, Beaudette et al. (2013) identified two major limitations of previous soil classification algorithms: (1) the lack of a method to account for comparisons between soil and non-soil material (i.e., deep vs. shallow profiles), and (2) lack of support for binary, nominal, and ordinal scale variables. To overcome these limitations, Beaudette et al. (2013) devised a new numerical soil classification algorithm, Numerical Comparison (of) Soil Profiles (NCSP) ('*NCSP*' function from the 'aqp' R package), where the between-profile dissimilarity is evaluated along regular depth slices (e.g., every slice or every *nth* slice). The treatment of sample depth as the basis for between-sample comparisons simplifies the process of comparing and classifying soils that have varying horizonation (e.g., genetic horizons) or that were collected using different sampling methods (e.g., genetic horizons vs soil depth intervals). Each slice-wise dissimilarity is calculated using Gower's generalized dissimilarity metric (Gower, 1971) which can accommodate any combination of binary, nominal, ordinal, ratio-scale, or continuous variables. Additionally, the algorithm accounts for comparisons between soil and non-soil material (which would result in an undefined dissimilarity) via replacement with the maximum between-slice dissimilarity from the collection of profiles being compared. The final between-profile dissimilarity is calculated by summing the collection of slice-wise dissimilarity matrices (Beaudette et al., 2013). The classification

output of the *NCSP* algorithm is influenced by both the sampling method used to describe the soil's vertical variability (i.e., depth-support: genetic horizons vs. depth intervals vs. depth functions) and the selection of soil properties used to compute the between-sample dissimilarity.

The first step in identifying an unknown soil is to describe, sample, and characterize its soil profile. The accurate characterization of soil properties and their intrinsic variability has historically required a level of training and experience only possessed by soil scientists (Rossiter, Liu, Carlisle, & Zhu, 2015; Salley et al., 2018). Recent efforts to simplify and standardize the process of soil profile characterization, however, have made it easier for non-soil scientists to collect accurate soil property data that can be used for soil identification. For example, the Land Potential Knowledge System (LandPKS) mobile app provides a complete digital interface for collecting and recording a small subset of standard morphologic data at standardized soil depths (Herrick et al., 2013, 2016, 2017). The current LandPKS app provides tools for estimating soil texture classes using a digital key that guides users through a series of texture-by-feel tests (Salley et al., 2018). It also allows users to measure Munsell soil color using the smartphone's camera and an external reference (e.g., post-it note) (Fan et al., 2017). With growing interest in the collection of soil data by non-soil scientists, there is a need to better understand how different levels of soil property generalization, with respect to the characterization of soil variability, affect identification accuracy. Specifically, there is a need to understand how the characterization of soil variability both with depth (i.e., depth-support) and in feature space (e.g., percent clay vs. texture class) affect our ability to correctly classify a soil. For example, a field scientist with little-to-no soil training would not be able to accurately describe a soil profile by genetic horizons or accurately determine the clay and sand percentages at each horizon. However, non-soil scientists can sample soil profiles at standardized depth intervals and estimate soil texture classes using tools like the LandPKS app.

With the proliferation of digital technologies an enormous quantity of soil data has been digitized, creating many national and global soil profile databases and digital soil maps (Brevik et al., 2015). The main objective of this study was to evaluate the potential accuracy of the *NCSP* algorithm using a simulated operational testing framework for matching the soil at a point location based on its numerical similarity to the soil components mapped and digitally available in that area. To implement this testing framework, we used a subset of the United States National Cooperative Soil Survey Soil Characterization Database (NCSS–SCD) as our simulated soil profile test dataset (i.e., measured soil profiles at point-locations) and the U.S. Soil Survey Geographic (SSURGO) database

as our soil profile reference library used for matching (i.e., mapped soil components in the area surrounding each test profile). Since the NCSS-SCD dataset was collected by soil scientists, this testing framework represents a simulation of the potential accuracy achievable by the citizen scientist when they have accurately measured their soil profile data. Furthermore, within this simulated testing framework, this study evaluated how soil property generalization across both depth and feature space affect the identification output. Specifically, this study evaluated three different forms of depth-support (i.e., genetic horizons, standardized depth intervals, and derived depth functions) and three soil property groups representing different degrees of generalization associated with the different levels of experience required for their measurement (e.g., texture class vs. percent sand, silt, and clay).

## 2 | METHODS

### 2.1 | Numerical soil identification framework

The numerical soil identification testing framework, described in detail in Figure 1, is comprised of four steps: (1) process and subset NCSS–SCD soil pedon data, (2) query SSURGO database for soil map unit components mapped in the area surrounding each NCSS–SCD soil pedon based on spatial filters, (3) resampling of soil profile data to different forms of depth-support, and (4) run the *NCSP* algorithm on soil profiles with different generalized datasets and depth-wise distributions and evaluate match results based on the soil component and several measures of soil functional similarity (i.e., ecological sites, land capability classification, and taxonomic distance).

### 2.2 | NCSS–SCD pedon data

Soil pedon data from NCSS–SCD were used as our simulated soil profile test dataset. NCSS–SCD is a comprehensive database of soil profile characterization data analyzed by the Kellogg Soil Survey Laboratory and cooperating laboratories. The NCSS–SCD contains data for more than 64,000 soil pedons collected from the 1950s to the present. NCSS-SCD pedons are targeted samples designed to represent the central concept of a soil series, the central concept of a soil map unit (SMU) component, and the range of properties within a series or landscape. NCSS-SCD also includes pedons sampled for different research projects (e.g., Rapid Carbon Assessment, NEON). NCSS–SCD data were queried from a Microsoft Access database (http://ncsslabdatamart.sc.egov.usda.gov/, accessed 14 Sept. 2018).
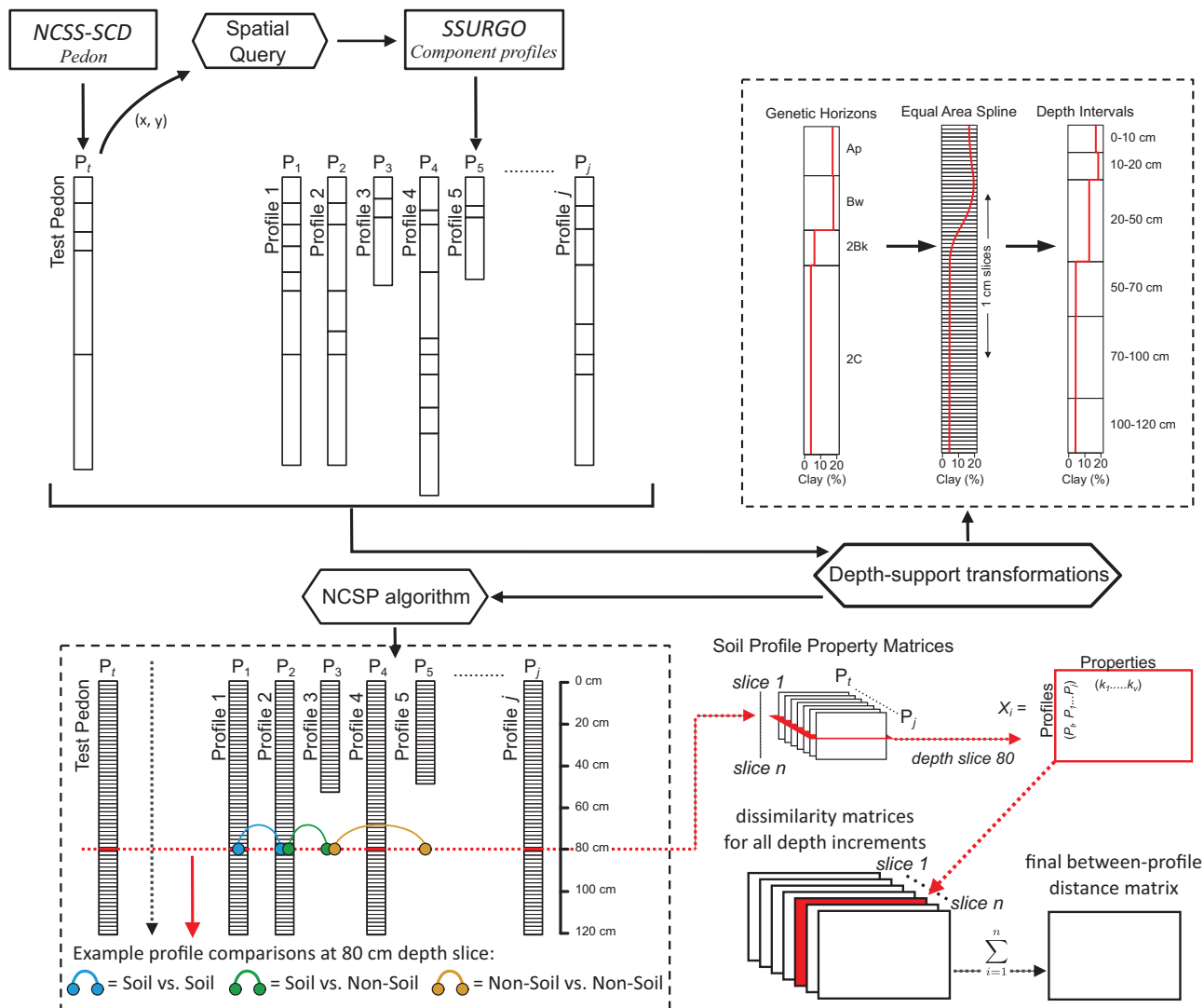
**FIGURE 1** Conceptual diagram of the numerical soil identification framework

Field-collected soil morphology data and records of the taxonomic history (e.g., sampling data, correlated taxonomic class, taxonomic updates) for each pedon are recorded and stored in the Natural Resources Conservation Service (NRCS) National Soil Information System (NASIS) database (NASIS database snapshot acquired 5 April 2017). Each NCSS–SCD pedon has been classified to a soil series class based on both the field described morphology data and laboratory analysis. A subset of 6,220 soil pedons were selected from the NCSS–SCD database based on the following filtering criteria: pedons must have (1) no missing horizons and have internally consistent depths (e.g., no overlapping horizon depths); (2) complete soil horizon property data for the following properties: percent sand, percent silt, percent clay, rock fragment volume (rfv), Munsell color-moist, pH, and percent organic matter (om); and (3) a correlated soil series name (i.e., *taxonname*) matching at

least one of the SSURGO component names returned from the spatial query. Soil color, spatial coordinates and taxonomy information were extracted from NASIS for each point and joined to NCSS–SCD. The spatial distribution of soil sampling points across the U.S. is presented in Figure 2.

## 2.3 | SSURGO component classes

The SSURGO database contains detailed soil information available for most areas of the U.S., as well as the Territories, Commonwealths, and Island Nations served by the USDA-NRCS, and is derived from detailed land surveys and hundreds of thousands of soil observations (e.g., profiles, augerings, partial excavations) (Hudson, 1992). The delineation of SSURGO map units is based on tacit soil-landscape and hillslope models and the attribution of
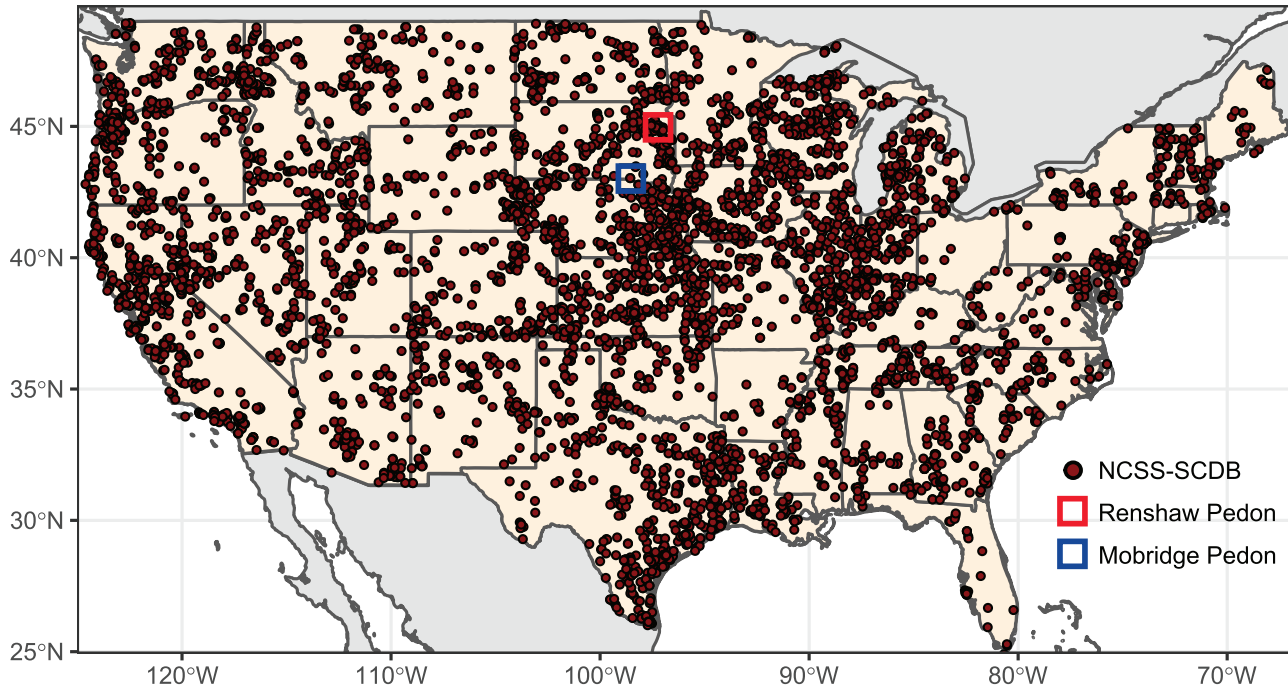
**FIGURE 2** Map of the 6,220 NCSS–SCD soil pedon locations evaluated in this study. The red and blue boxes indicate locations of the Renshaw and Mobridge example pedons (see Figure 4), respectively

those map units based on a limited number of representative pedons sampled within each soil map unit (Libohova et al., 2013, 2019). Consequently, SSURGO SMUs describe the modal concept of each soil class, reporting information aggregated from the collection of field-described soil observations hypothesized to represent the soil class concept. Soil properties that cannot be measured in the field are estimated by pedotransfer functions in NASIS, or occasionally informed by NCSS-SCD lab data when available. Since some NCSS-SCD pedons are sampled to represent the central concept of a SMU component, it is possible that in some comparisons the SSURGO SMU component data could be a direct copy of the NCSS-SCD pedon data. Analysis of the statistical distance between the two data sources, however, showed that for most comparisons the distances were relatively high (median distance = 37; Fig. S1-g). This suggests that in most cases the SMU component data is not a direct copy of the NCSS-SCD pedon data (distance = 0 if data sources are the same).

The SSURGO database contains over 300,000 SMU and over 1,000,000 SMU components and is mapped at scales typically ranging from 1:12,000 to 1:63,360. SMU delineation (e.g. polygon) area is controlled by the concept of a minimum legible area (MLA) which is the minimum ground area that is legible on a map (Forbes, Rossiter, & Van Wambeke, 1987). Agricultural lands have typically been surveyed at order 2 (MLA of 1–4 ha.), while rangeland, forest, and degraded landscapes have been surveyed at order 3 (MLA of 2–16 ha.) (Soil Survey Division Staff,

2017). Soil surveys in wilderness areas employ a mixed order 3/4 strategy, with more detailed mapping near riparian corridors or ecologically sensitive areas. The composition of SMU (e.g. SMU kind) are closely tied to the complexity of terrain contained by increasingly larger MLA: consociations (i.e., dominated by a single soil taxa) and complexes (i.e., two or more dissimilar soil taxa) at order 2, associations (i.e., two or more dissimilar soil taxa that could be separated at larger mapping scales) and complexes at order 3, and mostly associations at order 4 (Soil Survey Division Staff, 2017). Each SMU component contains a modal soil profile concept described by genetic horizons with representative soil property values assigned to each horizon. The majority of SSURGO components (approximately 88%) are mapped to the level of soil series or soil series-phase in U.S. soil taxonomy with approximately 24,000 unique soil series and approximately 29,000 unique soil series/series-phases within the U.S., and the Territories, Commonwealths, and Island Nations served by the USDA-NRCS. Because the SSURGO database does not include soil color estimates, we used soil color data from the NRCS Official Soil Series Descriptions (OSD) database. The OSD database contains a global collection of soil profile morphology descriptions (data sheets in narrative form) for the approximately 24,000 soil series.

Using the point location of each NCSS–SCD test pedon, we ran a spatial query on the SSURGO database to extract soil components within the intersecting soil map unit (i.e., home map unit) and all map units adjacent to the home
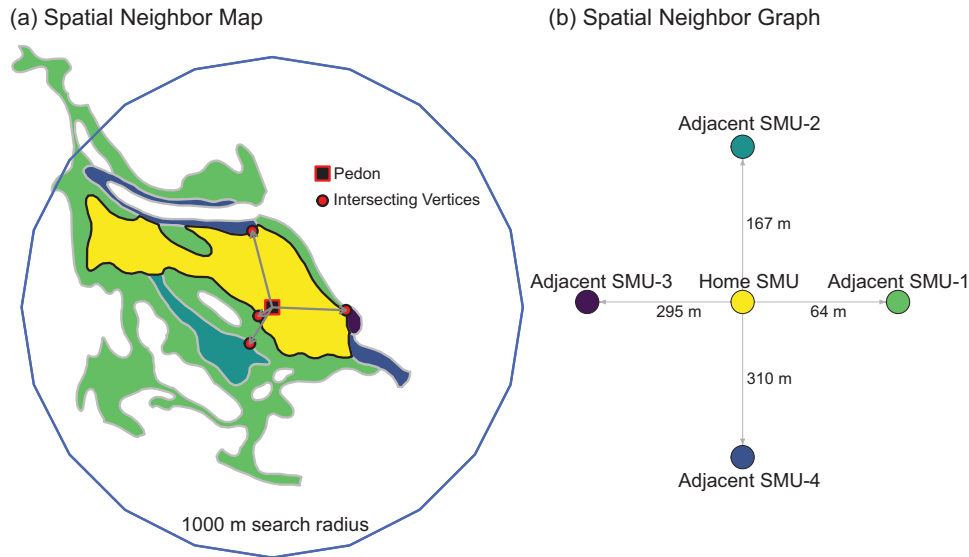
**F I G U R E 3** Schematic illustrating (a) the spatial query of SSURGO soil map units (SMU) that either intersect the NCSS–SCD pedon (i.e., Home SMU) or are adjacent to the Home SMU with an intersecting boundary within a 1 km search radius of the test pedon. The above example is from the Renshaw test pedon (see Figure 2 for geographic location and Figure 4a for queried SMU components) and illustrates the calculation of (a) distances between the test pedon and the closest intersecting vertices for all adjacent SMUs and (b) the distance relationships between the test pedon in the Home SMU and all adjacent SMUs

map unit whose closest intersecting boundary was within a 1 km search radius of the test pedon (Figure 3). We chose a 1 km search radius based on the range of MLAs for order 2 and 3 soil surveys (1:12,000 to 1:63,360) typically encountered in SSURGO. A soil map at 1:63,360 map scale has an MLA of 16 ha, which equates to a circle with a radius of 225 m. Order 4 soil maps (some wilderness areas within SSURGO) have map scales ranging from 1:63,360 to 1:250,000. At 1:250,000 map scale, a soil map has an MLA of 252 ha or circle with radius of 895 m. Thus, our 1 km search radius is approximately four time the MLA of a large-scale order 3 map and slightly greater than the MLA of a large-scale order 4 map. The requirement of spatial adjacency in our search criteria was based on a compromise between selecting a search radius suitable across a range of soil mapping orders (i.e., order 2–4) and the processing time required for spatial adjacency evaluation of the SSURGO polygons. SSURGO data were queried from NRCS Soil Data Access (http://sdmdataaccess.nrcs.usda.gov/, accessed on 16 Sept. 2019). For each queried SSURGO component, we extracted the OSD soil color data from the OSD matching the SSURGO component name (i.e., correlated soil series name). Due to differences in horizonation between OSD soil profiles and the SSURGO component instances of each soil series, we combined each SSURGO component profile with its matching OSD profile after slicing each profile into 1 cm increments. The merging of the two datasets along a common 1 cm depth basis minimized the number of processing steps and potential introduction of error, while also producing a format aligned with *NCSP's*

computation of slice-wise dissimilarity at each 1 cm increment. Horizonation differences were not an issue in the equal-area spline and depth interval datasets since separate equal area spline functions were modeled to each soil property.

## 2.4 | Soil property generalization

Two types of soil property generalization were evaluated: (1) different forms of depth-support and (2) different degrees of generalization in feature space. The vertical sampling of soil profiles typically employs one of three different types of depth-support: genetic horizons, standardized depth intervals, or continuous depth functions (e.g., sampling every 1 cm). Most soil classification systems characterize soil profiles by genetic horizons which are delineated using visual and tactile observations guided by current theories of soil development (IUSS - WRB, 2015; Soil Survey Staff, 1999). Sampling by either genetic horizon or soil depth interval transforms a theoretically continuous depth function into a stepped depth function, resulting in potential inaccuracies when attempting to predict soil property values at specific soil depths (Bishop, McBratney, & Laslett, 1999; Malone, McBratney, Minasny, & Laslett, 2009). To correct for this, many studies have used equal-area quadratic spline functions to model the 'true' depth-wise distribution of soil properties from profile data (Bishop et al., 1999; Malone et al., 2009; Odgers, McBratney, & Florence, 2018; Ponce-Hernandez,

**TABLE 1** Soil property groups based on user knowledge and experience (i.e., Novice, Expert, and Expert-Plus)

| | | | Soil property groups | | |
|---|---|---|---|---|---|
| Soil Property | Abrv. | Type | Novice | Expert | Expert-Plus |
| Soil color (CIE L*a*b*) | L, A, B | ratio | X | X | X |
| Rock fragment volume class | rfv-class | ordinal | X | X | X |
| Texture class | texture | ordinal | X | X | X |
| Rock fragment volume | rfv | ratio | | X | X |
| Clay | clay | ratio | | X | X |
| Sand | sand | ratio | | X | X |
| Silt | silt | ratio | | X | X |
| pH[a] | ph | ratio | | | X |
| Organic matter[a] | om | ratio | | | X |

[a] Require instruments for measurement

Marriott, & Beckett, 1986). This study used equal-area quadratic splines to transform soil profile data sampled by genetic horizon into two different soil profile datasets: (i) a sequence of standardized depth intervals following the depth breaks defined by LandPKS (i.e., 0–10, 10–20, 20–50, 50–70, 70–100, 100–120 cm), and (ii) continuous soil depth functions (i.e., predicted values at 1 cm increments) (Malone et al., 2009). The equal-area quadratic spline was implemented using the *ea-spline* function from the *ithir* R package (Malone, 2016). The *ea-spline* fits a mass preserving depth function where the original data is preserved and can be retrieved via integration of the continuous spline. The shape of the spline function (i.e., smoothness between input horizons) is controlled by a smoothing parameter, lambda ($\lambda$). We used lambda value of 0.1 which has been shown to provide reasonable results for most soil properties (Bishop et al., 1999). In addition to the three forms of depth-support, three soil property groups were identified based on the level of experience required for their measurement (i.e., Novice, Expert, and Expert-Plus soil property groups), with each group representing a different degree of generalization in feature space. Given the increased ease in measuring soil color and texture by non-soil scientists, as well as their general importance in U.S. Soil Taxonomy for differentiating and classifying soils, we selected soil texture class (derived from % sand, silt, clay), rock fragment volume class (i.e., 0–15%, 15–35%, 35–65%, and > 65%), and soil color (CIE L*a*b* coordinates) to represent our 'Novice' soil property dataset (Torrent & Barron, 1993). Our Expert and Expert-Plus datasets are composed of all properties in the Novice dataset plus additional soil properties that an expert soil scientist would be able to measure or estimate in the field. For the Expert dataset this included the addition of percent sand, silt, clay, and rock fragment volume. For the Expert-Plus dataset this included all properties in the Expert dataset plus the addition of percent organic matter and pH (Table 1). The combined effects of the different depth-wise distributions (i.e., genetic horizon, depth intervals, and depth functions) and soil property groups (i.e., Novice, Expert, and Expert-Plus properties) on soil profile identification accuracy were evaluated.

## 2.5 | Numerical classification of soil profiles

Numerical classifications between NCSS–SCD pedons and SSURGO soil map unit components were computed using the *NCSP* function from the *aqp* R package (Beaudette et al., 2013). At each point-location all queried soil profiles were segmented at 1 cm increments creating a soil property matrix for each profile, with rows representing depth slices and columns representing soil properties. At each depth slice the vector of soil property data is extracted from each soil profile property matrix, row-wise, and aggregated forming a new depth-slice matrix $X_i$:

$$X_i = \begin{vmatrix} X_{P_t,i,clay} & X_{P_t,i,sand} & X_{P_t,i,color} & \cdots & X_{P_t,i,k} \\ X_{P_1,i,clay} & X_{P_1,i,sand} & X_{P_1,i,color} & \cdots & X_{P_1,i,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{P_j,i,clay} & X_{P_j,i,sand} & X_{P_j,i,color} & \cdots & X_{P_j,i,k} \end{vmatrix}$$

where rows represent profiles ($P_j$) and columns represent properties ($k$) at depth slice $i$. Slice-wise comparisons were calculated at each 1 cm depth slice to a maximum depth of 120 cm using Gower's generalized dissimilarity metric (Gower, 1971). Gower's generalized dissimilarity calculates the dissimilarity (i.e., feature-space distance) between objects $t$ and $j$ by taking the average dissimilarity score over all possible k attribute comparisons:

$$S_{tj} = \frac{\sum_{k=1}^{v} S_{tjk} W_k}{\sum_{k=1}^{v} \delta_{tjk} W_k}$$ where $S_{tjk}$ is the contribution of variable $k$ to the similarity between objects $t$ and $j$, $W_k$ is the weighting factor for variable $k$, $\delta_{tjk}$ is an indicator of whether variable $k$ is present in both objects and thus

can be compared, and $v$ is the total number of variables. For continuous or ratio variables (e.g., % clay) $S_{tjk} = |X_{tk} - X_{jk}|/R_k$, where $X_{tk}$ and $X_{jk}$ are the values of variable $k$ for objects $t$ and $j$, respectively, and $R_k$ is the range of variable $k$ within the sample. For ordinal variables (e.g., texture class) all $X_{tk}$ are replaced by their ranks $r_{tk}$ determined over all objects, where $S_{tjk} = 0$ if the two objects agree on variable k, otherwise:

$$S_{tjk} = \frac{\left| r_{tk} - r_{jk} \right| - (T_{tk} - 1)/2 - (T_{jk} - 1)/2}{\max\{r_k\} - \min\{r_k\} - (T_{k,\,max} - 1)/2 - (T_{k,\,min} - 1)/2}.$$

where $T_{tk}$ is the number of objects which have the same rank score for variable k as object j; $T_{k,\,max}$ is the number of objects which have the maximum rank (i.e., $\{r_k\}$); and $T_{k,\,min}$ is the number of objects with the minimum rank (i.e., $\{r_k\}$), for variable k in the ordering (Podani, 1999). In the present application objects $t$ and $j$ represent depth slices from the NCSS–SCD test pedon and a SSURGO component profile, respectively. $S_{tj}$ is calculated between the NCSS–SCD test pedon and each SSURGO component profile producing a $(1 \times j)$ dissimilarity matrix $(D_i)$ at each depth slice. The pair-wise dissimilarity $D$ between the NCSS–SCD pedon and each SSURGO component profile is computed as the sum of the slice-wise dissimilarity:

$$D = \sum_{i=1}^{n} w_i D_i$$

where $n$ is the total number of slices and $w_i$ is an optional weighting coefficient. Dissimilarity values in the final pair-wise dissimilarity matrix were rescaled by dividing by the maximum dissimilarity and multiplied by 100, resulting in final dissimilarity values ranging between 0 and 100. No depth weighting or property weighting were performed, and undefined dissimilarities were retained for all non-soil (i.e., bedrock) vs. non-soil comparisons and replaced with the maximum between-slice dissimilarity or all soil vs. non-soil comparisons. This ensured that the dissimilarity between a slice of soil in one profile and a slice on non-soil in another profile is maximized (e.g., Profile 2: Bt horizon at 80 cm vs. Profile 3: R horizon at 80 cm; Figure 1), reflecting the fact that they are highly different. Similarly, in profile collections with both deep and shallow soils, when two shallow soils are compared at depths below their maximum profile depths (e.g., Profile 3 vs Profile 5 at 80 cm; Figure 1), undefined dissimilarities are retained to reflect the fact that no soil is being compared. All dissimilarity values calculated between the NCSS–SCD pedon and each SSURGO soil map unit component were extracted from the dissimilarity matrix. Since the SSURGO spatial query can include multiple soil map units (i.e., Home + adjacent within 1000 m; Figure 3), the occurrence of mul-

tiple instances of the same component class is common. In our ranking evaluation, the lowest dissimilarity value from each unique component class was used in the final ranking, and among these final values the component class with lowest dissimilarity value was used to evaluate the soil component match rate (i.e., match rate between top ranking SSURGO component series name and NCSS-SDC pedon series name).

To better account for functional similarity between the predicted component (i.e., top ranking component) and correct component (i.e., SSURGO component series name that matches the NCSS-SDC pedon series name) we evaluated match rates based on NRCS Land Capability Classification (LCC) (Klingebiel and Montgomery, 1966), NRCS Ecological Site Descriptions (ESD) (Caudle, 2013), and a measure of taxonomic distance (Minasny & McBratney, 2007; Rossiter, Zeng, & Zhang, 2017). Both LCC and ESD provide a higher order grouping structure relative to soil series classes (Salley, Talbot, & Brown, 2016b), where multiple soil series can belong to the same LCC or ESD class based on a similarity in land potential and thus provide insight into the functional similarity of misidentified soil profiles. Taxonomic distance values provide a quantitative metric for identifying instances where the misidentified soil component is highly similar to the correct component in terms of the soil properties evaluated and should therefore be reassigned as a correct match. In this study we used distance in property space as our measurement of taxonomic distance, computed from Gower's generalized dissimilarity metric results returned from the 'NCSP' function. Specifically, we calculated the difference in taxonomic distance ($DTD$) between the predicted component (i.e., smallest Gower's dissimilarity to NCSS-SDC pedon) and the correct component (i.e., SSURGO component that matches the NCSS-SDC test pedon):

$$DTD = S_{tc} - S_{tp}$$

where $S_{tc}$ and $S_{tp}$ are the Gower's generalized dissimilarity metrics between the NCSS-SDC test pedon and the correct component and predicted component, respectively. Using a heuristic approach, we defined two thresholds (i.e., DTD5: 5 taxonomic distance units; DTD10: 10 taxonomic distance units) for determining an acceptable range of taxonomic distance between the predicted and correct SSURGO components. Threshold values were derived using expert knowledge and interpretation of soil series similarity using functional indicators (e.g., LCC, ESD) to determine conservative DTD value ranges. Thus, DTD5 and DTD10 represent the number of taxonomic distance units between the predicted component and the correct component considered to be within an acceptable range of similarity.

**TABLE 2** Numerical soil profile identification results (moist color)

| Property generalization[a] | Depth generalization | | |
|---|---|---|---|
| | Genetic horizons | Depth intervals | Equal-area spline |
| | Identification accuracy (%) | | |
| Component match | | | |
| Novice | 53 | 46 | 53 |
| Expert | 57 | 48 | 57 |
| Expert-Plus | 59 | 53 | 60 |
| ESD match | | | |
| Novice | 82 | 78 | 82 |
| Expert | 84 | 79 | 83 |
| Expert-Plus | 85 | 83 | 85 |
| LCC match | | | |
| Novice | 75 | 70 | 76 |
| Expert | 77 | 71 | 77 |
| Expert-Plus | 79 | 74 | 79 |
| DTD5 | | | |
| Novice | 67 | 57 | 69 |
| Expert | 74 | 65 | 75 |
| Expert-Plus | 77 | 70 | 78 |
| DTD10 | | | |
| Novice | 79 | 65 | 82 |
| Expert | 85 | 76 | 86 |
| Expert-Plus | 88 | 81 | 89 |

[a] ESD, ecological site description; LCC, land capability class; DTD, difference in taxonomic distance.
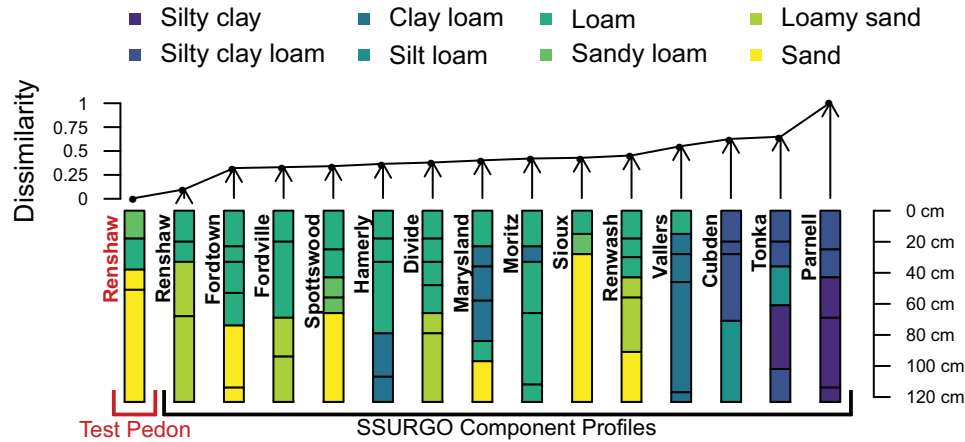
## 3 | RESULTS

The basic set of soil properties included in the Novice dataset resulted in nearly as high identification accuracy (46–53%) as that achieved with the Expert (48–57%) dataset, and only marginal improvement (53-60%) with the addition of pH and organic matter in the Expert-Plus dataset (Table 2). The type of depth-support (genetic horizons, standardized depth intervals, and equal-area splines) used to represent depth-wise soil variability had a small effect on identification accuracy, with slightly lower accuracies for the standardized depth intervals relative to genetic horizons or equal-area splines. Identification accuracies based on ESDs and LCC followed similar trends for the Novice (ESD: 78–82%; LCC: 70–76%), Expert (ESD: 79–84%; LCC: 71–77%), and Expert-Plus (ESD: 83–85%; LCC: 74–79%) datasets, with slightly lower accuracies for the standardized depth intervals (Table 2).

Examples of the numerical soil profile identification output for a correct component match and misidentification are shown in Figure 4. The Renshaw test case is an example of a correct component match where the most similar SSURGO component (i.e., lowest dissimilarity) matched the soil series classified to the NCSS–SCD test pedon (Figure 4a). In contrast, the Mobridge test case is an example of an apparent misidentification, where the NCSS–SCD test pedon matched the second most similar SSURGO component (Figure 4b). In US Soil Taxonomy, the Mobridge soil series is within the Fine-silty, mixed, superactive, mesic Pachic Argiustoll Family, while the Highmore soil series is in the Fine-silty, mixed, superactive, mesic Typic Argiustoll Family. Both series are classified with the same Family differentiae (i.e., fine-silty, mixed, superactive, mesic) but differ in their subgroup taxonomy (i.e., Pachic vs. Typic). This difference equates to a difference in mollic epipedon thickness, with the Mobridge's epipedon being greater than 50 cm (Pachic) and the Highmore's epipedon being less than 50 cm (Typic). While these types of difference result in separate subgroup taxa, from a land-use/land-potential perspective these differences may have little-to-no practical or applied meaning when a soil's properties fall within the tolerances of (for example) a single irrigation or fertilization schedule. This is supported by the fact that both components have the same LCC (2-c) and similar ESDs (Highmore ESD: Loamy; Mobridge ESD: Loamy Overflow).

We evaluated the use of taxonomic distance to account for functionally similar matches that were misidentified
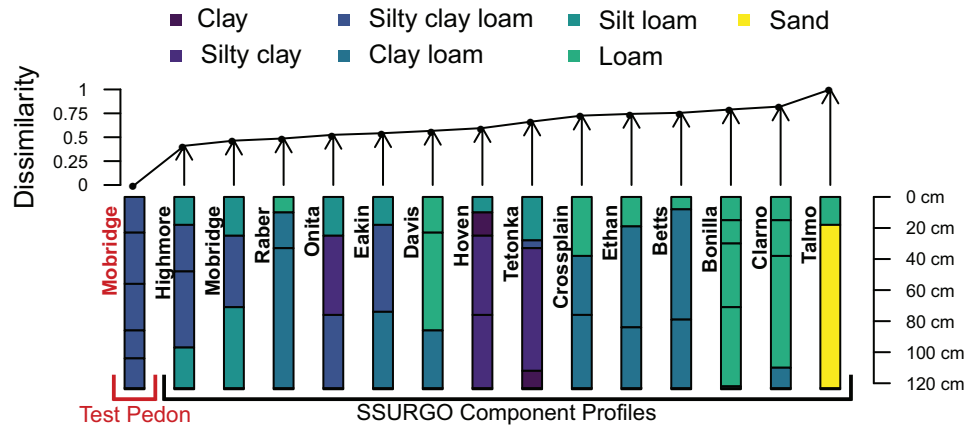
**FIGURE 4** Soil identification results using the Novice dataset with genetic horizon depth-support. SSURGO component profiles are ordered by increasing dissimilarity to the NCSS–SCD test pedon. The Renshaw test case (a) is an example of a component match where the NCSS–SCD test pedon series name and top ranking SSURGO component name match. The Mobridge test case (b) is an example of a misidentification (Mobridge = second rank) even though the test pedon exhibits a high taxonomic similarity amongst the top several SSURGO components. Based on the $DTD_5$ threshold, this case would be reassigned as correct due to its high taxonomic similarity. Soil texture classes are represented as an ordinal variable ranked according to approximate available water holding capacity given equal volume of coarse fragments and organic matter concentration. Soil texture classes are ranked in the following order: sand, loamy sand, sandy loam, loam, silt loam, silt, sandy clay loam, clay loam, silty clay loam, sandy clay, silty clay, and clay

due to minor taxonomic differences. The distribution of taxonomic distance values for the correctly classified SSURGO profiles (i.e., Match), the top ranking components (i.e., lowest dissimilarity) when a pedon was misidentified (No-Match – Top Component), and the correct component (i.e., SSURGO component matching the NCSS–SCD test pedon) when a pedon was misidentified (No-Match – Correct Component) are presented in Figures 5 and Supplemental Figure S1. These three groupings of taxonomic distance values were evaluated across the different soil property groups and depth-wise generalizations (Figures 5, Supplemental Figure S1). Similar to our component match rate results, the three different

soil property datasets, with their different degrees of soil property generalization, had a more pronounced effect on taxonomic distances compared to type of depth-support used (Figures 5, Supplemental Figure S1, Supplemental Table S1). One notable effect was the increase in taxonomic distances from the Novice to the Expert and from the Expert to the Expert-Plus datasets for both correctly classified (i.e., Match) and misidentified (i.e., No-Match) observations (Figure 5, Supplemental Table S1). This increase in taxonomic distance occurs due to the addition of several continuous soil properties (i.e., Sand, Silt, Clay, RFV, OM, pH), thus increasing the potential variability between soil components. However, despite this general
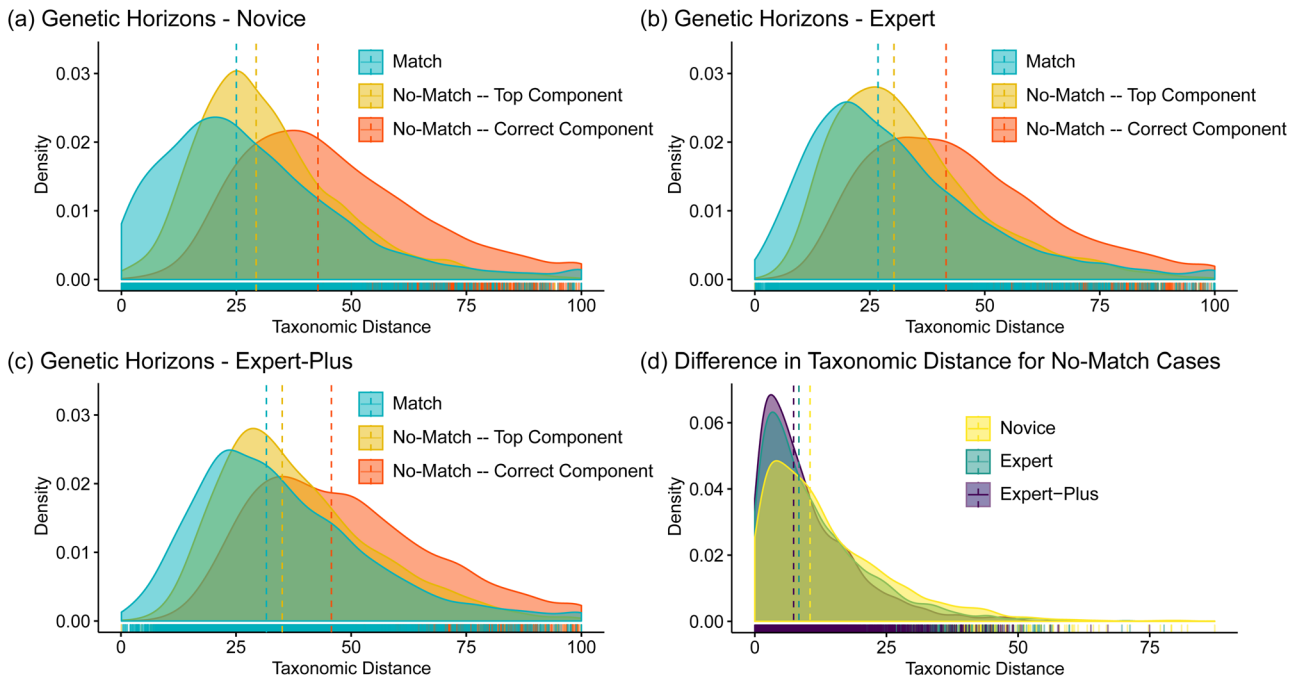
**FIGURE 5** Density plots of taxonomic distance for the Novice (a), Expert (b), and Expert-Plus (c) datasets, with depth-wise variability represented by genetic horizons. The distribution of taxonomic distance values are shown (a-c) for the correctly classified pedons (Match); the top ranking component when a pedon is misidentified (No-Match – Top Component); and the correct component when a pedon is misidentified (No-Match – Correct Component). For all misidentified (i.e, No-Match) cases, the distribution of the difference in the taxonomic distance between the predicted class (Top Component) and the correct component class (Correct Component) is also shown (d). Vertical lines indicate the median value for each group

increase in taxonomic distance, there was a decrease in the taxonomic distance between the top ranking component (i.e., smallest taxonomic distance to the test pedon) and the correct component class (i.e., SSURGO component matching the NCSS–SCD test pedon) with the addition of soil information (Figure 5d). In other words, for the misidentified pedons the use of more specific soil property data (e.g., percentage of clay vs. texture class) and the addition of new soil properties (i.e., OM, pH) resulted in predicted classes more similar to the correct soil classes. Additionally, the increase in taxonomic distance values from the Match to the No-Match – Top Component indicates that the misidentified NCSS–SCD test pedons are generally more dissimilar to the queried SSURGO components relative to the NCSS–SCD test pedons that were correctly matched. This suggests that the misidentified pedons were not well represented by the modal SSURGO profiles and thus likely represent extragrades within the queried soil components or possibly a component not represented by the mapped components within our search radius.

The use of our heuristic taxonomic distance thresholds raised the identification accuracy from 46–60% for component matches to 57–78% for the $DTD_5$ threshold and 65–89% for the $DTD_{10}$ threshold across all datasets. In the Mobridge test case the taxonomic difference between the predicted SSURGO class (Highmore) and the correct SSURGO class

(Mobridge) was 4.3 distance units. Based on our $DTD_5$ threshold, this test case would be reassigned as correct due to its high taxonomic similarity. Although the $DTD_5$ and $DTD_{10}$ thresholds are largely subjective, their match rates were very similar to the ESD (78–85%) and LCC (70–79%) match rates (Table 2).

## 4 | DISCUSSION

### 4.1 | Limits and sources of error

The error rate of the numerical soil identification framework presented in this study (40–54%) can be attributed to several sources, including those associated with: (1) our numerical soil identification framework, (2) the pedon point-location test dataset (NCSS–SCD), (3) soil survey map unit delineations and concepts (SSURGO), and (4) soil series concept purity/overlap. Each potential source of error is discussed in detail below.

### 4.1.1 | Numerical soil identification framework

In the hierarchical structure of Soil Taxonomy, criteria for the splitting of soil taxa often does not directly align with

the quantitative data used in our numerical soil identification. For example, at the highest classification level, differentiation between soil orders is largely based on diagnostic horizons which encapsulate depth dependent changes in specific soil properties (e.g., vertical arrangement and aggregate measures). At lower levels (i.e., suborder, great group, subgroup) most breaks are often based on soil limitations including those relating to the soil moisture regimes (Aquic), shallow bedrock (Lithic), or zones of mineral accumulation (Calcic). The depth dependent nature of these differentiating criteria makes it difficult to approximate using a slice-wise comparison approach. The *NCSP* algorithm can incorporate site or profile-specific attributes by calculating a separate profile dissimilarity matrix that can be averaged with the final summed slice-wise dissimilarity matrix. While all soil properties analyzed in this study were horizon or depth-specific, the addition of profile-specific (i.e., "aggregate") soil properties may help improve our identification accuracy. For example, Young and Hammer (2000) used profile-specific variables in a clustering analysis of soil classes including: depth to maximum clay content, difference between the minimum and maximum soil profile clay content, and thickness of soil profile colors meeting mollic criteria. In many cases the scale of aggregate profile-specific properties, which can be derived from existing depth-wise property data, is much closer to the scale of breaks and criteria used in Soil Taxonomy and thus may help to improve identification results.

Any numerical comparison of soil profiles requires that the collection of horizons for each soil profile be treated as an ordered group. In the *NCSP* algorithm this is partially accomplished by calculating a separate dissimilarity matrix at each soil layer or slice, followed by the summation of all slice-wise dissimilarity matrices. A major limitation of this approach is that it treats all depth increments as independent of each other, and thus fails to account for the depth dependent relationships that exist between soil properties that are heavily relied upon in Soil Taxonomy. Alternative approaches that can account for these depth dependent relationships, like incorporating soil diagnostic features in a separate profile dissimilarity matrix, would likely enhance our ability to correctly identify soil classes.

### 4.1.2 | NCSS–SCD

The NCSS–SCD is comprised of pedons collected since the 1950s. Such a wide time span introduces three sources of potential error. The first relates to changes in Soil Taxonomy that might have occurred since the pedon was initially sampled and whether the soil series initially classified to the pedon was updated during those changes. We attempted to minimize this type of error by only selecting NCSS–SCD pedons whose correlated soil series name (i.e., taxonname) matched a SSUROGO component name within our spatial query. This would prevent cases where a soil series name was discontinued; however, this may not prevent errors where a series was split into two or more separate series but the NCSS–SCD pedon was not updated to match the correct series. The second source of error has to do with GPS positional accuracy associated with each pedon. Since many of the pedons in NCSS–SCD were sampled before the development of GPS technology, their locations were recorded using the Public Land Survey System (PLSS) and later digitized using the Township and Range designations. Libohova, Seybold, Adhikari, Wills, & Beaudette, 2019 showed how low GPS positional accuracy associated with a subset of NCSS–SCD pedons containing both GPS and PLSS-derived coordinates resulted in very different soil pH values returned from SSURGO spatial queries. Errors associated with positional accuracy were also minimized by our spatial query, where the NCSS–SCD pedon matched a SSURGO component within the 1000 m search radius. Finally, the third source of potential error resulting from the temporal nature of the dataset can occur due to methodological changes in the measurement soil properties (e.g., pH, soil carbon) over time.

An additional source of error unrelated to the temporal nature of the dataset can occur due to sampling bias. In general, most soil surveys employ a purposive sampling design for the collection of field data, where a soil scientist will select sampling points in locations thought to, (i) represent the central concept of mapped soil series, (ii) delineate map unit edges, or (iii) validate soil-landscape models. This can result in a confirmation bias, where the soil mapper, based on their expert judgment, is expecting to characterize a particular soil series before even digging a soil pit (Libohova et al., 2019). This is more likely to occur on consociations where the dominant soil component comprises a large areal fraction of the map unit. For example, the Mobridge test case is in the 'Highmore silt loam, 2 to 6 percent slopes' soil map unit. This map unit is a consociation with the Highmore component mapped on 85% of the map unit and the Mobridge component mapped on 6% of the map unit. Based on these estimated distributions a general expectation to encounter the Highmore component is likely to be high.

### 4.1.3 | SSURGO

The SSURGO database consists of more than 3,000 soil surveys carried out over multiple decades. Due to their varying goals, these soil surveys have differences in the total area mapped as well as the level of support from field surveys and pedon data which generate component concepts

and delineate their spatial footprint across the landscape (Libohova et al., 2019). Initial soil surveys were conducted using political boundaries (e.g., county), while more recent soil surveys are conducted using natural boundaries (i.e., Major Land Resource Areas, MLRA; or Land Resource Units, LRU) (Indorante, McLeese, Hammer, Thompson, & Alexander, 1996). Soil surveys based on political boundaries often resulted in duplicative thematic classes (i.e., map unit and component concepts) between spatially adjacent survey areas. Mapping to MLRA boundaries has helped minimize this thematic redundancy since the spatial footprint of a particular set of unique soil components is typically contained within a resource area's natural boundary. Soil survey updates, however, generally occur on a 30-year cycle and many survey areas have not yet been updated or recorrelated along resource area boundaries (Indorante et al., 1996). Therefore, depending on the survey vintage and MLRA update progress, SSURGO map units can vary widely in their consistency/redundancy. Similarly, soil survey vintage also affects the spatial accuracy of SSURGO data, with older surveys (rangeland, forested lands, and wilderness areas) often mapped at coarser spatial resolutions (i.e., soil survey mapping order; e.g., Order 2: 1:12,000 vs. Order 4: 1:64,000), which affects the level of detail used to characterize soil spatial heterogeneity (i.e., consociation vs. association/complex and the mapping of minor components). Furthermore, most original soil surveys used rectified, high-altitude photographs as their standard base maps. Prior to digitization, all line work from these maps had to be transferred by hand to an acceptable base map (e.g., USGS 7.5′ topographic quadrangle), adding additional uncertainty on the final placement of the digitized map unit boundaries (D'Avello & McLeese, 1998).

The continuous nature of soil variability can result in a gradual gradation from one soil component to another within a SMU or collection of related SMUs. Consequently, there are regions of the landscape where the diagnostic characteristics of a soil closely match one soil component relative to others, while in other regions the component identity of the soil is less well defined (Young, Hammer, & Williams, 1997). Within SSURGO, each map unit component is described by its modal concept, reporting information aggregated from a collection of field-described soil profiles that represent the component concept. This study used the representative values for all SSURGO soil data. Depending upon the value range for each property within a given component class, the representative value may fail to adequately differentiate an instance of that class from other surrounding soil components. Thus, in many cases our inability to account for variability within component classes may limit our ability to accurately differentiate between-component differences (Soil Survey Division Staff, 2017).

### 4.1.4 | Soil taxonomy

Soil Taxonomy is a hierarchical classification system that consists of six levels: Order, Suborder, Great Group, Subgroup, Family, and Series. Classification from higher to lower levels occurs according to a bifurcated key based on specific rules and criteria. Common criteria used to separate soils at different levels of the hierarchy include diagnostic horizons (i.e., soil layers whose structure and genesis is correlated to specific soil-forming processes), soil temperature and moisture regimes, and specific soil property values or ranges (e.g., soil color, texture, and chemical properties). The rigid hierarchical structure of Soil Taxonomy can result in a high diversity of soil classes in areas where the range of diagnostic values spans multiple taxonomic thresholds. Numerous studies have reported a high degree of taxonomic variability within individual soil map units (McCormack & Wilding, 1969; Nordt, Jacob, & Wilding, 1991; Wilding, Jones, & Schafer, 1965; Young et al., 1997). For example, Young et al. (1997) examined the map unit composition of a single map unit on the Missouri River floodplain based on soil transects. The map unit, Eudora silt loam, was correlated as a consociation (i.e., one dominant soil type with only minor inclusions of dissimilar soils). In their analysis of the Eudora silt loam consociation, Young et al. (1997) identified 28 unique soil families. Furthermore, of the 120 observed pedons, only 25% (30) were in the named family taxa of the map unit, but 83% (100) were in either the named family taxa or a similar family taxa. High taxonomic diversity coupled with a high degree of interpretive similarity among soil taxon results in high levels of functional similarity and high misidentification rates.

At the lowest level of Soil Taxonomy, soil series concepts are developed based on a combination of soil taxonomic differences at the Family level and perceived management differences. In monothetic classification systems, such as Soil Taxonomy, many misidentifications are a result of small differences in the soil properties used to define diagnostic properties and horizons, where values slightly above or below each set threshold determines which path down the hierarchy a profile moves at each taxonomic break (i.e., 'taxonomic chop') (Butler, 1980; Rossiter et al., 2017). Thus, individual soils classified to separate taxa may only differ slightly with respect to one or more taxonomic limits, as was illustrated in the misidentification of the Mobridge soil series (Figure 4b). Taxonomic distance provides an alternative approach that can correct for cases where functionally similar soils are classified separately due to taxonomic thresholds. However, the utility of taxonomic distance as an alternative to the component match rate depends on the selection of soil properties. Different sets of soil properties that relate to specific soil functions or management

objectives can be used to provide specific measures of interpretive accuracy. For this study we selected soil properties that are considered important diagnostic indicators of land potential, and thus provide a general measure of interpretive accuracy based on the calculated distance values.

## 4.2 | Effects of soil property generalization and depth-support on identification accuracy

Generalizing soil property data and its effect on numerical soil identification has important implications for citizen soil science initiatives. Our evaluation of different soil property groups provided an assessment of both the effects of generalizing soil properties in feature space and the influence of different soil properties based on the level of experience required for their measurement. Among our soil property groups, the comparison between the Novice and Expert directly evaluated the effects of generalizing soil properties by comparing the use of texture and rock fragment volume classes (Novice group) to a more detailed characterization that measured the percent sand, silt, clay, and rock fragment volume. Adding these additional continuous soil properties only produced an incremental increase in accuracy. The evaluation of the Expert-Plus dataset provides insight into the potential improvement of identification accuracy with the collection of additional soil properties that typically require a high level of technical training or equipment (e.g., soil scientists, pH meters). Similar to the differences between the Novice and Expert datasets, there was a small improvement in the Expert-Plus datasets both in terms of the component match rate and a decrease in the taxonomic distance between the predicted class and correct class. This general trend suggests that additional, modest improvements in identification accuracy could be made by incorporating additional soil properties.

Our results showed that the form of depth-support used to characterize soil profiles had a small effect on identification accuracy, with the depth interval support producing slightly lower match rates (Table 2). For most soil properties, genetic horizon support is generally believed to provide a more accurate representation of depth-wise variability relative to the use of depth interval support. This is particularly true in highly stratified soils, soils with textural discontinuities, or soils with diagnostic horizons that overlap set depth intervals. In these types of soils the use of depth interval support can lead to very different soil property distributions relative to genetic horizon support. It is important to note that both soil datasets in this study (NCSS-SCD and SSURGO) use genetic horizons to describe soil profiles and thus the modeling of continuous depth functions in any soils with abrupt, highly contrasting genetic horizons could result in the creation of artifacts when resampling to standard depth intervals (e.g., averaging percentage of clay from Bqkm + Bt). However, any comparisons between genetic horizons and depth intervals must consider the number and sequence of depth intervals. In this study we used LandPKS depth intervals (i.e., 0–1, 1–10, 10–20, 20–50, 50–70, 70–100, 100–120 cm) which were selected to provide a general approximation of the vertical variation found in most soils. Our results confirm the utility of LandPKS depth intervals in approximating vertical soil variability across our test dataset (n = 6,220), where the use of depth interval support resulted in only a small drop in the match rates relative to genetic horizon support.

Evaluating identification accuracy based on exact component matching assumes that all misidentified observations are equally serious errors, resulting in the under-representation of the functional or management-relevant similarity. For example, in SSURGO many of the soil components returned from a spatial query of the home and adjacent map units are often taxonomically similar (often within the same Subgroup or Family) and have similar soil limitations and management potential but may only differ slightly with regard to a particular diagnostic criteria (e.g., color depth range requirement for mollic epipedons, Figure 4b). The use of taxonomic distance to evaluate identification accuracy can mitigate this but first requires identifying specific criteria (e.g., thresholds) for defining what is and is not similar. In our evaluation of taxonomic distance, we defined two taxonomic distance thresholds using a conservative value range based on the observed distribution of taxonomic distance values. Based on these subjective thresholds, we observed an appreciable increase in identification accuracy (from 46–60% to 65–89%) which we believe represents an acceptable range of variation while remaining sufficiently similar. Determining a threshold of taxonomic similarity that also provides a meaningful indication of similarity in land-use potential is challenging. The high accuracies based on the $DTD_{10}$ threshold are in a similar range as our management relevant identification accuracies using ESDs (78–85%) and LCC (70–79%). At the taxonomic classification level of the soil series or soil series-phase, differentiating criteria are often based on management relevant properties and conditions (e.g., slope, water table). Although the soil properties evaluated in this study are highly relevant for land management, they may not provide the necessary information to differentiate soils with other limiting physical or chemical properties.

## 4.3 | Improving numerical soil classification

Allocating soil profiles to existing classes using numerical methods is challenging in monothetic classification

systems. This is particularly true at highly detailed levels of classification such as the soil series in Soil Taxonomy. As we have shown, class confusion often occurs due to small differences in diagnostic properties and horizons that span across rigid thresholds, producing the taxonomic chop described by Butler (1980). This was shown with the Mobridge and Highmore series where small differences in mollic epipedon thickness separated the two series. These results support previous criticisms on the use of elaborate hierarchies of rationalized classes (i.e., divisions based on taxonomic chop) for soil mapping (Butler, 1980; Fitz-Patrick, 1971; Rossiter et al., 2017; Webster, 1968). When soil series are created through a rationalized subdivision of classes within a soil-taxonomic system, the differentiae used to subdivide classes may not reflect locally important differences in soil properties and conditions, thus resulting in poor predictive ability. A solution to this problem was proposed by Butler (1980) were he argues for a 'taxonomic hiatus' or a separation between the top-down, rationalized creation of classes within soil-taxonomic systems and the bottom-up creation of classes by the soil surveyor that conform to the landscape based on natural boundaries and local soil conditions. This hiatus would allow soil series to be created that reflect management relevant differentiae within a survey area (Salley, Curtis Monger, & Brown, 2016a), while still allowing soil series to be associated with one or more of the higher-level taxonomic classes.

Current soil-taxonomic systems develop soil series by rationalized subdivision of higher order classes, often resulting in functionally similar soil classes that are prone to misidentified. Furthermore, high levels of undocumented within-map unit taxonomic variability further complicate the ability to correctly match a field-described soil profile to estimated SSURGO component profiles that represent modal concepts. The extent to which a soil survey can characterize or minimize within-map unit variability depends on, (i) the level of intrinsic soil variability in a survey area, (ii) the logistical and monetary resources allocated to that survey effort, (iii) how well the current soil series classes fit the taxonomic variability, soil property modes and natural boundaries within that landscape, and (iv) the survey area objectives and specifications (USDA-NRCS, 2020). In light of the limitations imposed by current soil-taxonomic systems, three main factors currently limit the efficacy of soil identification models: (1) the ability to quantify and incorporate within-class variability; (2) the ability to evaluate each soil profile as an ordered group rather than a sequence of separate horizons or slices; and (3) the ability to define and measure metrics of soil functional similarity that can provide a more robust evaluation of model error.

The limitations imposed by rigid hierarchical classification systems are widely recognized (Lagacherie, 2005;

McBratney & de Gruijter, 1992; Odeh, McBratney, & Chittleborough, 1992) and several modeling approaches have been developed to address these issues, most notably those based on fuzzy systems (i.e., fuzzy set theory and fuzzy logic) and machine learning (ML) techniques (Burrough, Van Gaans, & Hootsmans, 1997; Chaney et al., 2016; Hengl, Toomanian, Reuter, & Malakouti, 2007; Heung et al., 2016; Hughes, McBratney, Minasny, & Campbell, 2014). While many numerical soil classification algorithms return a distance metric, these distances are not the same as measures of classification uncertainty (e.g., fuzzy or probabilistic class membership) returned from fuzzy clustering or ML methods. Techniques like fuzzy-k means with extragrades (McBratney & de Gruijter, 1992) or more recently the *akromeson* algorithm which employs endmembers to identify extragrades (Hughes et al., 2014) are able to quantify non-hierarchical continuous classes. This means that each soil profile in the classification can exist in more than one class based on its numeric distance to the different class centroids within the measured property space. Soil classification problems (i.e., arranging soil profiles into classes) employ unsupervised classification techniques where the number of unlabeled classes is unknown. While unsupervised ML techniques (e.g., artificial neural networks, self-organizing maps) are widely used for classification problems, they are not commonly used in soil classification. Soil identification problems (i.e., allocating a soil profile into an existing class), however, can use supervised classification techniques since the class labels are known. Spatial predictions of soil taxonomic classes have been made using supervised ML techniques (Chaney et al., 2016; Hengl et al., 2016; Ramcharan et al., 2017), however, these predictive soil mapping (PSM) applications are restricted to using spatially exhaustive environmental covariates (e.g., remote sensing-based indices) which can produce high prediction uncertainty. An interesting example of these efforts is the POLARIS SMU component map of the contiguous US which uses the DSMART algorithm to predict SSURGO components at a 30 m grid cell size (Chaney et al., 2016). A main limitation of the POLARIS modeling framework, however, lies in the random nature of the point sampling and component assignment within each soil map unit. This results in significant overlap in covariate space for the points assigned to each component, resulting in significant class confusion within each random forest model (e.g., 17% and 55% match rates for the top predicted and top ten predicted SMU component, respectively). Site-based soil identification (i.e., a single sampling site) has an advantage over PSM in that it is not restricted to spatially exhaustive covariates and therefore able to incorporate soil property data into the classification model.

Most soil classification systems use genetic horizon depth-support to characterize soil profiles which

complicates direct numerical comparisons between profiles due to differing horizonation. The *aqp NCSP* algorithm solves this issue by performing a series of separate slice-wise comparisons at standardized depths. However, when calculating distance based upon a series of separate slice-wise comparisons the depth-dependent relationships intrinsic to soil profile development (i.e., pedogensis) are largely lost. These depth-dependent soil property relationships are incorporated into most diagnostic criteria used in Soil Taxonomy. In addition to the potential benefits of using profile-specific attributes (e.g., diagnostic features: presence/absence, thickness/depth to top) which was previously addressed, two statistical approaches may further help quantify these depth-dependent soil relationships. The first involves the use of soil depth and soil property weights within the slice-wise dissimilarity calculations. The *NCSP* algorithm can accommodate variable depth weighting and Gower's generalized dissimilarity can accommodate variable property weighting. The combination of both depth and property weights would allow for the creation of custom weighting schemes that could be implemented using a rule-based framework built upon Soil Taxonomy. For example, if the dominant differentiating criteria among soils returned from a SSURGO query was soil color in the top 30 cm, we could assign higher property weights to color in the top 30 cm and higher depth weights to the slices in the top 30 cm. Similarly, if texture differences in the particle size control section were important, property weights for soil texture and depth weights for slices within the control section could be increased. The second approach involves the use of statistical models (e.g., ML models) that can accommodate large numbers of covariates and thus analyze the complete set of soil property values across all slices rather than evaluating each slice separately. While this approach solves the issue of treating all profile slices as a group, it does not retain the order of the slices within each profile group which results in some information loss. This can be partially corrected for by incorporating profile-specific attributes in addition to the collection of slice-wise attributes.

Functional similarity among soil series can complicate the evaluation of soil identification accuracy. Future work is needed to develop robust metrics of soil functional similarity among soil series so that the accuracy of soil identification models may be better evaluated (Rossiter et al., 2017).

## 5 | CONCLUSIONS

Soil components allow land managers to infer a general range of soil behavior in response to management actions and disturbance effects, indexed according to labels that do not require an understanding of Soil Taxonomy. Correct identification of soil components at sampling locations is a challenging task, even for many soil scientists. A main source of confusion lies in the fact that soils are complex, heterogeneous systems that exhibit high variability in both geographic and feature space. Despite these challenges, this study demonstrated that numerical soil classification can be used to support soil identification by citizen scientists using limited, simple soil observations. Numerical soil classification algorithms provide a quantitative method for describing differences between soils and when coupled with accurate soil property data provides an objective framework for allocating an unknown soil to an existing soil class (e.g., soil component). This study showed that the use of highly generalized soil property data (e.g., soil texture class and rock fragment volume class, soil color) resulted in nearly as high identification accuracy as that achieved with more precise soil property data (e.g., percentage Sand, pH, OM) that requires a high level of technical training and/or equipment. Furthermore, this study showed minimal effect from the type of depth-support used to characterize vertical soil variability, with slightly lower identification accuracies using standardized depth intervals. The ability for land managers and other non-soil specialists to accurately identify their soil is critical for implementing sustainable soil management. Results from this study support the feasibility of soil identification by citizen scientists through the collection of soil profile data using generalized property classes and standardized depth intervals, provided the interval ranges sufficiently characterize locally important soil features (e.g. abrupt changes in soil texture or pH, diagnostic features such as duripans or petrocalcic horizons, bedrock contact, etc.). While this study supports the use of existing data and technology to assist non-soil scientists in identifying and understanding the soil resource, we do not suggest that the role of the pedologist is on a path towards obsolescence. Rather, given the complexity of soil systems and their central role in land health, the job of the pedologist is needed now more than ever.

## ORCID
*Jonathan J. Maynard* https://orcid.org/0000-0003-2546-0073
*Shawn W. Salley* https://orcid.org/0000-0002-6092-0154

## REFERENCES

Beaudette, D. E., Roudier, P., & O'Geen, A. T. (2013). Algorithms for quantitative pedology: A toolkit for soil scientists. *Computers & Geosciences*, *52*, 258–268. https://doi.org/10.1016/j.cageo.2012.10.020.

Bishop, T. F. A., McBratney, A. B., & Laslett, G. M. (1999). Modeling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma*, *91*(1–2), 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8.

Brevik, E. C., Calzolari, C., Miller, B. A., Pereira, P., Kabala, C., Baumgarten, A., & Jordán, A. (2015). Soil mapping, classification, and pedologic modeling: History and future directions. *Geoderma*, *264*, 256–274(November 2015). https://doi.org/10.1016/j.geoderma.2015.05.017.

Burrough, P. A., Van Gaans, P. F. M., & Hootsmans, R. (1997). Continuous classification in soil survey: Spatial correlation, confusion and boundaries. *Geoderma*, *77*(2–4), 115–135. https://doi.org/10.1016/S0016-7061(97)00018-9.

Butler, B. E. (1980). *Soil classification for soil survey*. UK: Oxford University Press.

Carré, F., & Jacobson, M. (2009). Numerical classification of soil profile data using distance metrics. *Geoderma*, *148*(3–4), 336–345. https://doi.org/10.1016/j.geoderma.2008.11.008.

Caudle, D. (2013). Interagency ecological site handbook for rangelands. U.S. Department of the Interior, Bureau of Land Management.

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, *274*, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025.

D'Avello, T. P., & McLeese, R. L. (1998). Why are those lines placed where they are?: An investigation of soil map recompilation methods. *Soil Horizons*, *39*(4), 119. https://doi.org/10.2136/sh1998.4.0119.

Duniway, M. C., Bestelmeyer, B. T., & Tugel, A. (2010). Soil processes and properties that distinguish ecological sites and states. *Rangelands*, *32*(6), 9–15. https://doi.org/10.2111/RANGELANDS-D-10-00090.1.

Fan, Z., Herrick, J. E., Saltzman, R., Matteis, C., Yudina, A., Yudina, A., ... Van Zee, J. (2017). Measurement of soil color: A comparison between smartphone camera and the Munsell color charts. *Soil Science Society of America Journal*, *81*(5), 1139. https://doi.org/10.2136/sssaj2017.01.0009.

Fan, Z., Herrick, J. E., Saltzman, R., Matteis, C., Yudina, A., Yudina, A., ... Van Zee, J. (2018). Approaches for improving field soil identification. *Soil Science Society of America Journal*, *82*(4), 871–877. https://doi.org/10.2136/sssaj2017.09.0337.

FitzPatrick, E. A. (1971). Pedology: A systematic approach to soil science. *The Journal of Ecology*, *62*(1), 344.

Forbes, T. R., Rossiter, D., & Van Wambeke, A. (1987). *Guidelines for evaluating the adequacy of soil resource inventories*. Cornell University: Department of Agronomy.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857. https://doi.org/10.2307/2528823.

de Gruijter, J. J. (1977). Numerical classification of soils and its application in survey. *Agriculture Research Reports*, *1977*.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., & Al, E. (2016). SoilGrids250m: Global gridded soil information based on Machine Learning. *Journal of Earth System Science Data in review*, *12*(2), e0169748

Hengl, T., Toomanian, N., Reuter, H. I., & Malakouti, M. J. (2007). Methods to interpolate soil categorical variables from profile observations: Lessons from Iran. *Geoderma*, *140*(4), 417–427. https://doi.org/10.1016/j.geoderma.2007.04.022.

Herrick, J. E., Beh, A., Barrios, E., Bouvier, I., Coetzee, M., Dent, D., ... Karl, J. W. (2016). The Land-Potential Knowledge System (LandPKS): Mobile apps and collaboration for optimizing climate change investments. *Ecosystem Health and Sustainability*, *2*(3), e01209. https://doi.org/10.1002/ehs2.1209.

Herrick, J. E., Karl, J. W., McCord, S. E., Buenemann, M., Riginos, C., Courtrigh, E., ... Bestelmeyer, B. (2017). Two new mobile apps for rangeland inventory and monitoring by landowners and land managers. *Rangelands*, *39*(2), 46–55. https://doi.org/10.1016/j.rala.2016.12.003.

Herrick, J. E., Urama, K. C., Karl, J. W., Boos, J., Johnson, M. V. V., Courtright, E., ... Ganguli, A. C. (2013). The global Land-Potential Knowledge System (LandPKS): Supporting evidence-based, site-specific land use and management through cloud computing, mobile applications, and crowdsourcing. *Journal of Soil and Water Conservation*, *68*(1), 5A–12A. https://doi.org/10.2489/jswc.68.1.5A.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, *265*(March), 62–77. https://doi.org/10.1016/j.geoderma.2015.11.014.

Hudson, B. D. (1992). The soil survey as paradigm-based science. *Soil Science Society of America Journal*, *56*(3), 836–841.

Hughes, P. A., McBratney, A. B., Minasny, B., & Campbell, S. (2014). End members, end points and extragrades in numerical soil classification. *Geoderma*, *226–227*(1), 365–375. https://doi.org/10.1016/j.geoderma.2014.03.010.

Indorante, S. J., McLeese, R. L., Hammer, R. D., Thompson, B. W., & Alexander, D. L. (1996). Positioning soil survey for the 21st century. *Journal of Soil and Water Conservation*, *51*(1), 21–28.

IUSS - WRB. 2015. International soil classification system for naming soils and creating legends for soil maps.

Klingebiel, A. A., & Montgomery, P. H. (1966). Land-capability classification. Soil Conservation Service, U.S. Department of Agriculture.

Lagacherie, P. (2005). An algorithm for fuzzy pattern matching to allocate soil individuals to pre-existing soil classes. *Geoderma*, *128*(3-4 SPEC. ISS.), 274–288. https://doi.org/10.1016/j.geoderma.2005.04.009.

Libohova, Z., Seybold, C., Adhikari, K., Wills, S., Beaudette, D., Peaslee, S., ... Owens, P. R. (2019). The anatomy of uncertainty for soil pH measurements and predictions: Implications for modellers and practitioners. *European Journal of Soil Science*, *70*(1), 185–199. https://doi.org/10.1111/ejss.12770.

Libohova, Z., Wills, S., & Odgers, N. P. 2013. Legacy data quality and uncertainty estimation for United States GlobalSoilMap products. GlobalSoilMap: Basis of the Global Spatial Soil Information System - Proceedings of the 1st GlobalSoilMap Conference. Taylor & Francis London. p. 63–68

Little, I. P., & Ross, D. R. (1985). The levenshtein metric, a new means for soil classification tested by data from a sand-podzol

chronosequence and evaluated by discriminant function analysis. *Australian Journal of Soil Research*, *23*(2), 319–324. https://doi.org/10.1071/SR9850115.

Malone, B. (2016). Ithir: Functions and algorithms specific to pedometrics.

Malone, B. P., McBratney, B., Minasny, B., & Laslett, G. M. (2009). Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma*, *154*(1–2), 138–152. https://doi.org/10.1016/j.geoderma.2009.10.007.

McBratney, A. B., & de Gruijter, J. J. (1992). A continuum approach to soil classification by modified fuzzy k-means with extragrades. *Journal of Soil Science*, *43*(1), 159–175. https://doi.org/10.1111/j.1365-2389.1992.tb00127.x.

McCormack, D. E., & Wilding, L. P. (1969). Variation of soil properties within mapping units of soils with contrasting substrata in northwestern Ohio. *Soil Science Society of America Journal*, *33*(4), 587. https://doi.org/10.2136/sssaj1969.03615995003300040030x.

Minasny, B., & McBratney, A. B. (2007). Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma*, *142*(3–4), 285–293. https://doi.org/10.1016/j.geoderma.2007.08.022.

Moore, A. W., Russell, J. S., & Ward, W. T. (1972). Numerical analysis of soils: A comparison of three soil profile models with field classification. *Journal of Soil Science*, *23*(2), 193–209.

Nordt, L. C., Jacob, J. S., & Wilding, L. P. (1991). Quantifying map unit composition for quality control in soil survey. *Spatial Variability in Soils & Landforms*, *1991*, 183–198.

Norris, J., & Dale, M. (1971). Transition matrix approach to numerical classification of soil profiles. *Soil Science Society of America Journal*, *35*(3), 487–491.

Odeh, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1992). Fuzzy-c-means and kriging for mapping soil as a continuous system. *Soil Science Society of America Journal*, *56*(6), 1848–1854. https://doi.org/10.2136/sssaj1992.03615995005600060033x.

Odgers, N. P., McBratney, A. B., & Florence, C. (2018). Soil Profile Classes. *Pedometrics* (pp. 265–288). New York: Springer.

Podani, J. (1999). Extending Gower's general coefficient of similarity to ordinal characters. *Taxon*, *48*(2), 331–340. https://doi.org/10.2307/1224438.

Ponce-Hernandez, R., Marriott, F. H. C., & Beckett, P. H. T. (1986). An improved method for reconstructing a soil profile from analyses of a small number of samples. *Journal of Soil Science*, *37*(3), 455–467.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J. (2017). Soil property and class maps of the conterminous US at 100 meter spatial resolution based on a compilation of National Soil Point Observations and Machine Learning. *Soil Science Society of America Journal (2001)*, *82*(1), 186–201

Rayner, J. H. (1966). Classification of soils by numerical methods. *Journal of Soil Science*, *17*(1), 79–92.

Rossiter, D. G., Liu, J., Carlisle, S., & Zhu, A. X. (2015). Can citizen science assist digital soil mapping? *Geoderma*, *259–260*, 71–80. https://doi.org/10.1016/j.geoderma.2015.05.006.

Rossiter, D. G., Zeng, R., & Zhang, G. L. (2017). Accounting for taxonomic distance in accuracy assessment of soil class predictions. *Geoderma*, *292*, 118–127. https://doi.org/10.1016/j.geoderma.2017.01.012.

Salley, S. W., Curtis Monger, H., & Brown, J. R. (2016a). Completing the land resource hierarchy. *Rangelands*, *38*(6), 313–317. https://doi.org/10.1016/j.rala.2016.10.003.

Salley, S. W., Herrick, J. E., Holmes, C. V., Karl, J. W., Levi, M. R., McCord, S. E., … Van Zee, J. W. (2018). A comparison of soil texture-by-feel estimates: Implications for the citizen soil scientist. *Soil Science Society of America Journal*, *82*(6), 1526–1537. https://doi.org/10.2136/sssaj2018.04.0137.

Salley, S. W., Talbot, C. J., & Brown, J. R. (2016b). The Natural Resources Conservation Service land resource hierarchy and ecological sites. *Soil Science Society of America Journal*, *80*(1), 1. https://doi.org/10.2136/sssaj2015.05.0305.

Soil Survey Division Staff. (2017). *Soil survey manual* (pp. 120–125). Handbook No. 18. Ditzler C., & Scheffe K., and Monger H.C., (Eds.). Washington, DC: United States Department of Agriculture.

Soil Survey Staff. (1999). Soil taxonomy. A basic system of soil classification for making and interpreting soil surveys. Agriculture Handbook No. 436. Soil Conservation Service, U.S. Department of Agriculture. https://doi.org/10.1017/S0016756800045489.

Torrent, J., & Barron, V. (1993). Laboratory measurement of soil color: Theory and practice. In *Soil color* (Vol. *31*, pp. 21–33). Madison, WI: SSSA Spec. Publ. Series.

USDA-NRCS. 2020. National soil survey handbook, title 430-VI.

Webster, R. (1968). Fundamental objections to the 7th approximation. *Journal of Soil Science*, *19*(2), 354–366. https://doi.org/10.1111/j.1365-2389.1968.tb01546.x.

Wilding, L. P., Jones, R. B., & Schafer, G. M. (1965). Variation of soil morphological properties within Miami, Celina, and Crosby mapping units in west-central Ohio. *Soil Science Society of America Journal*, *29*(6), 711. https://doi.org/10.2136/sssaj1965.03615995002900060033x.

Young, F. J., & Hammer, R. D. (2000). Defining geographic soil bodies by landscape position, soil taxonomy, and cluster analysis. *Soil Science Society of America Journal*, *64*(3), 989. https://doi.org/10.2136/sssaj2000.643989x.

Young, F. J., Hammer, R. D., & Williams, F. (1997). Estimation of map unit composition from transect data. *Soil Science Society of America Journal*, *61*(3), 854–861. https://doi.org/10.2136/sssaj1997.03615995006100030020x.

## SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.