Published online May 17, 2018

Approaches for Improving Field Soil Identification

Zhaosheng Fan*
USDA-ARS Research Unit @ the Jornada
2995 Knox St.
Las Cruces, NM 88003

Skye A. Wills
USDA-NRCS
National Soil Survey Center
Lincoln, NE 68508

Jeffrey E. Herrick
USDA-ARS Research Unit @ the Jornada
2995 Knox St.
Las Cruces, NM 88003

Travis W. Nauman
U.S. Geological Survey Southwest
Biological Science Center
Moab, UT 84532

Colby W. Brungard
Dep. of Plant and Environmental Sciences
New Mexico State Univ.
Las Cruces, NM 88003

Dylan E. Beaudette
USDA-NRCS Soil Science Division
Sonora, CA 95370

Matthew R. Levi
USDA-ARS Research Unit @ the Jornada
2995 Knox St.
Las Cruces, NM 88003

Anthony T. O'Geen
Dep. of Land, Air, and Water Resources
Univ. of California
Davis, CA 95616

Use of soil survey information by non-soil-scientists is often limited by their inability to select the correct soil map unit component (COMP). Here, we developed two approaches that can be deployed to smartphones for non-soil-scientists to identify COMP from the location alone or location together with easily observed field data (i.e., slope, depth to the restrictive layer, and soil texture by depth). In addition, we also compared the two newly developed approaches with a traditional approach identifying COMP based on the dominant COMP (DC-based approach). All three approaches were tested with the Rapid Assessment of US Soil Carbon database and the combined USDA- NRCS Soil Survey Geographic database and the USDA-NRCS State Soil Geographic Database. The results indicated that the observation-based approach performed significantly better than the other two approaches, suggesting that a small set of easy-to-measure site-specific observations could significantly improve COMP identification. The location- and DC-based approaches had similar low performance overall. However, the location-based approach slightly improved identifications over the DC-based approach for cases where (i) there were multiple possible components within the soil map unit and (ii) the components were located in close proximity to a boundary of a different soil map unit polygon. The benefit of using the location-based approach may be greater in specific soil survey areas where topography was the major factor leading to the creation of the map unit legend.

Abbreviations: COMP, soil map unit component; CoSSGO, Combined SSURGO and STATSGO2 Geographic Database; DC, dominant component (approach); NSE, Nash–Sutcliffe efficiency; RaCA, Rapid Assessment of US Soil Carbon database; SMU, soil map unit; SMUP, soil map unit polygon; SSURGO, soil survey geographic database; STATSGO2, state soil geographic database.

Core Ideas

- **The traditional dominant-component-based approach is not reliable for identifying soils.**
- **We developed two approaches that can be used to identify soils with mobile devices.**
- **A small set of easily collected field data could greatly improve soil identification.**

Soil surveys can provide access to a virtual treasure trove of knowledge and information (Adhikari and Hartemink, 2016; Hudson, 1992; Sanchez et al., 2009). Mobile apps such as SoilWeb (O'Geen et al., 2017; http://casoilresource.lawr.ucdavis.edu, accessed 29 Mar. 2018) and mySoil (http://www.bgs.ac.uk/mySoil, accessed 29 Mar. 2018) now allow anyone with access to a global positioning system-enabled mobile phone or a computer to determine which soil map unit (SMU) they are in (Beaudette and O'Geen, 2009). These tools, however, still require an understanding of soil survey and map unit concepts to correctly identify the soil series COMP or to determine that the named COMP is not included in the map unit (Rossiter et al., 2015). This is particularly challenging in SMUs that consist of multiple components mapped together (i.e., an association or complex).

Therefore, individuals wishing to access soil information for a particular location most commonly assume that the dominant COMP exists across the entire SMU (Anderson et al., 2006; Miller and White, 1998). Here, we refer to this simplification of the soil landscape model as the DC-based approach. The DC-based approach is simple and straightforward, but is often unreliable where: (i) there are multiple possible components within the map unit (Subburayalu et al., 2014) and (ii) a location of interest is close to a boundary of a map unit consisting of a different dominant com-

Soil Sci. Soc. Am. J.
doi:10.2136/sssaj2017.09.0337
Received 26 Sep. 2017.
Accepted 2 Mar. 2018.
*Corresponding author (zhaosheng.fan@ars.usda.gov).

ponent (Gatzke et al., 2011).In addition, the DC-based approach potentially misses important soil variations, which may result in inefficient or incorrect management decisions. Using only a DC-based approach ultimately leads to diminished confidence in soil survey products, leading to wasted resources, crop failure, or structural damage to infrastructure where the identified soil is quite different from the correct one.

Many consumers of soil surveys lack the training or time to validate map unit concepts in the field. Therefore, the objective of this study was to develop an approach that can be used to help improve the ability of non-soil-scientists to identify the COMP correctly. Two new approaches were developed from information in traditional soil maps based on either location alone or location together with soil properties that can be relatively easily determined with guidance provided through smartphone and other mobile platforms. Therefore, end-users of our approaches would only need to make simple observations in the field to identify the most likely COMP, increasing the accessibility of soil survey information to those with little or no soil science training.

MATERIALS AND METHODS

Soil Geodatabase

The digital soils geographic database used in this study was derived from the USDA–NRCS Soil Survey Geographic database (SSURGO) and the USDA-NRCS State Soil Geographic Database (STATSGO2). Both SSURGO and STATSGO2 are available for download from Web Soil Survey at http://websoilsurvey.nrcs.usda.gov (accessed 29 Mar. 2018). The SSURGO database has a higher spatial resolution with a map scale from 1:12,000 to 1:63,360 and covers ~95% of the contiguous United States. By contrast, STATSGO2 has a lower spatial resolution with a map scale of 1:250,000 and complete coverage of the contiguous United States (the lower 48 states).

We first combined SSURGO and STATSGO2 to generate a soil property geographic database with complete coverage across the contiguous United States. This was done by first removing the coverage areas of STATSGO2 that overlapped with the more detailed SSURGO and then combining the remaining STATSGO2 with SSURGO: the Combined SSURGO and STATSGO2 Geographic Database (CoSSGO) database. One polygon in the CoSSGO database represents one map unit; one map unit consists of one or more COMPs; one COMP consists of several soil horizons. Site-specific (e.g., slope) and horizon-specific information (e.g., sand, silt, and clay contents) are included in the CoSSGO database.

Approach Development

The goal of this study was to develop approaches suitable for nonexperts that can be easily implemented on mobile devices. Therefore, we assumed that the following field observations are important and also feasible to collect with mobile devices by nonexperts: spatial location (latitude and longitude), slope, soil texture (0–200 cm) by feel (Vos et al., 2016), and depth to restrictive layer.

Accordingly, we developed two approaches that used different types of observations (Fig. 1). The first approach (hereafter called the location-based approach) was developed to identify the COMP solely based on the geographic location within the map unit. The second approach (hereafter called the observation-based approach) was developed to identify the COMP on the basis of easily observed soil properties (e.g., soil texture by depth). Detailed descriptions of the two approaches are given in the following sections.

The Location-Based Approach

The location-based approach was developed to identify the COMP on the basis of the location (i.e., latitude and longitude) of a point in consideration of all soil map unit polygons (SMUPs) surrounding that location. For a given location, the approach involves the following steps in order (Fig. 1):

(i) Identify the internal "home" SMUP that the location falls within;

(ii) Identify the external "neighbor" (contiguous) SMUPs that fall within a 100-m radius of the location (we also examined radii greater than 100 m; however, the results were not different from those with the 100-m radius);

(iii) Query the COMPs of both the neighbor and home SMUPs,

(iv) Calculate the conditional probability of each COMP (as presented later), and

(v) Identify the COMP with the highest conditional probability.

The conditional probability of a COMP is a function of: (i) the distance-weighted probability of the SMUP containing the COMP and (ii) the probability of the COMP within its SMUP. For a given location, the conditional probability of a COMP was defined as:

$$P_{comp,j} = \frac{\sum_{i=1}^{i_n} \left(p_{mu,i} \times p_{comp,i,j} \right)}{\sum_{i=1}^{i_t} \sum_{k=1}^{k_{t,i}} \left(p_{mu,i} \times p_{comp,i,k} \right)} \quad [1]$$

where $P_{comp,j}$ is the conditional probability of COMP j, i_n is the total number of home and neighbor SMUPs that contain COMP j, $P_{mu,i}$ is the distance-weighted probability of SMUP i, and $p_{comp,i,j}$ is the probability (i.e., the area coverage of COMP in the CoSSGO database) of COMP j in SMUP i, i_t is the total number of home and neighbor SMUPs, $k_{t,i}$ is the total number of COMPs in the SMUP i, and $P_{comp,i,k}$ is the probability of COMP k in the SMUP i. The distance-weighted probability of SMUP i (i.e., $P_{mu,i}$) is defined as:

$$P_{mu,i} = e^{\beta \cdot D_{mu,i}} \text{ for the neighbor map unit;} \quad [2]$$

$$P_{mu,i} = 1.0 \text{ for the home map unit,} \quad [3]$$

where β is the exponential coefficient (m^{-1}) and $D_{mu,i}$ is the shortest distance (m) from the given location to SMUP i. The exponential coefficient, β, is unknown and is estimated as presented later.

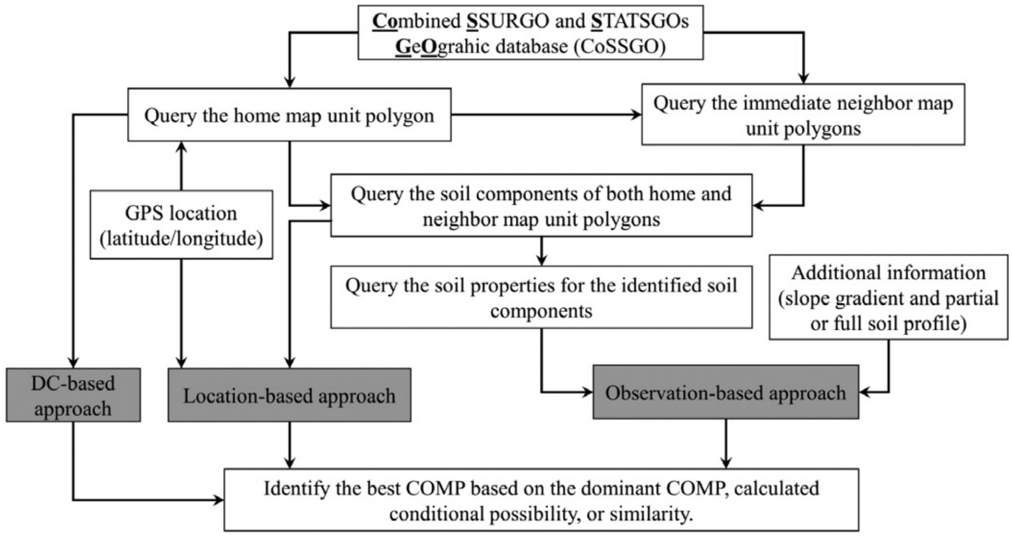

Fig. 1. The workflow of the two developed approaches (location- and observation-based) and the traditional dominant component based approach (the DC-based approach). COMP = soil map unit component.

The Observation-Based Approach

The observation-based approach identifies the COMP by calculating the similarity between the observations and the likely COMP. After the home and neighbor COMPs are identified (as discussed in the previous section), the similarities in slope, soil texture by depth, and depth to restrictive layer (e.g., bedrock, if a full soil profile is described) between observations and COMPs are then calculated. We chose these soil properties to identify COMP because they are relatively easily observed in field; however; they are not the only properties that can be used to differentiate soils. Many other soil properties (e.g., cation exchange capacity, and pH) are also important for differentiating soils but are not easily observed by end-users (e.g., the general public), so these properties are not included in our approaches.

Slope and depth to restrictive layer similarities are defined as:

$$S_{slope,j} = 1.0 - \frac{\left|Slope_o - Slope_{C,j}\right|}{Slope_R}; \quad [4]$$

$$S_{DtR,j} = 1.0 - \frac{\left|DtR_o - DtR_{C,j}\right|}{DtR_R}, \quad [5]$$

where $S_{slope,j}$ is the similarity between the observed slope and that of COMP j; $S_{DtR,j}$ is the similarity between the observed depth to restrictive layer and the representative value of COMP j; $Slope_O$ (in %) and DtR_O (in cm) are the observed slope and depth to the restrictive layer, respectively; $Slope_{C,j}$ (in %) and $DtR_{C,j}$ (in cm) are the slope and depth to the restrictive layer of COMP j in the CoSSGO database, respectively; and $Slope_R$ and DtR_R are the ranges of slope and depth to the restrictive layer in the CoSSGO database, which are set to 200% and 457 cm, respectively. We expect that these similarities would be automatically calculated and accessible to end-users (soil scientists and non-soil-scientists) via mobile devices (e.g., smartphones) and/or web browsers in the future.

It is reasonable for end-users to estimate soil texture by feel and then record the soil texture class (Vos et al., 2016) via mobile devices with guidance embedded in mobile applications. For example, LandInfo, one of the LandPKS app modules, has video demonstrations on how to conduct field estimates of soil texture by feel (Herrick et al., 2016). The similarity in soil texture between a measured soil profile and a COMP profile can be calculated via the following steps:

(i) Identify the texture classes of the observed soil profile;

(ii) Find the representative sand and clay contents for the texture classes of the observed soil profile (Table 1);

(iii) Find the texture classes of the profile of the COMP;

(iv) Find the representative sand and clay contents for the profile of the COMP (Table 1);

(v) Slice the observed soil profile and the profile of the COMP into layers with 1 cm thickness (Beaudette et al., 2013);

(vi) Calculate the similarities of sand and clay contents as:

$$S_{sand,j} = \frac{\sum_{z=1}^{z_b}\left(1.0 - \frac{\left|Sand_{O,z} - Sand_{C,z,j}\right|}{Sand_R}\right)}{z_b}; \quad [6]$$

$$S_{clay,j} = \frac{\sum_{z=1}^{z_b}\left(1.0 - \frac{\left|Clay_{O,z} - Clay_{C,z,j}\right|}{Clay_R}\right)}{z_b}; \quad [7]$$

$$z_b = \min(z_{b,p}, z_{b,c,j}), \quad [8]$$

where $S_{sand,j}$ and $S_{clay,j}$ are the similarity of sand and clay between the observed soil profile and COMP j; z is the soil depth (cm); z_b is the calculation domain (cm); $z_{b,p}$ and $z_{b,c}$ are the depths to the lower boundary of the observed soil profile and profile of COMP j, respectively (in cm); $Sand_{O,z}$ and $Clay_{O,z}$ are the representative sand and clay contents (in %) at soil depth z (cm) based on the field-observed soil texture class, respectively; $Sand_{C,z,j}$ and $Clay_{C,z,j}$ are the representative sand

Table 1. The representative sand and clay contents for the texture classes developed by the USDA.

Soil texture class	Range		Representative content	
	Sand	Clay	Sand	Clay
	%		%	
Sand	85–100	0–10	92.5	5
Loamy sand	70–90	0–15	80	7.5
Sandy loam	43–80	0–20	61.5	10
Sandy clay loam	45–80	20–35	62.5	27.5
Loam	23–52	7–27	37.5	17
Silt	0–20	0–12	10	6
Silt loam	0–50	0–27	25	13.5
Silty clay loam	0–20	27–40	10	33.5
Clay loam	20–45	27–40	32.5	33.5
Sandy clay	45–65	35–55	55	45
Silty clay	0–20	40–60	10	50
Clay	0–45	40–100	22.5	70

and clay contents (in %) at depth z based on the soil texture class of COMP j; and $Sand_R$ and $Clay_R$ are the ranges of representative sand and clay contents (in %) in Table 1,which are set to 82.5% and 65%, respectively.

The final step is to calculate the similarity of the soil texture as:

$$S_{text,j}=\frac{S_{sand,j}+S_{clay,j}}{2}, \qquad [9]$$

where $S_{text,j}$ is the similarity between the observed soil texture by depth and that of COMP j.

The total similarity is calculated as:

$$\begin{cases} S_{total,j}=\dfrac{S_{slope,j}+S_{DtR,j}+S_{text,j}}{3} \\ \text{if a full soil profile is described in the field} \\ S_{total,j}=\dfrac{S_{slope,j}+S_{text,j}}{2} \\ \text{if a partial soil profile is described in the field} \end{cases}, \qquad [10]$$

where $S_{total,j}$ is the total similarity with COMP j. A higher total similarity indicates a better match between a COMP and an observation.

The DC-Based Approach

For comparison, the COMP was also identified using the traditional DC-based approach, which is based on the dominant COMP of a map unit. For a given location, the DC-based approach first queries all of the COMPs in the home SMUP that the location falls within and then identifies the dominant COMP that has the maximum area coverage in the home SMUP (Fig. 1). Mathematically, the location-based approach can be simplified to the DC-based approach by assuming that the distance-weighted probability of the neighbor SMUP (i.e., $P_{mu,i}$ in Eq. [2]) is equal to zero.

Performance Evaluation

We used the Rapid Assessment of US Soil Carbon (RaCA) database to examine the performance of the three approaches. The RaCA database contains 32,084 pedons and 144,833 samples from 6148 US locations (Fig. 2; Soil Survey Staff and Loecke, 2016; Wills et al., 2014), so there are approximately five pedons at each location and one or more horizons for each pedon. The pedons and COMPs were excluded from the test dataset if they: (i) did not have soil texture information, (ii) had missing soil horizons between the surface and lower boundaries, (iii) had no slope information, or (iv) had no location (latitude and longitude) information. Data used as inputs included the spatial location (for all of the three approaches) and field-determined texture by depth, depth to the restrictive layer (if available), and slope (which was necessary for the observation-based approach).

The Nash–Sutcliffe efficiency (NSE) was used to evaluate and compare the performance of different approaches. This was defined as (Nash and Sutcliffe, 1970):

$$\text{NSE}=1-\left[\frac{\sum_{n=1}^{m}\left(X_n-Y_n\right)^2}{\sum_{n=1}^{m}\left(X_n-\bar{X}\right)^2}\right], \qquad [11]$$

where m is the number of pedons; X_n and Y_n are the observed and predicted soil properties (e.g., representative sand percentage) for pedon n, respectively; and $\bar{X}$ is the mean value of observed soil properties from the RaCA pedons. The NSE ranges from ∞ to 1.0 with NSE = 1 indicating a perfect match between the observed and predicted properties. The performance of the three approaches was graded using the following criteria: “very good” if the NSE was greater than 0.65, “good” if the NSE was between 0.65 and 0.50, “satisfactory” if the NSE was between 0.50 and 0.30, and “unsatisfactory” if the NSE was less than 0.30 (Moriasi et al., 2007).

The NSE was calculated separately for each soil property (i.e., soil texture by depth, slope, and depth to the restrictive layer). For the DC-based and observation-based approaches, NSE was calculated of the basis of the predicted and observed soil properties. For the location-based approach, the exponential coefficient (i.e., β in Eq. [2]) was unknown and was first estimated by maximizing the total NSE for sand percentage, clay percentage, slope, and depth to the restrictive layer. After β was estimated, the NSE for the location-based approach was then calculated on the basis of the predicted and observed soil properties.

RESULTS AND DISCUSSION

Comparisons among Approaches

For all of the pedons, the observation-based approach performed best among the three approaches used for identifying COMPs whose properties most closely resembled the properties of the test RaCA pedons (Table 2, Table 3). This was a result of the site-specific input data (i.e., slope and soil texture by depth) required by the observation-based approach to constrain soil identifications. As noted above, these properties were selected because they are relatively easy and reasonable (compared with other soil properties) to collect in the field. For the observation-based approach, ~57% of the calculated NSEs fell within the performance category “very good” (NSE > 0.65), ~18% within the category “good” (0.5 < NSE ≤ 0.65), 2% within the category “satisfactory”, and ~23% within the category “unsatisfactory” (Table 2, Table 3).

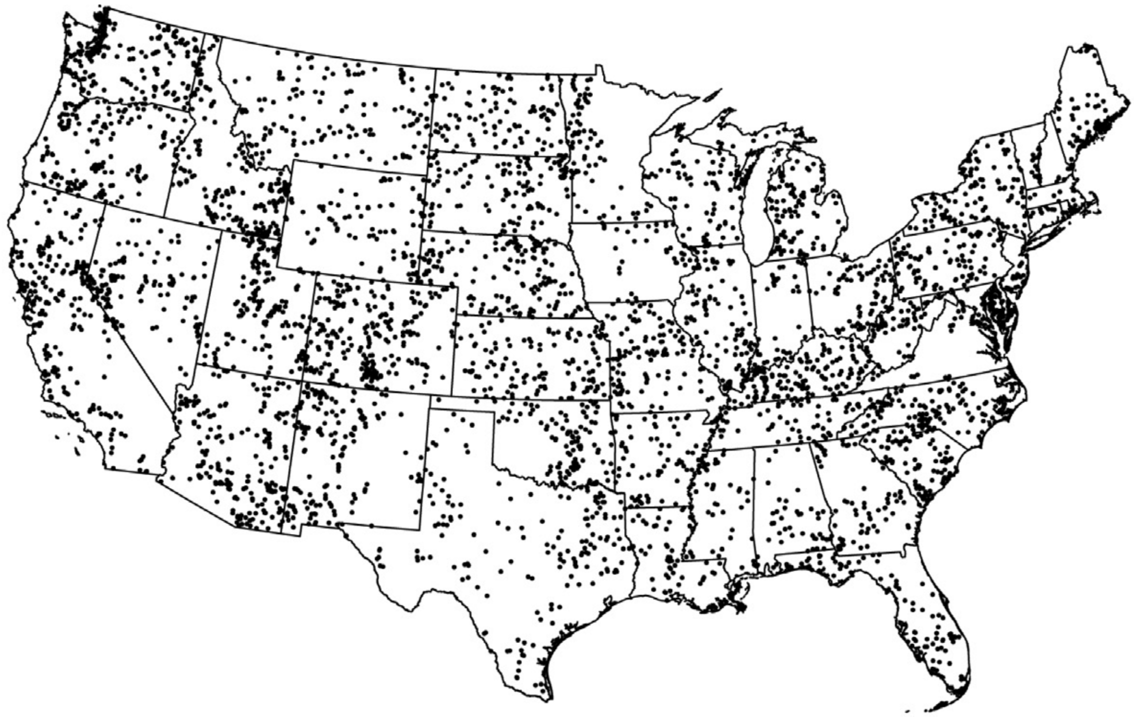

Fig. 2. Spatial distribution of the pedons in the Rapid Assessment of US Soil Carbon (RaCA) database. Each dot denotes a location and each location has approximately five pedons.

The performance of the location- and DC-based approaches, overall, were similar to each other and significantly lower than that of the observation-based approach. For the location-based approach, ~2% of the calculated NSEs fell within the performance category "very good" (NSE > 0.65), ~17% within the category "good" (0.5 < NSE ≤ 0.65), ~20% within the category "satisfactory" (0.3 < NSE ≤ 0.5), and 63% within the category "unsatisfactory" (NSE ≤ 0.3) (Table 2, Table 3). For the DC-based approach, ~2% of the calculated NSEs (except for depth to the restrictive layer) fell within the performance category "very good", ~14% within the category "good", ~14% within the category "satisfactory", and ~70% within the category "unsatisfactory" (Table 2, Table 3).

For areas where the dominant COMP was <51% of the map unit, the location-based approach was slightly better than the DC-based approach (Table 2) when the pedons were close to the SMU boundaries (i.e., ≤15 m). The location- and DC-based approaches performed similarly when the pedons were further from the SMU boundaries (i.e., >15 m). In contrast, when the dominant COMP comprised >51% of a SMU, the location-based and DC-based approaches performed similarly (Table 3) no matter whether the pedons were close to or away from the SMUP boundaries (Table 3).

The RaCA dataset has some strengths and limitations when assessing accuracy. The strengths included the fact that the pedons were geographically dispersed and represented all major soil groups and most of the land use and cover classes (Soil Survey Staff and Loecke, 2016). However, the pedons were selected to represent the dominant component in the SMUP in which they occurred and thus represented the most likely and most common conditions to be observed. Therefore, the dataset was biased to favor the DC-based approach.

Another limitation was that the sample data only included basic descriptions and texture class information without detailed laboratory analysis (the implications of this are discussed below). Although soil series names were assigned to each RaCA pedon that might correspond with COMPs in the CoSSGO database, initial exploratory analysis indicated that the comparison was not informative. This is probably because of incomplete pedon information for soil series identification (e.g., pedons were only sampled to 1 m) and because of changes in soil series and COMPs between the time when the RaCA pedons were sampled and the time when the CoSSGO spatial and tabular data were obtained.

Comparison among the Soil Properties

All three approaches more accurately predicted the slope and sand content of the pedons than the clay content and were unable to "satisfactorily" (i.e., NSE ≤ 0.3) match the depth to the restrictive layer (Table 2, Table 3). This might be partially because the depth to the restrictive layer can be difficult to determine and tends to vary significantly within short distances (e.g., a few meters), whereas the slope tends to be relatively invariable within short distances and easier to measure with a high degree of accuracy (compared with depth to the restrictive layer and soil texture). This results in a better match between slope observations and COMP slope by the three approaches.

In addition, the representative value of sand and clay from the soil texture classes (instead of the actual sand and clay contents; Table 1) were used by the observation-based approach to

Table 2. The calculated Nash–Sutcliffe efficiency (NSE) for the pedons where the dominant component coverage in the home map units is <51% for pedons <5 to 100 m from the nearest map unit boundary†.

	DC-based §				Location-based					Observation-based			
x‡	Sand	Clay	Slope	DtoR	Sand	Clay	Slope	DtoR	β	Sand	Clay	Slope	DtoR
m	———%———			cm	———%———			cm		———%———			cm
<5	0.26	<0	<0	<0	0.33	0.04	<0	<0	-0.483	**0.67**	**0.72**	**0.79**	<0
5–10	<0	<0	0.29	<0	<0	<0	0.52	<0	-0.051	**0.70**	0.51	**0.78**	<0
10–15	0.07	0.16	<0	<0	0.31	<0	0.37	<0	-0.035	**0.80**	**0.73**	**0.72**	0.23
15–20	<0	<0	**0.67**	<0	<0	<0	**0.67**	<0	-0.052	0.60	0.60	**0.78**	<0
20–25	0.20	<0	0.52	<0	0.20	<0	0.53	<0	-0.062	**0.67**	0.54	**0.68**	<0
25–50	0.27	0.00	0.52	<0	0.28	0.01	0.51	<0	-0.031	**0.81**	**0.67**	**0.76**	0.01
50–100	0.33	0.07	0.02	<0	0.40	0.23	0.36	<0	-0.024	**0.80**	**0.79**	0.60	0.10

† The bold, single underlined, and dotted underlined numbers indicate performance ratings of "very good" (NSE > 0.65), "good" (0.5 < NSE ≤ 0.65), and "satisfactory" (0.3 < NSE ≤ 0.5), respectively. The remaining numbers indicate "unsatisfactory" performance (NSE ≤ 0.3).

‡ The shortest distance from a given location to the bounds of its neighboring map unit polygons.

§ DC, dominant component; DtoR, depth to the restrictive layer; β, exponential coefficient [Eq. 2].

perform soil profile matching and to evaluate the performance of the three approaches (i.e., calculating the NSE). Using the texture class to identify the representative sand and clay contents may reduce the variation in actual sand and clay contents. For example, two different soils that have different actual sand and clay contents may have the same representative sand and clay content, because the two soils fall within the same soil texture class. This results in a relatively better match of soil texture – compared with depth to the restrictive layer – between the observations and COMPs by the three approaches.

Moreover, the soil texture class in the RaCA database was determined by well-trained soil scientists. The soil texture class determined by non-soil-scientists would be expected to have lower accuracy than that in the RaCA database; therefore, the performance of the observation-based approach might be lower in reality when the soil texture class is determined by non-soil-scientists.

Implications for Modeling

One of the most challenging issues facing most (if not all) ecosystem and crop models is the inaccuracy of site-specific soil information (e.g., soil texture) (Luo et al., 2012), which substantially affects the quality of model simulations and projections. This is because soil information is critical for initializing, calibrating, or validating these models at site, ecosystem, and regional scales. The observation-based approach greatly improves the accuracy of COMP identification. The identified COMPs can then be used to retrieve and estimate other important soil properties associated with components within a soil survey that are difficult or time-consuming to measure (e.g., soil organic C content). Therefore, using user-provided texture and our approaches to generate more accurate soil predications for use with ecosystem and crop models has great potential to improve the reliability of model simulations and predictions by substantially reducing the uncertainties of the representations of soil properties as model inputs or validation datasets.

Implications for Increasing the Use of Soil Information by Land Managers

The general public (including farmers, gardeners, natural resource managers, and policymakers) requires soil information to make decisions regarding land suitability, productivity, profitability, stability, and sustainability. These approaches not only are easily implemented and deployed on smartphones and other mobile devices but they also do not require the general public to have an extensive knowledge of soil science or soil surveys to collect the necessary data that drive the approaches. With the rapid growth in the number of smartphone users, these approaches and any subsequent mobile applications developed from these approaches could be a powerful tool for the general public to identify the correct

Table 3. The calculated Nash–Sutcliffe efficiency (NSE) for the pedons where the dominant component coverage in the home map units is >51% for pedons <5 to 100 m from the nearest map unit boundary†.

	DC-based §				Location-based					Observation-based			
x‡	Sand	Clay%	Slope	DtoR	Sand	Clay	Slope	DtoR	β	Sand	Clay	Slope	DtoR
m	———%———			cm	———%———			cm		———%———			cm
<5	0.31	0.16	<0	<0	0.31	0.18	<0	<0	-0.173	**0.71**	0.61	**0.79**	<0
5–10	0.60	0.10	0.35	<0	0.60	0.10	0.35	<0	-0.081	**0.80**	0.66	**0.76**	0.41
10–15	0.52	0.02	0.51	<0	0.52	0.01	0.51	<0	-0.043	**0.78**	0.59	**0.79**	0.30
15–20	0.35	0.16	0.16	<0	0.36	0.18	0.18	<0	-0.012	**0.74**	**0.71**	**0.76**	<0
20–25	0.56	0.24	0.33	<0	0.56	0.24	0.33	<0	-0.031	**0.84**	**0.68**	0.59	<0
25–50	0.48	0.29	0.56	<0	0.48	0.29	0.56	<0	-0.021	**0.79**	**0.70**	**0.68**	<0
50–100	0.48	0.32	0.65	<0	0.48	0.32	0.65	<0	-0.013	**0.79**	**0.70**	**0.79**	<0

† The bold, single underlined, and dotted underlined numbers indicate performance ratings of "very good" (NSE > 0.65), "good" (0.5 < NSE ≤ 0.65), and "satisfactory" (0.3 < NSE ≤ 0.5), respectively. The remaining numbers indicate "unsatisfactory" performance (NSE ≤ 0.3).

‡ The shortest distance from a given location to the bounds of the neighboring map unit polygons.

§ DC, dominant component; DtoR, depth to the restrictive layer; β, exponential coefficient [Eq. 2].

COMP accurately and thus make efficient and effective decisions related to environmental, agricultural, and ecological issues.

CONCLUSIONS

Accurate and accessible soil information is critical for natural resource management. Here, we developed two approaches that can be used to identify COMPs with either location alone, or location together with soil properties that are easily determined in the field. Our results indicated that the observation-based approach could significantly increase the reliability of COMP identification by using a small set of easily-collected field data, suggesting that the observation-based approach should be used whenever possible. However, if field observations of soil texture and slope are unavailable, the location-based approach may be useful for slightly improving the predictions (relative to simply using the dominant COMP) for locations that are close (<15 m) to the SMUP boundaries, although the performance of both the location- and DC-based approaches were significantly lower and less satisfactory than that of the observation-based approach.

Further improvements of the location-based approach are likely to be made possible by including some readily available raster environmental covariates (e.g., high-resolution digital elevation and vegetation maps; Nauman and Duniway, 2016; Zhu et al., 2010), which might be helpful for further constraining the selection and identification of COMPs without field observations. Similarly, with the rapid development of mobile sensors (e.g., Delgado et al., 2013; Gómez-Robledo et al., 2013; Han et al., 2016; Zerger et al., 2010), more observations (e.g., soil color and pH) could potentially be included to further improve the performance of the observation-based approach (for example, further differentiating soil components that have overlapping soil properties) as collecting those data with smartphones and other mobile devices becomes easier in the near future.

ACKNOWLEDGMENTS

This research was based on work supported by the United States Agency for International Development, which is supporting the development of a global Land-Potential Knowledge System in cooperation with the USDA-ARS. Portions of work were supported by the US Geological Survey Ecosystems Mission Area. The authors express their sincere gratitude the Jornada Experimental Range, New Mexico State University, and the Sustainability Innovation Lab at Colorado for providing logistical support and coordination of this study. We also thank Shawn W. Salley (USDA-ARS) for constructive suggestions in the preparation of the manuscript. The use of trade, product, industry, or firm names is for descriptive purposes only and does not imply endorsements by the US government.

REFERENCES

Adhikari, K., and A.E. Hartemink. 2016. Linking soils to ecosystem services—A global review. Geoderma 262:101–111. doi:10.1016/j.geoderma.2015.08.009

Anderson, R.M., V.I. Koren, and S.M. Reed. 2006. Using SSURGO data to improve Sacramento model a priori parameter estimates. J. Hydrol. 320:103–116. doi:10.1016/j.jhydrol.2005.07.020

Beaudette, D.E., and A.T. O'Geen. 2009. Soil-Web: An online soil survey for California, Arizona, and Nevada. Comput. Geosci. 35:2119–2128. doi:10.1016/j.cageo.2008.10.016

Beaudette, D.E., P. Roudier, and A.T. O'Geen. 2013. Algorithms for quantitative pedology: A toolkit for soil scientists. Comput. Geosci. 52:258–268. doi:10.1016/j.cageo.2012.10.020

Delgado, J.A., K. Kowalski, and C. Tebbe. 2013. The first nitrogen index app for mobile devices: Using portable technology for smart agricultural management. Comput. Electron. Agric. 91:121–123. doi:10.1016/j.compag.2012.12.008

Gatzke, S.E., D.E. Beaudette, D.L. Ficklin, Y. Luo, A.T. O'Geen, and M. Zhang. 2011. Aggregation strategies for SSURGO data: Effects on SWAT soil inputs and hydrologic outputs. Soil Sci. Soc. Am. J. 75:1908–1921. doi:10.2136/sssaj2010.0418

Gómez-Robledo, L., N. Lopez-Ruiz, M. Melgosa, A.J. Palma, L.F. Capitan-Vallvey, and M. Sanchez-Maranon. 2013. Using the mobile phone as Munsell soil-colour sensor: An experiment under controlled illumination conditions. Comput. Electron. Agric. 99:200–208. doi:10.1016/j.compag.2013.10.002

Han, P., D. Dong, X. Zhao, L. Jiao, and Y. Lang. 2016. A smartphone-based soil color sensor: For soil type classification. Comput. Electron. Agric. 123:232–241. doi:10.1016/j.compag.2016.02.024

Herrick, J.E., A. Beh, E. Barrios, I. Bouvier, M. Coetzee, D. Dent, et al. 2016. The land-potential knowledge system (landpks): Mobile apps and collaboration for optimizing climate change investments. Ecosyst. Health Sustain. 2:3. doi:10.1002/ehs2.1209

Hudson, B.D. 1992. The soil survey as paradigm-based science. Soil Sci. Soc. Am. J. 56:836–841. doi:10.2136/sssaj1992.03615995005600030027x

Luo, Y., J.T. Randerson, G. Abramowitz, C. Bacour, E. Blyth, N. Carvalhais, et al. 2012. A framework for benchmarking land models. Biogeosciences 9:3857–3874. doi:10.5194/bg-9-3857-2012

Miller, D.A., and R.A. White. 1998. A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. Earth Interact. 2:1–26. doi:10.1175/1087-3562(1998)002<0001:ACUSMS>2.3.CO;2

Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50:885–900. doi:10.13031/2013.23153

Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models part I—A discussion of principles. J. Hydrol. 10:282–290. doi:10.1016/0022-1694(70)90255-6

Nauman, T.W., and M.C. Duniway. 2016. The automated reference toolset: A soil–geomorphic ecological potential matching algorithm. Soil Sci. Soc. Am. J. 80:1317–1328. doi:10.2136/sssaj2016.05.0151

O'Geen, A.T., M. Walkinshaw, and D.E. Beaudette. 2017. SoilWeb: A multifaceted interface to soil survey information. Soil Sci. Soc. Am. J. 81:853–862. doi:10.2136/sssaj2016.11.0386n

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

Sanchez, P.A., S. Ahamed, F. Carre, A.E. Hartemink, J. Hempel, J. Huising, et al. 2009. Digital soil map of the world. Science 325:680–681. doi:10.1126/science.1175084

Soil Survey Staff and T. Loecke. 2016. Rapid assessment of carbon: Methodology, sampling, and summary. USDA-NRCS. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/home/?cid=nrcs142p2_054164#citation (accessed 29 Mar. 2018).

Subburayalu, S.K., I. Jenhani, and B.K. Slater. 2014. Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. Geoderma 213:334–345. doi:10.1016/j.geoderma.2013.08.018

Vos, C., C.V. Axel Don, R. Prietz, A. Heidkamp, and A. Freibauer. 2016. Field-based soil texture estimate could replace laboratory analysis. Geoderma 267:215–219. doi:10.1016/j.geoderma.2015.12.022

Wills, S., T. Loecke, C. Sequeira, G. Teachman, S. Grunwald, and L.T. West. 2014. Overview of the U.S. rapid assessment project: Sampling design, initial summary and uncertainty estimates. In: A.E. Hartemink and K. McSweeney, editors, Soil carbon. Springer International Publishing, Dordrecht, The Netherlands. p. 95–104.

Zerger, A., R.A.V. Rossel, D.L. Swain, T. Wark, R.N. Handcock, V.A.J. Doerr, et al. 2010. Environmental sensor networks for vegetation, animal and soil sciences. Int. J. Appl. Earth Obs. Geoinf. 12:303–316.

Zhu, A.-X., F. Qi, A. Moore, and J.E. Burt. 2010. Prediction of soil properties using fuzzy membership values. Geoderma 158:199–206. doi:10.1016/j.geoderma.2010.05.001