Geoderma 397 (2021) 114998

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Regional ensemble modeling reduces uncertainty for digital soil mapping

Colby Brungard [a,*], Travis Nauman [b], Mike Duniway [b], Kari Veblen [c], Kyle Nehring [c], David White [d], Shawn Salley [e], Julius Anchang [a]

[a] Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM 88003, USA
[b] US Geological Survey, Southwest Biological Science Center, Moab, UT 84532, USA
[c] Department of Wildland Resources, Utah State University, Logan, UT 84322, USA
[d] Las Cruces Soil Survey Office, Natural Resources Conservation Service, United States Department of Agriculture, Las Cruces, NM 88005, USA
[e] USDA-NRCS National Ecological Site Team, Jornada Experimental Range, Las Cruces, NM 88005, USA

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:
Machine learning
Regionalization
Soil survey
Major land resource areas
Ecoregions
Landforms

ABSTRACT

Recent country and continental-scale digital soil mapping efforts have used a single model to predict soil properties across large regions. However, different ecophysiographic regions within large-extent areas are likely to have different soil-landscape relationships so models built specifically for these regions may more accurately capture these relationships relative to a 'global' model. We ask the question: Is a single 'global' model sufficient or are regionally-specific models useful for accurate digital soil mapping? We test this question by modeling soil depth classes across the 432,000 km^2 upper Colorado River Basin in the Western USA using a single global model, multiple ecophysiographic models, and ensembles of the ecophysiographic models.

Effective soil depth class observations (n = 12,194) were derived from multiple soil databases. Fifty-seven environmental covariates were derived from a 30 m digital elevation model, climate data, satellite imagery, and aeroradiometric data. Three independent land classifications were used to stratify the area. Two expert-derived land classifications, USDA Major Land Resource Areas (MLRA) and US-EPA Level III ecoregions, divided the study area into multiple ecophysiographic regions based on vegetation and broad-scale physiographic differences. The third land classification divided the study area into broad landforms.

Soil depth observations were split into separate training (n = 10,470) and validation (n = 1,724) datasets. First, a 'global' random forest model was used to model soil depth classes using all training observations and covariates. 'Global' denotes a model built with all training data across the extent of the area, not a model at world extent. Second, the land classifications were used to subset the observations into ecophysiographic sub-datasets and random forest models were refit for each region. Models fit by ecophysiographic region are referred to as regional models. Thirdly, predictions from each regional model were fused into regional-ensemble models. Accuracy, Brier scores, and Shannon's entropy were used to compare model accuracy and uncertainty. Regional ecophysiographic models were also compared to models built for geographic areas that were defined solely to be approximately equal in area. Training dataset density and the imbalance ratio were investigated to determine if data characteristics influenced regional accuracy/uncertainty metrics.

Accuracy for the global model using the validation set was 62.8%. Regional model accuracies ranged between 56.1% and 75.0%. We found: 1) useful inter-regional differences in global model accuracy were revealed when the global model was validated by region, 2) no consistent relationship between training observation density and accuracy/uncertainty metrics, 3) no meaningful differences in accuracy and uncertainty metrics between physiographic and geographic regions, 4) ensembles of regionally-specific models were approximately as accurate as global models, and 5) both region-specific models and ensembles of regional models were less uncertain than the global model. Overall, we recommend the use of soil depth class predictions made from MLRA regional ensemble models because this prediction had higher accuracy than the ecoregion ensemble model prediction, but lower uncertainty than both the global model and the landform ensemble model predictions. We answer our question: Ensembles of regionally-specific models are approximately as accurate as global models, but result in less uncertainty.

* Corresponding author.
E-mail address: cbrung@nmsu.edu (C. Brungard).

https://doi.org/10.1016/j.geoderma.2021.114998
Received 25 September 2020; Received in revised form 20 January 2021; Accepted 3 February 2021
Available online 31 March 2021

1. Introduction

Current country, continental, and global scale digital soil mapping (DSM) approaches have generally used a single model to predict soil properties and classes across the entire region of interest. For example, at the country scale, Adhikari et al. (2013) and Adhikari et al. (2014) developed soil texture and soil organic carbon (SOC) content predictions for Denmark using a Cubist model for each property, Ramcharan et al. (2018) used a single model for predicting soil classes using a random forest and properties with boosted regression trees across the continental USA, and Padarian et al. (2017) used a single regression tree for predicting soil organic carbon contents across Chile. At the continental scale, Viscarra-Rossel et al. (2014) made a baseline map of soil organic carbon in Australia using a single piecewise linear decision tree, Ballabio et al. (2016) modeled top soil texture fractions across Europe using a single multivariate adaptive regression spline model, and Hengl et al. (2015) modeled multiple soil properties across Africa using a single random forest model for each property. At the global scale, Hengl et al. (2014) and Hengl et al. (2017) used linear-regression kriging and an ensemble of machine learners for modeling multiple soil properties and classes.

While many of these studies have achieved at least moderate predictive performance, models over large geographic areas likely encompass significant heterogeneity in soil forming factors and processes and certainly will include areas with fundamentally different soil-landscape-covariate relationships. For example, the USA is a pedologically diverse country (Guo et al., 2003), which encompasses cold soils of alpine tundra environments to warm soils of arid environments, and much in-between. Thus it seems possible that applying DSM by sub-regions within larger geographic areas could constrain soil-landscape relationships relative to the complex multivariate spaces of larger country or continental models that implicitly assume a similar relationship between soil and environmental covariates across all the area being modeled (Guevara et al., 2018). In particular, regional modeling may make it easier to justify model assumptions or to identify and select limited but regionally-specific biophysical variables controlling soil variability (Guevara et al., 2018).

The benefit of region splitting to improve model performance has been recognized for some time (McBratney et al., 1991), but this topic has returned with the application of DSM over larger geographic areas. Recent literature suggests that regionally-specific models may be beneficial for improving DSM predictions. For example, both Mulder et al. (2016) and Ross et al. (2020) found that regional SOC models had much higher predictive accuracy than global models. Guevara et al.

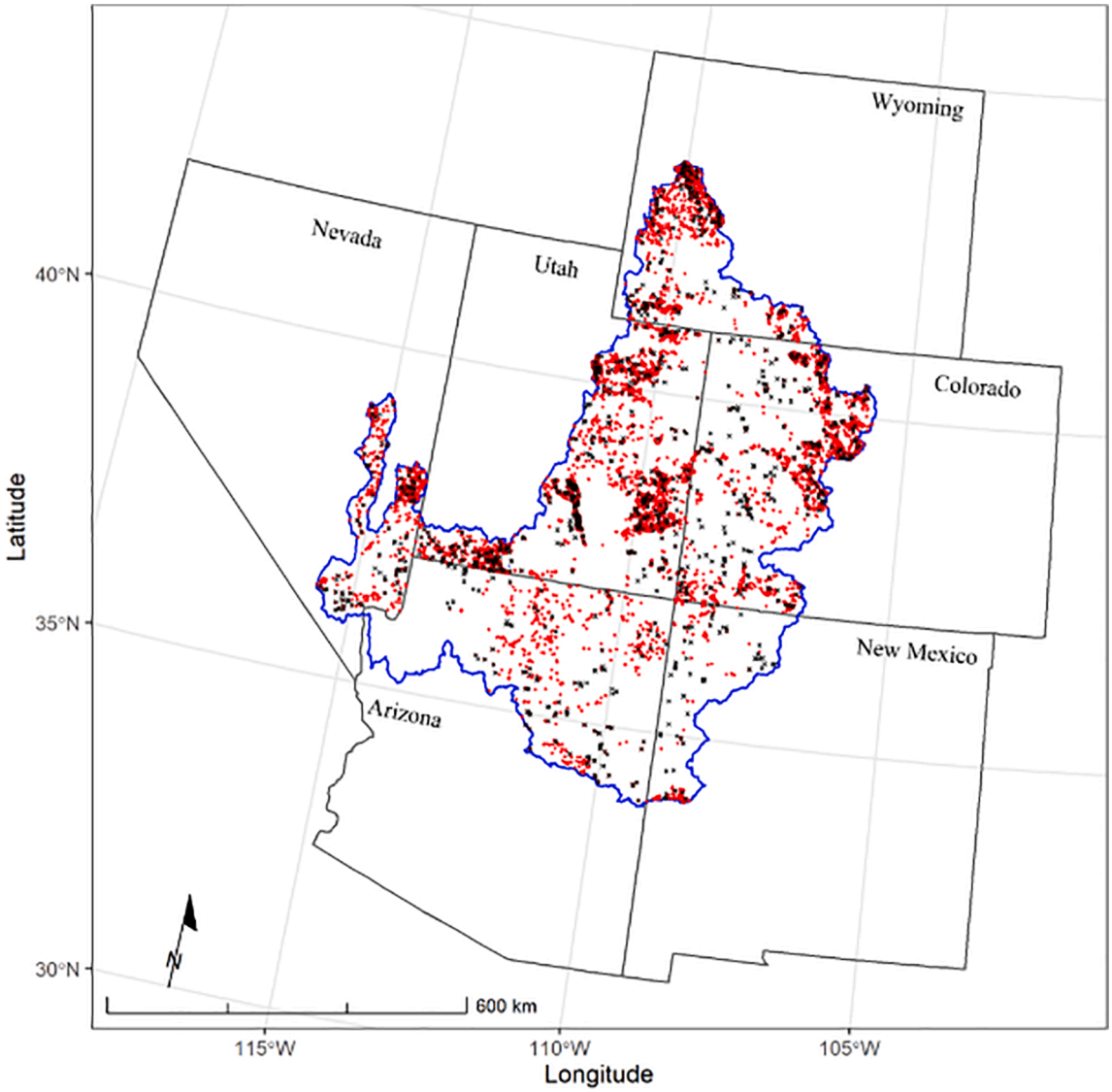

Fig. 1. The Colorado River Basin above Lake Mead in the western USA. Red symbols are training observations. Black symbols are validation observations.

(2018) found that country-specific models produced border artifacts when combined into an ensemble, but suggested that modeling by ecophysiographic region could reduce this effect. At a finer geographic scale, studies that partitioned areas into upland and wetland areas (Peng et al., 2015), or slope-aspect class (Thompson and Kolka, 2005) improved models.

Here, we explore the benefits and limitations of modeling by physiographic regions compared to a single 'global' model for DSM. We define 'global' models as those built using all training observations contained in the entire extent of the study area. We define 'regional' models as those built using only a subset of training observations that are constrained in a specific ecophysiographic portion of the study area. We explore this approach by building a 'global' model for effective soil depth classes in the Colorado River Basin above Lake Mead in the western USA (Fig. 1) and comparing this against regional models built for individual ecophysiographic regions within the study area. We also compared regional ecophysiographic models against arbitrary geographic (i.e., not ecophysiographic) models to test if division by ecophysiographic region constrained model error and uncertainty relative to simple divisions of the dataset. We also build ensembles of ecophysiographic regional models and compare these against the global model. Model accuracy, Brier scores, entropy, predicted spatial patterns, and uncertainty are explored. We demonstrate a reduction in model uncertainty in both individual regional models and ensembles of regional models. Results have application to GlobalSoilMap and Global Soil Partnership organizations that are attempting to develop global soil map products (FAO and ITPS, 2017) and to active national soil survey programs (Thompson et al., 2020).

2. Material and methods

2.1. Study area

The Colorado River Basin above Lake Mead is a 432,000 km^2 watershed that contains a wide variety of elevations, climates, geology, geomorphology, and associated soil systems. The basin starts in the Rocky Mountains at ~ 4300 m and is the water source for 40+ million people in the arid southwestern USA (Udall and Overpeck, 2017). The basin includes the Colorado Plateau, a cool desert area of uplifted sedimentary lithology with large canyon systems and extends into the Mojave Desert basin and range province near Las Vegas. The region is a popular outdoor recreational destination, has extensive energy and mining extraction, and is facing potential mega-drought under current climate trajectories (Ault et al., 2016, Copeland et al., 2017). Both anthropogenic land uses and drought have resulted in negative vegetation changes and increased wind and water erosion which are likely to intensify (Goldstein et al., 2008; Neff et al., 2008; Munson et al., 2011b, 2011a; Miller et al., 2011; Nauman et al., 2018; Fick et al., 2020; among others).

The region includes dominantly public land ownership with some privately owned and Native American lands. The sensitive environmental issues of the region and variety of interests and public entities involved in land management often results in heated debate about land use regulation. Many of the management issues require contextual soil information at a landscape scale (Nauman and Duniway, 2016), but there are still several areas that have not been mapped by the US National Cooperative Soil Survey and existing surveys vary in spatial and thematic detail (Soil Survey Staff, 2019)

2.2. Soil observations

Soil depth is a key soil property as it influences plant and animal habitat (Belcher et al., 1995; Bernard-Verdier et al., 2012; Fuhlendorf and Smeins, 1998; Nussear et al., 2009). In this study, soil depth was defined as effective soil depth which is the distance from the soil surface to a soil horizon where the physical characteristics of the soil strongly inhibit root penetration. The assumption was made that the upper depth of the first horizon designated as having a root restricting layer was the effective depth of the soil. Root restricting horizons as designated by Schoeneberger et al. (2012) are those with a master R (bedrock) or any of the following subordinate designations (cemented (m), densic (d), fragic (x), paralithic (r)).

Soil depth is a censored variable (Chen et al., 2019). That is, for some observations the depth of excavation stops before a physically root restricting horizon is encountered (i.e., excavation stopped at 150 cm even if the total soil depth is $>$ 150 cm). This study converted soil depth into soil depth classes to mitigate the difficulty of censored data and because soil depth classes are often used in land management decision processes (e.g., in identifying vulnerable soils) as a fundamental aspect of land potential-based classification systems used by land and resource managers in the US (Ecological Sites; Duniway et al., 2010). Depth classes were defined as Bedrock (exposed bedrock outcrop), Very Shallow ($<$25 cm), Shallow (25–50 cm), Moderately Deep (50–100 cm), Deep (100–150 cm), and Very Deep ($\geq$150 cm).

Observations of soil depth classes were collated from multiple data sources including the NRCS National Soil Information System and National Cooperative Soil Characterization Database (National Cooperative Soil Survey, 2019), Forest Service Natural Resource Manager (https://www.fs.fed.us/nrm/) Inventory and Mapping database, and multiple independent research projects. Observational data were extensively cleaned to ensure that soil depth classes in each database matched the associated horizon data. Observations were not included in the analysis if they were not excavated to 150 cm and no other information existed that indicated a possible root restricting layer. In addition to soil depth observations from point observations, observations of rock outcrop were hand digitized where rock outcrops were extremely obvious in aerial photos. Additional rock outcrop points were generated from the SSURGO spatial polygon-based database (Soil Survey Staff, 2019) by selecting every map unit that was noted as a consociation and where the first component was designated as 100% rock outcrop. Within each selected map unit, a single bedrock location was then randomly chosen. This resulted in a total of 12,194 observations (including 919 observations of rock outcrop derived from SSURGO and 70 observations derived from visual identification of rock outcrop in aerial photography) available for modeling.

2.3. Training and validation datasets

Observations from the Kellogg Soil Survey Lab in the study area (KSSL, $n = 565$) dataset were reserved as a separate validation dataset because an independent dataset is the gold standard for evaluating model performance (Brus et al., 2011). While not independent, the KSSL dataset is considered the highest quality soil dataset in the USA because of extensive review and represents the best approximation of an independent validation dataset available for this research. However, we supplemented the KSSL dataset with 10% of the observations in each depth class from the available non-KSSL data for two practical reasons: 1) the KSSL data did not contain any rock outcrop observations which would affect validation and uncertainty metrics, and 2) the KSSL data is a relatively small dataset in this study area and a larger validation dataset was valued to provide more robust accuracy metrics. This approach resulted in 10,470 training observations (86% of the entire dataset) and 1724 validation observations (14% of the entire dataset). The spatial distribution of training and test observations are shown on Fig. 1. All reported accuracy and uncertainty metrics were based on the validation dataset and not on cross validation.

2.4. Covariates

Fifty-seven covariates representing six of the seven SCORPAN factors (McBratney et al., 2003) were derived from a 30-m digital elevation model (DEM), satellite imagery, climate datasets, gamma

aeroradiometric datasets, and existing sedimentary deposit thickness maps, (Table 1). Covariates at resolutions > 30 m were resampled to 30-m resolution as needed using bilinear interpolation. All covariates derived from satellite imagery were derived from cloud free images over a 3-year time period (1/1/2006 to 12/31/2009) of Landsat 5 imagery from Google Earth Engine (Gorelick et al., 2017).

Gridded estimates of sedimentary deposit thickness were obtained from Pelletier et al. (2016). Gridded estimates of annual temperature, precipitation, and the ratio of summer (June-September) to annual (precipitation ratio) were derived from the PRISM 2010 30-year normal dataset (PRISM Climate Group, 2010) and downscaled from 800-m resolution with cubic convolution to 30-m and used to capture patterns in climate variables. Additionally, the 3-year median and standard deviation top of atmosphere brightness temperature (Chander et al., 2009) were used as a proxy for land surface temperature and microclimate. The 3-year median and standard deviation NDVI were used to capture spatial vegetation patterns as well as variability in vegetation patterns.

Relative elevation (subtraction of neighborhood minimum from cell value) and relative mean elevation (subtraction of neighborhood mean from cell value) were calculated over neighborhood sizes of 2, 4, 8, 16, 32, 64, and 128 pixel radii using a custom python script. Slope aspect was represented for all both cardinal (N-S, E-W) and both intercardinal (NW-SE, NE-SW) axes by taking the cosine of the cell aspect minus the axis of interest (e.g. southness = cosine[aspect-180]). All other terrain covariates were derived from a 30-m DEM of the project area in SAGA GIS (Conrad et al., 2015). Briefly, a 30-m DEM covering the project area was subset by individual HUC 6 watersheds (USDA-NRCS et al., 2019), terrain derivatives were calculated for each watershed, and then each terrain derivative was mosaicked back together to create seamless derivatives across the study area.

Variations in parent material were represented by Landsat 5 Tier 1 median clay and ferrous indices (Boettinger et al., 2008) and a normalized band ratio of band 5 (short-wave IR) to band 1 (blue) which has been visually observed to highlight differences in sedimentary rock type. Aeroradiometric gamma uranium, thorium, potassium, and absorbed dose were also used to capture broad differences in parent material (Hill et al., 2009).

Table 1
Covariates used for modeling effective soil depth. All covariates were at a 30-m resolution.

SCORPAN factor	Covariate
Soil	Sedimentary deposit thickness
Climate	Annual precipitation and temperature
	Brightness temperature 3-year median and standard deviation
	Precipitation ratio
Organisms	NDVI 3-year median and standard deviation
Relief	Convergence index and mass balance index
	Diurnal anisotropic heating
	Aspect: Eastness, Northwestness, Northeastness, Southness
	Elevation
	Longitudinal, tangential, compound, total, minimum, plan, profile curvatures
	Catchment and modified catchment area
	Multi-resolution valley bottom and ridge top flatness
	Positive openness and protection index
	Relative height (2, 4, 8, 16, 32, 64, & 128 cell radius)
	Relative mean height (1, 2, 8, 16, 32, 64, & 128 cell radius)
	Topographic and Saga wetness indices
	Slope and catchment slope
	Terrain surface convexity
	Topographic position and ruggedness indices
Parent Material	Clay and Ferrous normalized ratios
	γ absorbed dose
	γ potassium, thorium, and uranium
	Normalized ratio SWIR/blue

2.5. Physiographic regions

Three different physiographic-based land classifications were used to divide the study area into ecophysiographic regions: 1) USDA-NRCS Major Land Resource Areas (MLRA) (USDA-NRCS, 2006), 2) US Environmental Protection Agency Ecoregion level III (Omernik and Griffith, 2014), and 3) major landforms (Iwahashi et al., 2018).

The MLRA (Fig. 2) are expert-defined divisions of the land area of the USA based on differences in physiography, geology, climate, water, soils, biological resources, and land use (Salley et al., 2016a) Level III Ecoregions (Fig. 2) are groupings of expert-defined ecosystems based on similarities in soils, physiography, vegetation, land use, geology, and hydrology from 1:7,500,000 scale maps (Omernik and Griffith, 2014). Because the overall study area was defined by the Colorado River Basin watershed above Lake Mead, the project area included narrow sections of several MLRA and Ecoregions around the study area periphery. This resulted in slivers of several MLRA and Ecoregions which had many fewer observations than the other regions that were completely within in the study area. To overcome this imbalance in the number of observations by MLRA/Ecoregion we combined some MLRA and Ecoregions (Table 2). The Central Nevada Basin and Range and Southern Nevada Basin and Range MLRAs were combined with the Great Salt Lake MLRA. The Arizona and NM Mountains MLRA was combined with the Colorado Plateau MLRA. The Southern Rocky Mountain Parks MLRA was joined with the Southern Rocky Mountains MLRA. The Arizona/New Mexico Mountains Ecoregion was combined with the Arizona/New Mexico Plateaus Ecoregion.

Major landforms (Fig. 2) were derived from overview landform groups derived from a 280 m global DEM (Iwahashi et al., 2018). Because of the scarcity of the 'plains' group in the project area, the plains landform class was combined with the 'Plateau, Terrace, and Large lowland slope' class resulting in four broad landform categories (Table 1).

2.6. Geographic areas

The study area was also divided into geographic areas to test the effect of region definitions. As opposed to physiographic regions that attempt to delineate areas of ecosystem similarity, geographic regions are delineated based only on coordinates. These geographic regions were created by dividing the study area into a number of areas equal to the number of MLRA's ($n = 9$), Ecoregions ($n = 8$), and landforms ($n = 4$) based solely on the criteria that regions should be approximately equal in size. The intent of this approach was to dilute any effect of constraining soil-covariate relationships by physiographic region. If modeling by physiographic area does constrain soil-covariate relationships, then models fit by physiographic region should be more accurate than models fit by geographic area. Geographic areas were created using the sppcosa R package (Walvoort et al., 2010).

2.7. Global, regional, and ensemble modeling strategy

We initially assumed that including ecophysiographic regions as covariates would improve model accuracy as it would constrain soil-covariate relationships. A global model using all observations and all covariates (including the MLRA, Ecoregion, and Landform ecophysiographic regions) was initially built and resulted in an accuracy of 62.8%. Physiographic regions were then removed as covariates from the global model but model accuracy remained 62.8%. We concluded that ecophysiographic regions were either unimportant covariates or that the same information was captured by other covariates. This suggests the need to explicitly model by area rather than relying on covariates to capture inter-region variability. Subsequently our modeling approach followed Fig. 3. First the training observations were used to build a global model that covered the entire extent of the study area. Training observations were then split by region and used to tune each regional

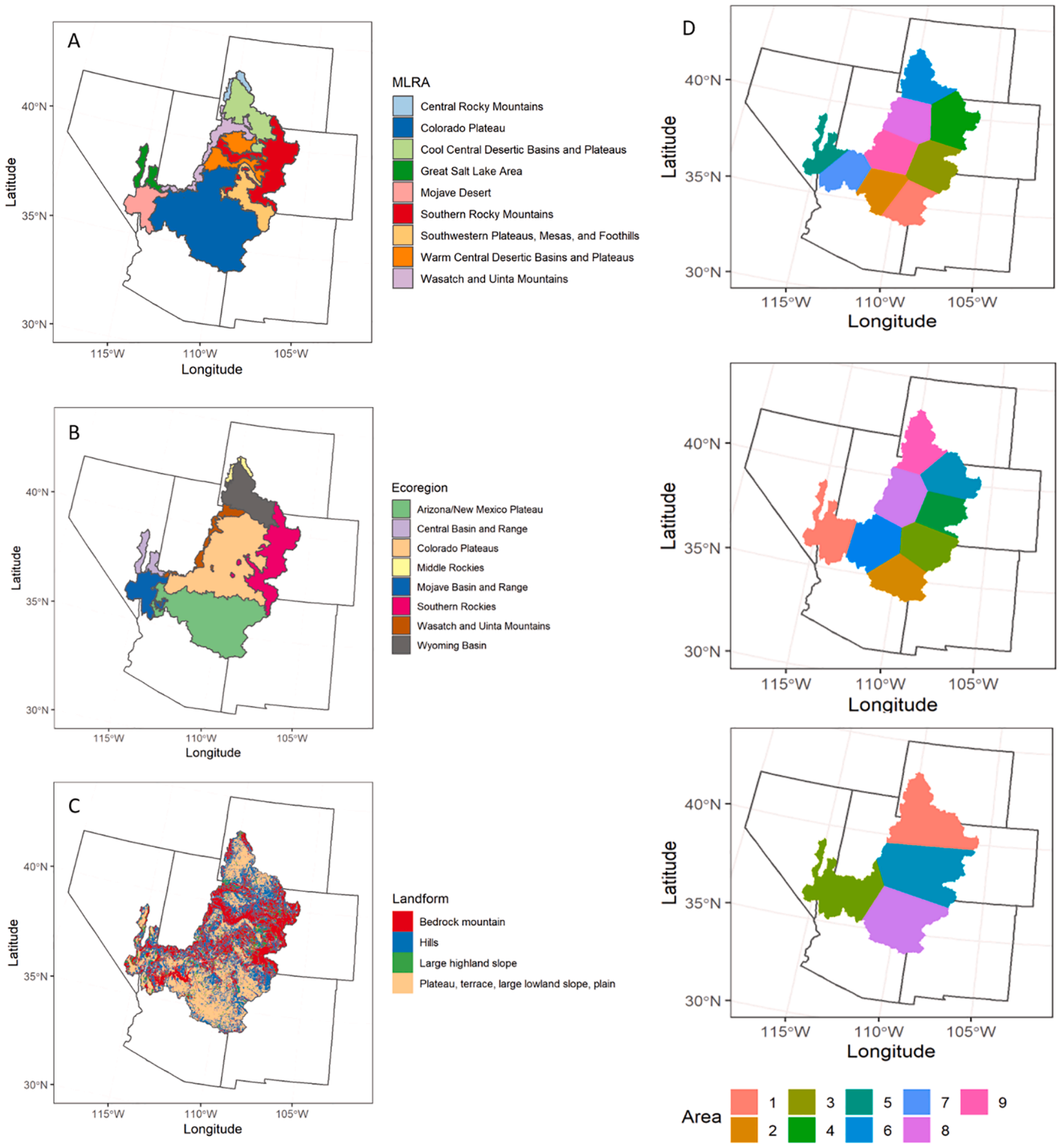

Fig. 2. Physiographic and geographic divisions of the study area. A. Major Land Resource Areas (MLRA's) in the Upper Colorado River Basin. MLRA's are expertly defined physiographic regions based on considering differences in physiography, geology, climate, water, soils, biological resources, and land use. B. Level III ecoregions in the Upper Colorado River Basin. Ecoregions are groupings of expert-defined ecosystems based on similarities in soils, physiography, vegetation, land use, geology, and hydrology from 1:7,500,000 scale maps. C. Broad landform classes created from a 280 m global DEM modified from Iwahashi et al. (2018). D. Geographic areas based on approximately equal area sizes; top = 9 areas, middle = 8 areas, bottom = 4 areas.

model. Random forests (RFs) were used for each model because they are relatively quick, easy to tune, and have been demonstrated to be among the more accurate available models for DSM (Brungard et al., 2015; Hengl et al., 2015). All models were tuned using 10-fold cross validation. The one-standard-error-rule (Hastie et al., 2001; Kuhn and Johnson, 2013), that is the tuning parameters that resulted in model accuracy within one standard error of the model with the absolute highest accuracy, was used to select the final model. This process was repeated for each physiographic region (MLRA, Ecoregion, and Landform) and geographic region.

Regional modeling resulted in multiple models, one for each region. This required joining the models and/or predictions because only a single, most accurate, soil depth class prediction was desired. Initial results demonstrated that applying a regional model (e.g., Colorado Plateau MLRA) to other regions (e.g., Mojave Desert MLRA) resulted in a significant drop in accuracy (up to 15% decrease) and simply

Table 2
Physiographic regions used for regional modeling.

Major Land Resource Areas	Ecoregion	Major Landforms
Great Salt Lake Area	Central Basin & Range	Bedrock mountain
Mojave Desert	Mojave Basin & Range	Hills
Cool Central Desertic Basins & Plateaus	Middle Rockies	Large highland slope
Warm Central Desertic Basins & Plateaus	Wyoming Basin	Plateau, Terrace, Large lowland slope, and Plains
Colorado Plateau	Wasatch & Uinta Mountains	
Southwestern Plateaus Mesas & Foothills	Colorado Plateaus	
Central Rocky Mountains	Southern Rockies	
Wasatch & Uinta Mountains	Arizona & New Mexico Plateau	
Southern Rocky Mountains		

mosaicking the regional predictions resulted in boundary artifacts (results not shown). Thus, we employed an ensemble approach. The utility of ensemble modeling has been demonstrated by (Guevara et al., 2018; Malone et al., 2014; Ramcharan et al., 2018; Taghizadeh-Mehrjardi et al., 2019; Taghizadeh-mehrjardi et al., 2020a; Taghizadeh-Mehrjardi et al., 2020b). However, previous ensemble approaches have modeled continuous (e.g., pH, C) rather than categorical (e.g., soil depth class) data. Additionally, most available ensemble approaches such as caretEnsemble (Deane-Mayer and Knowles, 2019; Guevara et al., 2018) define an ensemble as different forms of statistical models for the same dataset, rather than models that use different training data.

Thus, we took the approach outlined in Fig. 3 to produce an ensemble of soil depth class predictions. First, each regional model was used to predict both the soil depth class and the probability of occurrence of each soil depth class across all regions. The maximum class probability at each cell was then found for each regional model prediction and combined into a new raster stack termed the 'max stack'. Subsequently the maximum value from this 'max stack' was again found, which resulted in a single raster layer where each cell represented the highest probability from any regional model. This was used to quantify prediction confidence, a surrogate measure of uncertainty, where higher values indicate that the regional ensemble models had higher predicted probabilities and were more 'confident' in their predictions. Final class predictions were made by stacking the class predictions from all models and choosing the class with the highest probability. This required a 'model identification' raster which identified which regional model made the highest probability prediction at each cell. The 'model identification' raster also allowed insight into where each model contributed to the final class prediction and prediction confidence.

The above approach was repeated twice; first using the regional

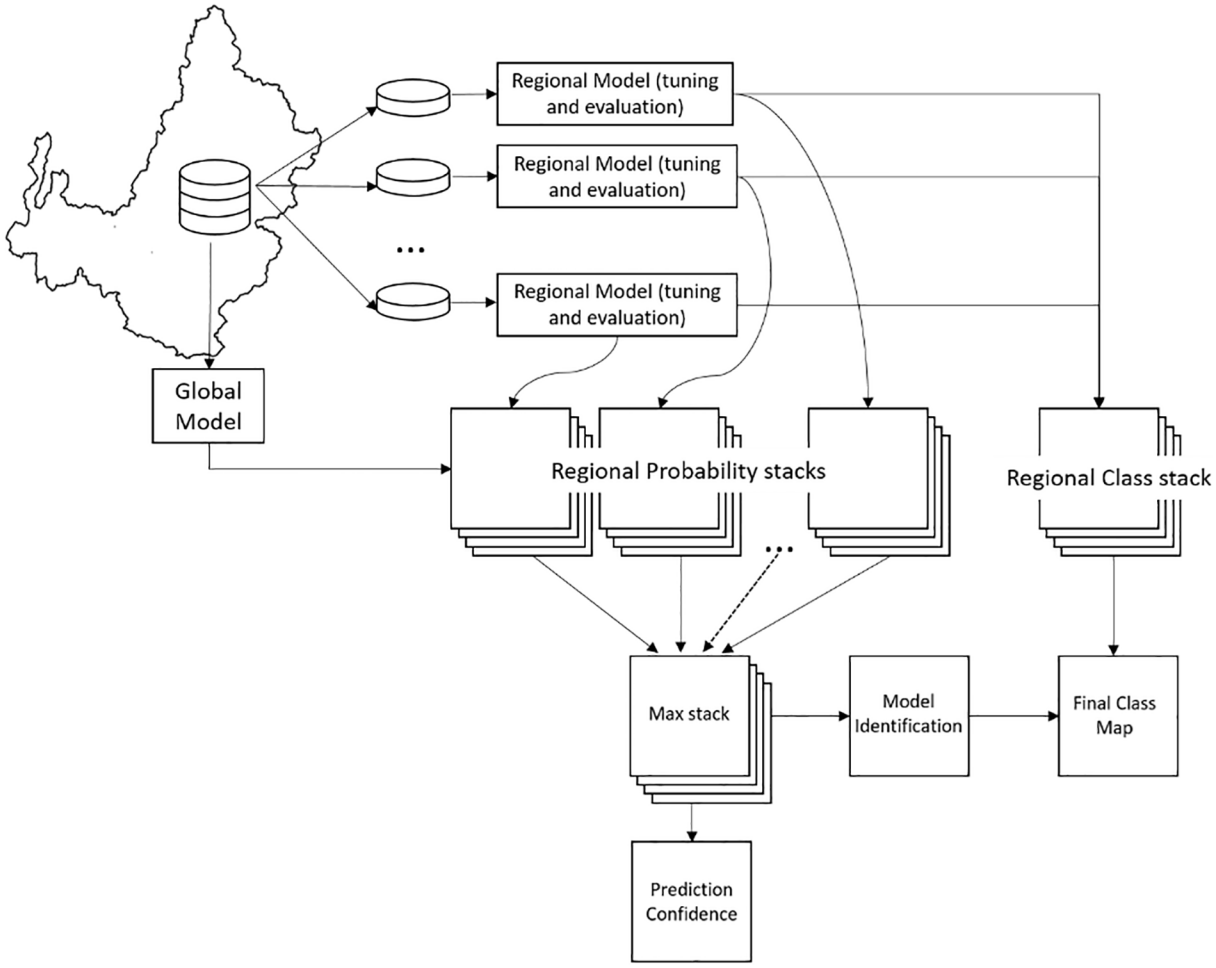

Fig. 3. Regional and global modeling strategy. All observations from the study area were used to build a global model. The observations were then split by region and used to tune and validate regional models. Each regional model was then used to predict soil depth classes and the probability of occurance of soil depth classes across the entire UCRB. The maximum probability for each regional prediction was extracted from each probability stack and combined into a single probability stack. The maximum value was again found from this stack and used to quantify prediction confidence. The model that produced the highest probability was also identified and used to extract the class prediction with the highest probability from the regional class stack.

models only (which is termed the regional prediction ensemble), and secondly using the regional models plus the global model (which is termed the regional + global prediction ensemble).

2.8. Model evaluation

All models, including the global model, were evaluated by region. This was done by dividing the validation dataset into regionally-specific validation sub-datasets, one for each region. Each regional model was then evaluated with the corresponding regional evaluation dataset. The global model was also evaluated using each regionally-specific evaluation dataset. This resulted in two evaluations for each region: a validation based solely on the regional model and an evaluation of the global model in that region. Regional ensemble models were evaluated in the same fashion as the individual regional models.

Three metrics were used to evaluate model performance: overall accuracy, Shannon's entropy (Kempen et al., 2009), and Brier scores (Brungard et al., 2015). Each metric was calculated from the validation observations. Overall accuracy evaluates classification accuracy but can mask nuances in per-class accuracy. Shannon's entropy is a measure of model uncertainty (Hengl et al., 2017). Larger values indicate greater uncertainty. The Brier score is a 'skill' score and quantifies the ability of a model to predict the correct class with high confidence. A lower Brier score (range [0, 1]) indicates greater accuracy with higher confidence. For example, a model that predicts the correct class with a high probability (e.g., 90%) has a lower Brier score than a model that predicted the correct class with a lower probability (e.g., 51%).

Regional and regional + global model ensembles were evaluated using the entire validation dataset. Only overall accuracy was calculated for ensemble predictions because Brier scores and entropy require the probability distribution for each class and the approach to model ensembling taken here preserved only the maximum probability values. We used this approach rather than other model ensemble approaches (Caubet et al., 2019; Guevara et al., 2018; Malone et al., 2014; Taghizadeh-Mehrjardi et al., 2019) because this study was modeling categorical rather than continuous soil properties, and because it was not clear how models trained in one region where physically different soil-covariate relationships were expected could be extended to a different region.

Modeling by ecophysiographic region required splitting the training data into sub-datasets; one for each region. This introduced the possibility that differences in model validation results between regions were a result of differences in training data structure and availability rather than fundamental differences in soil-landscape relationships between ecophysiographic regions. To test if differences in model validation metrics (accuracy, Brier scores, and entropy) were a result of differences in the distribution of training observations, rather than underlying fundamental soil-covariate relationships, we investigated the relationship between model fit metrics and two metrics of training data availability and structure: 1) the density of training observations and 2) the imbalance ratio. The density of training observations was calculated as the number of observations divided by the total area (km^2). Training observation density quantifies differences in the number of available training observations between region. Previous studies have indicated that the frequency distribution of training observations between classes influences predictive accuracy (Brungard et al., 2015; Taghizadeh-mehrjardi et al., 2020a; Taghizadeh-Mehrjardi et al., 2020b). The imbalance ratio (IR) is a simple measure that quantifies differences in the number of observations per class and is calculated as:

$$IR = min_{class}/max_{class}$$

where min_{class} is the number of observations in the class with the fewest observations and max_{class} is the number of observations in the class with the most observations. The imbalance ratio ranges between 0 and 1. A value of 1 indicates that all classes are perfectly balanced, while values closer to zero indicate that one class has more observations than the other classes. If model validation metrics were related to either the training observation density or the IR, then the conclusion would be that regional model validation metrics solely resulted from differences in training data availability and structure rather than fundamental soil-landscape differences in ecophysiographic regions

All modeling and validation was performed using the Rstudio IDE (R Core Team, 2019; RStudio Team, 2018) and the following packages: caret (Kuhn, 2019); tidyverse (Wickham, 2017); raster (Hijmans, 2019); forcats (Wickham, 2019a); reshape2 (Wickham, 2007); measures (Probst, 2018); ggplot2 (Wickham, 2016); sp (Bivand et al., 2013); aqp (Beaudette et al., 2013); soilDB (Skovlin and Roecker, 2019); rgdal (Bivand et al., 2019); ggpubr (Kassambara, 2019), stringr (Wickham, 2019b); rgeos (Bivand and Rundel, 2019); openxlsx (Schauberger and Walker, 2019), ggpubr (Kassambara, 2019), sf (Pebesma, 2018) and rnaturalearth (South, 2017a, 2017b), and Rcolorbrewer (Neuwirth, 2014). All code for this research can be found at: https://github.com/ColbyBrungard/DSM_Regional_Modeling.

3. Results

Global model accuracy was 62.8%. Validation accuracies for the global model, the global model applied to each region, and the regional models are presented in Fig. 4. Three patterns in this figure are evident. First, important regional differences are masked when validation data from the global area is used to assess global model accuracy (compare circles to the dashed line). Secondly, the global model accuracy is approximately the mean accuracy of all regional models so that about half of the regional models are more accurate and about half of the regional models equal or are less accurate than the overall global model accuracy (compare triangles to the dashed line). Thirdly, the accuracy of regional models and the global model applied in each region are approximately equal, although there are notable exceptions such as the "Large highland slope" landform for which the global model was about 10% more accurate than the regional model (compare the triangles to the circles). The general similarity in accuracy between the regional models and the global model in each region is perhaps unsurprising given that each regional model was constructed from a subset of the data used to train the global model.

Brier scores are presented in Fig. 5. In general, regional models had slightly improved (i.e., lower) Brier scores than did the global models; eight of the nine MLRA models and all of the Ecoregion models had equal or better Brier scores compared to the global model applied to that region (compare triangles to circles). However, all Landform regional models had equal or worse (i.e., higher) Brier scores than global models. We are unsure exactly why regional models based on broad landform classes are less skilled than regional models from other ecophysiographic classifications, but it is likely because dividing an area solely by landform neglects important factors such as climate, lithology, or age that help constrain soil-covariate relationships.

Shannon's entropy metrics are shown in Fig. 6. This figure reveals that in general the regional models predicted soil classes with less uncertainty (lower entropy) than did the global model in each region. (compare circles to triangles). Additionally, the majority of MLRA and Ecoregion models had lower uncertainty than the overall global model (compare triangles to dashed line). However, similarly to the Brier scores, modeling by broad landform had equal or worse entropy than global models and we attribute this to the same reasons.

No effect between the density of training observations on model accuracy, Brier scores, or entropy was detected (Fig. 7). Similarly, no consistent relationship between the imbalance ratio and model accuracy, Brier scores, or entropy was detected, except possibly for the regional Landform models (Fig. 7).

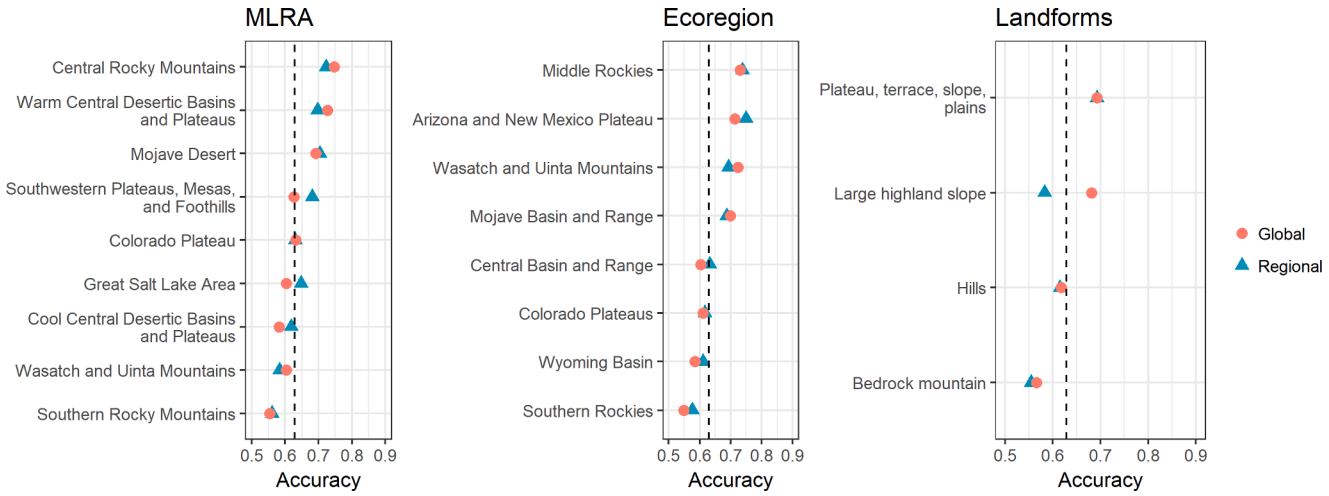

Fig. 4. Model accuracy of physiographic areas as assessed by a separate validation dataset. The red circles are the accuracy of the global model in each region. The blue triangles are the accuracy of each regional model. The dashed line is the overall global model accuracy.

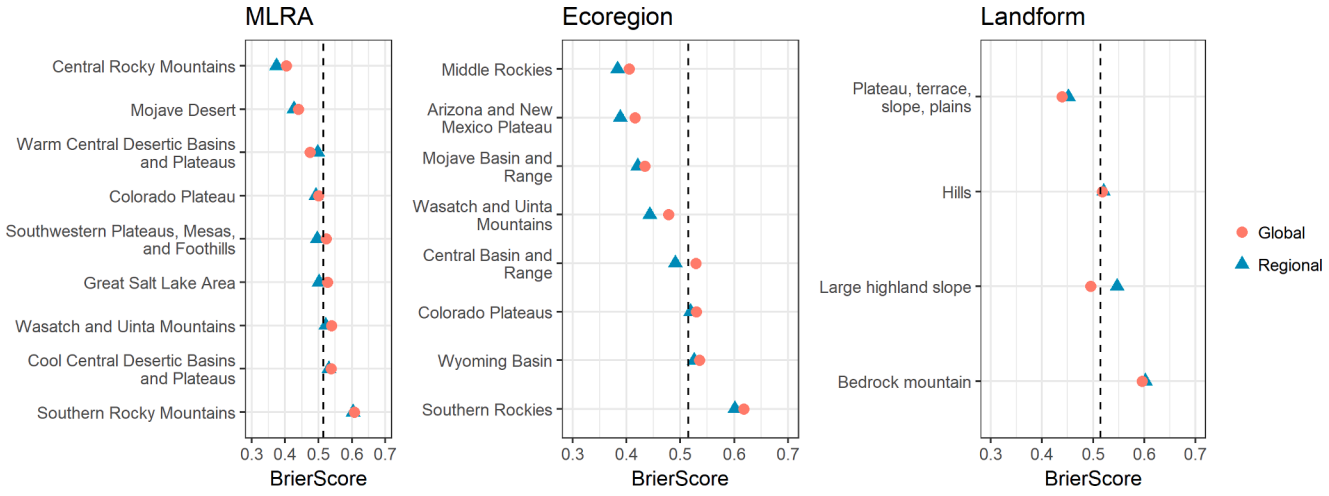

Fig. 5. Brier Scores for physiographic models. Brier scores indicate which model most often predicted the correct class with the highest probability. More confidence can be placed in the model that predicts the correct class with a higher probability than one that predicts the correct class with a lower probability. In this sense the Brier score is a 'skill' score. A model with lower Brier score is more skilled at making the correct predictions.

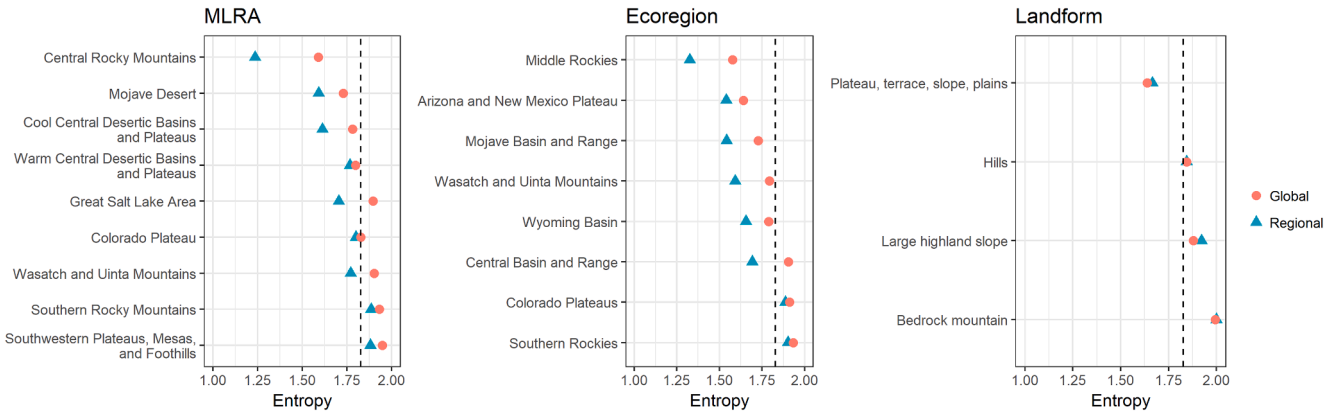

Fig. 6. Shannon's entropy for physiographic models. Larger values indicate greater uncertainty. A value of zero indicates that one class was predicted with a probability of 1 and the other classes were predicted with a probability of 0.

3.1. Physiographic and geographic models

The mean accuracy, Brier scores, and entropy were approximately equal between physiographic and geographic regions regardless of the number of areas (Fig. 8).

3.2. Regional model ensemble accuracy

Ensembles of regional models that included the global model were slightly more accurate on average than ensembles of regional models only (Fig. 9). However, both regional ensembles and regional + global

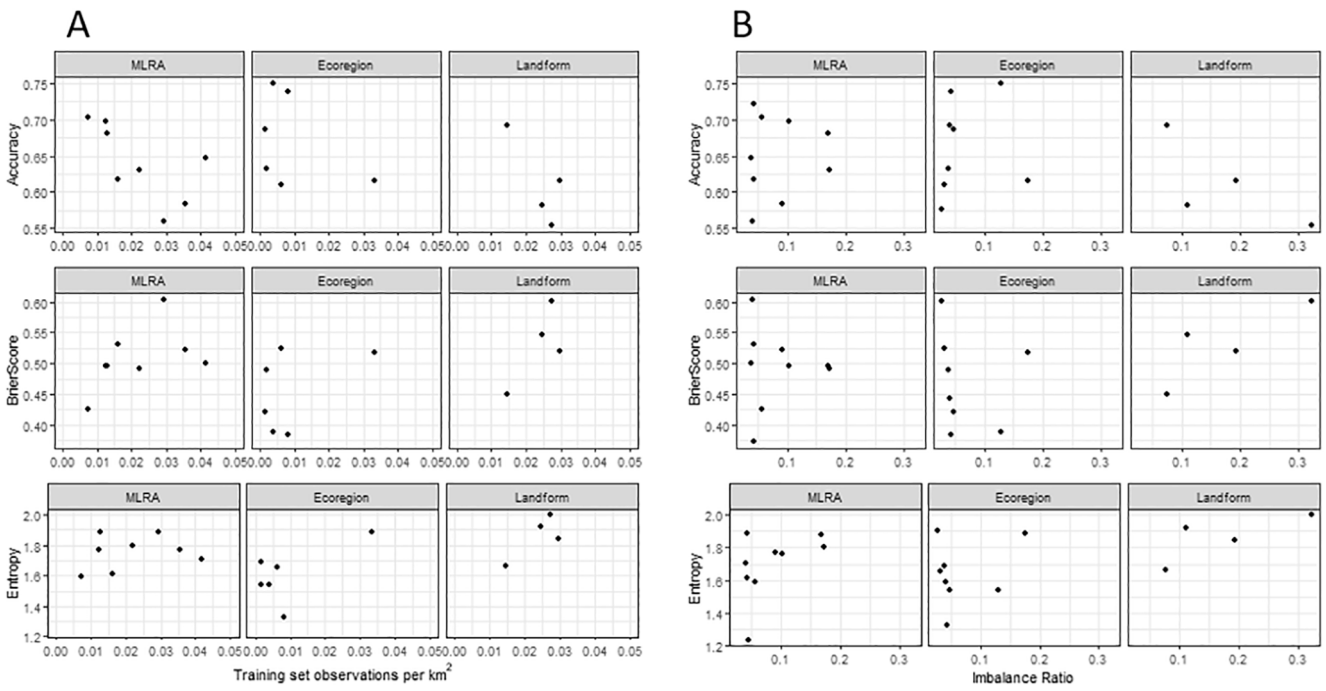

Fig. 7. Training observation density (A) and imbalance ratio (B) plotted against accuracy, Brier scores, and entropy for regional models. The imbalance ratio measures the distribution of observations between classes. An imbalance ratio of 1 indicates that the number of observations between all classes are equal. Points in the plots are the values from the regional models. Note that several ecoregions with training area density > 0.05 observations per km^2 were removed for visual clarity, but this did not affect the overall conclusion that no consistent relationship was evident. No consistent relationship is evident suggesting that training data availability and structure did not affect model performance.

ensembles were slightly less accurate on average than the global model. Full confusion matrices for the global and the regional + global ensemble models are presented in Appendix A.

Average prediction confidence was approximately equal (0.60) for MLRA and Ecoregion regional + global ensemble predictions, while it was less (0.49) for the Landform regional + global ensemble predictions (Fig. 10). It is clear from Figs. 9–11 that while the Landform regional + global ensemble predictions were slightly more accurate on average than the global model or the other ecoregional models, they also produced the predictions with highest uncertainty. The low model confidence results because the regional Landform models had the highest entropy (Fig. 6) which is likely because these landforms classes occur across areas with different lithology and climates.

3.3. Regional model ensemble predictions

Spatial predictions of soil depth classes and associated prediction confidence are presented in Fig. 11. Differences in the spatial patterns of soil depth class predictions are relatively minor between regional model types, although the ecoregion ensemble model predicted a greater spatial extent of moderately deep soils and the landform ensemble model predicted a greater spatial extent of very deep soils compared to the MLRA ensemble.

Regional model identifications are presented in Fig. 12 and show the location where each regional model was used to make the final ensemble prediction. The key point from this figure is that individual regional models contributed to predictions outside of the region in which the model was trained.

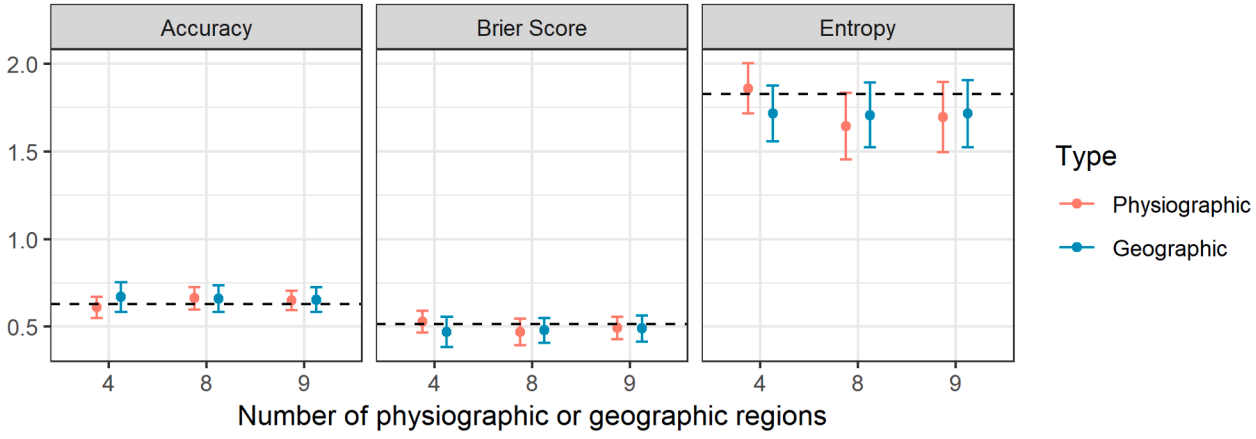

Fig. 8. Comparison of physiographic vs. geographic divisions. Mean and standard deviation of accuracy, Brier scores, and entropy averaged over all physiographic and geographic regions. The x-axis is the number of geographic and physiographic regions (MLRA = 9, Ecoregion = 8, Landform = 4). The dashed lines are the corresponding metrics of the global model.

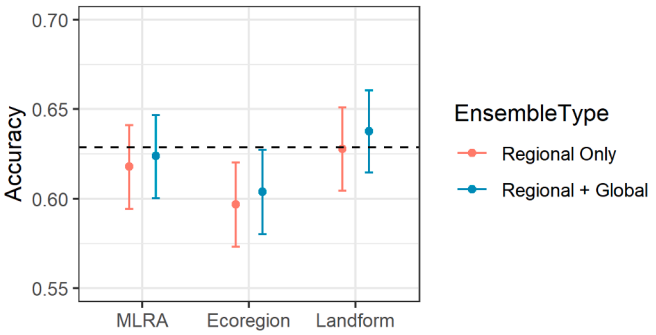

Fig. 9. Regional ensemble prediction accuracy. Regional ensembles were created by selecting the class prediction from the regional model with the highest class probability at each cell. Regional + global models included the global model with the regional models. The dashed line is the global model accuracy.

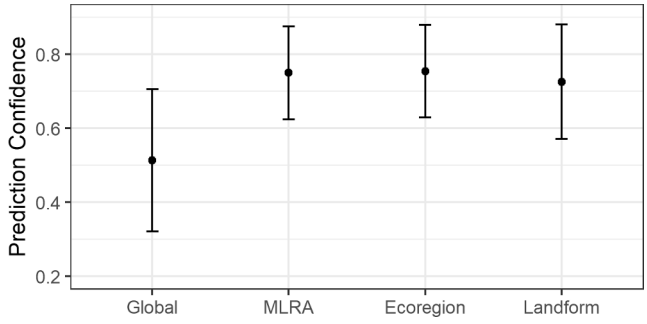

Fig. 10. Prediction confidence average and standard deviation values for regional + global ensemble predictions. Confidence values calculated from validation observations. Lower values indicate lower values of predicted class probability values which are more uncertain.

4. Discussion

4.1. Regional variation in accuracy and uncertainty metrics

We discovered useful inter-regional differences in global model accuracy and uncertainty metrics when the global model was assessed using validation data from individual regions (Fig. 4). While the overall accuracy of the global model was approximately the mean of the global model in each region, in some regions the accuracy of the global model was actually much higher (e.g., Central Rocky Mountains MLRA) or lower (e.g., Southern Rocky Mountains MLRA) than validation of the global model suggested. This finding is also supported by the Brier and entropy scores (Fig. 5 and Fig. 6) where some of the regions had much different metrics than the overall global model would suggest. These findings highlight the importance of validating (or at least cross-validating) models by region as suggested by Brenning (2012); Meyer et al. (2018); and Schratz et al. (2019). Validating models by region has several practical implications for digital soil mapping; primarily the identification of regions with low model accuracy in which additional resources and effort (field sampling, additional covariates, additional modeling efforts) could be preferentially allocated to improve predictive accuracy.

The generally small accuracy differences between the regional models and the global model could result from the specific model used. Random forests recursively partitions feature space into homogenous areas and it could be that other model types that fit different (e.g., curvilinear) relationships could result in greater differences between regional and global models.

4.2. Effect of data distribution

We found that the density of training observations was not predictive of model accuracy or uncertainty (Fig. 7 A) and we interpret these results to indicate that inter-regional differences in accuracy are a result of differences in the strength of soil-covariate relationships because of inherent pedodiversity (Guo et al., 2003). Similarly, recent soil property mapping work in the same study area found a surprisingly low spatial correspondence between model uncertainty and observation density (Nauman and Duniway, 2020). Also, Guevara et al., (2018) found that

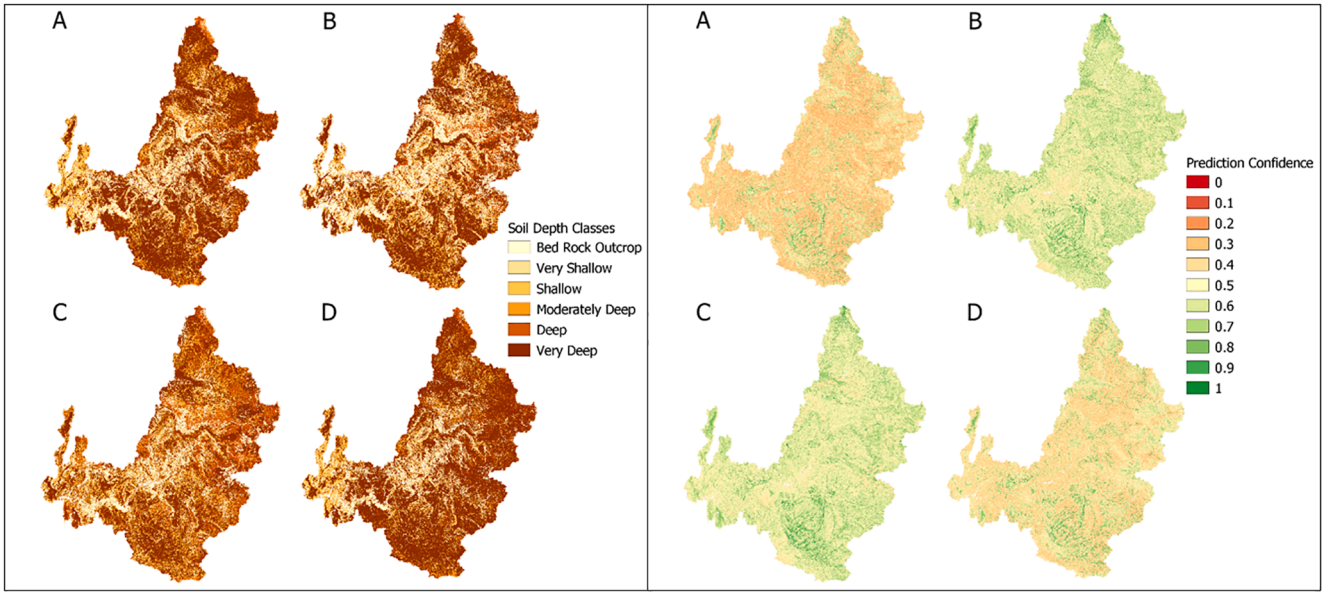

Fig. 11. Soil depth class predictions (left) and associated prediction confidence (right). A. Global model only. B. MLRA regional + global model. C. Ecoregion regional + global model. D. Landform regional + global model. Prediction confidence is derived from the maximum probability value at each cell from any of the regional or the global models. Higher prediction confidence indicates less uncertainty in the class predictions.

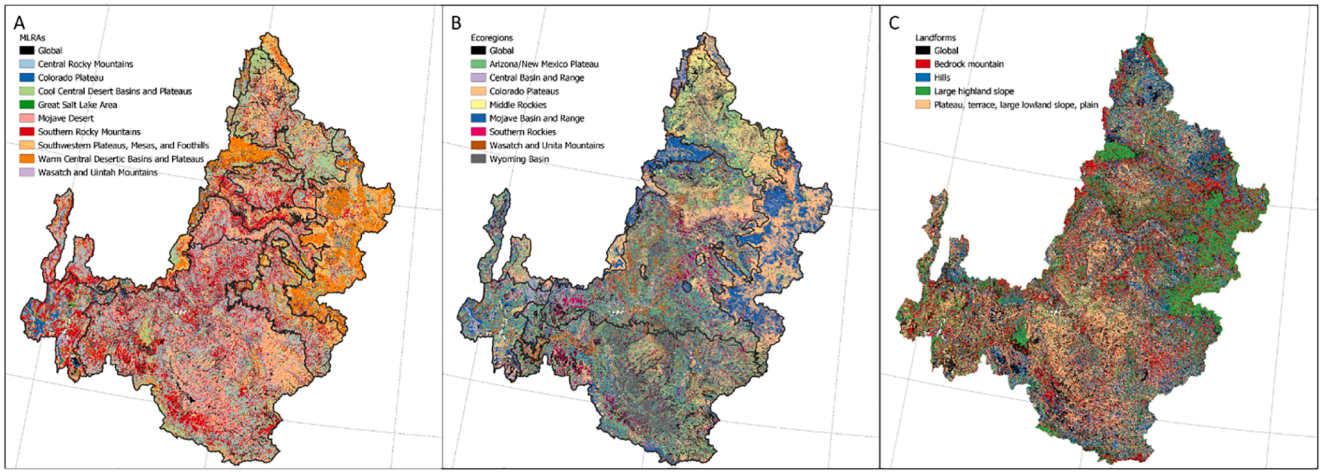

Fig. 12. Regional model identification from regional + global models. Each color corresponds to the regional (or global) model that made the prediction at each cell. Lines on MLRA (A) and Ecoregion (B) are the specific regions. Landform regions (C) are not shown for visual clarity. The specific colors match the colors of the region which made the prediction (Fig. 2). Note that rasters were resampled to 1 km for visual clarity, but this did not significantly change the pattern at this scale. The key point from this analysis is that individual regional models often contributed to predictions outside of the region in which the model was trained.

countries with large areas (e.g., Brazil, Mexico) - and presumably greater pedodiversity - had lower model performance despite the larger data density. These findings may contextualize the results of Somarathna et al. (2017) who found that increasing the number of observations resulted in a more accurate model, regardless of the specific model tested. Their study was conducted within a single physiographic region and it is likely that increasing the number of observations in a single region is likely to result in a better model, but the number of observations in a region is not the sole influence on model accuracy. However, we temper this interpretation with the acknowledgement that the observed differences in accuracy and uncertainty metrics between regions could also be due to bias in the geographical and feature-space distribution of observations inherent in the datasets that were used.

Additionally, we found that the imbalance ratio was not predictive of model accuracy or uncertainty metrics (Fig. 7 B). This is also surprising as class imbalance has been found to affect model accuracy (Brungard et al., 2015; Taghizadeh-mehrjardi et al., 2020a; Taghizadeh-Mehrjardi et al., 2020b). However, previous investigations into class imbalance was found when modeling taxonomic classes rather than soil property classes (i.e., soil depth classes) and it is possible that taxonomic classes have weaker covariate correlations thus possibly contributing to the relative importance of data distribution on soil taxonomic class modeling.

4.3. Physiographic vs. geographic areas

Is there a fundamental modeling domain over which DSM should be applied? The mean accuracy, Brier scores, and entropy were approximately equal between physiographic and geographic regions regardless of the number of areas in this study (Fig. 8). This suggests that our choice of physiographic region to constrain soil-covariate relationships may still be too large in extent because just dividing the region into geographic areas (which should dilute any soil-covariate relationships) results in approximately the same model metrics. It is possible that had physiographic regions been smaller in geographic space or had they been defined specifically to constrain soil-landscape relationships that larger differences between physiographic and geographic regions may have been found.

We conjecture that soil systems (also termed soil landscapes, or soilscapes) are likely candidates for defining fundamental modeling domains (Hewitt et al., 2010; Hole, 1978; Lagacherie et al., 2001; Schmidt et al., 2010). Soil systems are areas with repeating sequences of soils and conceptually could provide a needed connection between an explicit pedological understanding of the landscape and DSM techniques. We do caution here that establishing modeling domains solely by broadly defined landforms may be insufficient, but one potential way forward would be to model within a nested hierarchy (Salley et al., 2016b) such as modeling by landforms within an ecophysiographic region.

4.4. Ensembles of regional models

Any attempt at modeling multiple spatial regions at a local, regional, or global extents (e.g., FAO and ITPS, 2017) must eventually consider the necessity of seamlessly joining predictions. Although regional ensemble model accuracies were approximately the same as the global model (Fig. 9), we feel that continued investigation into regional modeling is warranted because ensembles of regional models consistently resulted in lower uncertainty than the global model (Fig. 10 & Fig. 11). Individual regional models (Fig. 5 and Fig. 6) also demonstrate that regional models made predictions with lower uncertainty than did global models, even if the accuracy between the regional and global models was approximately equal. When entropy of the regional models (Fig. 6) is compared with the corresponding Brier scores (Fig. 5) it is clear that, in general, regional models are equal to or more skilled than global models at making the right prediction (if only slightly) and make less uncertain predictions than do global models in each region. This same pattern is also evident in the regional ensemble models, which had higher prediction confidence than did the global model (Fig. 11).

We attribute the reduction in ensemble prediction accuracy compared to the global model accuracy to our specific methodological approach (Fig. 3). Several methods exist for merging predictions of different types of models from the same area (Malone et al., 2014; Román Dobarco et al., 2017; Caubet et al., 2019), but no method exists for joining models created in different areas. While it is likely that model averaging could potentially be used for continuous soil properties (e.g., pH, carbon), a different method is required for soil classes. We attempted several approaches including mosaicking regional predictions and joining the individual models into one 'supramodel', but found that these produced boundary artifacts or were very inaccurate (results not shown). Instead, we developed a regional modeling strategy that relied on the predicted probability raster stacks from each model. While conceptually simple, this approach had several limitations including high computational demands and the lack of a full probability distribution for each class in each cell. Instead, recent developments in Bayesian soil class data fusion could potentially be used which would

preserve the probabilities for each class (Rasaei and Bogaert, 2019).

It was surprising that models developed for one region often predicted the most probable class in areas far outside of the region in which each model was trained (Fig. 12). For example, predictions from the Southern Rocky Mountains MLRA produced predictions for large areas outside of the region but relatively small areas inside that specific region (Fig. 12A). We did not detect a consistent pattern in these results and cannot definitively infer why models trained in one region would make the highest predictions in other areas, but attribute this to the method of joining regional models and conclude that different methods of regional model ensembles are likely needed.

Overall, the utility of regional modeling appears to depend upon the method used to fuse predictions. We clearly acknowledge that better methods for fusing regional model predictions are needed, but we believe there is value in regional modeling; particularly because it appears to reduce uncertainty which has implications for the ultimate use of the soil spatial information in environmental modeling and monitoring (Carré et al., 2007; Nauman et al., 2019).

5. Conclusions

We explored the benefits and limitations of modeling effective soil depth classes by physiographic regions. We compared the accuracy, Brier scores, and entropy of regional models built with training data from specific physiographic regions to a single 'global' model that was built using all available observations. We also compared regional models to models built by geographic (rather than physiographic) areas and ensembles of regional and global models. Overall, soil depth class predictions made from ensembles of MLRA models has higher accuracy than the ecoregion ensemble model prediction but lower uncertainty than both the global model and the landform ensemble model predictions.

Specific findings include:

1. Useful inter-regional differences in global model accuracy were revealed when the global model was validated by region. We recommend validating (or at least cross-validating) models by region. Understanding how model accuracy and uncertainty metrics change by physiographic region has practical implications for digital soil mapping such as identifying where additional resources and effort could be preferentially allocated.
2. Regional model accuracy and uncertainty metrics did not seem to be affected by data availability or structure (density or imbalance ratio), suggesting that differences in regional model performance were driven by differences in underlying soil-covariate relationships.
3. No meaningful difference in accuracy and uncertainty metrics was found between physiographic versus geographic divisions. This may have been because of the size and definition of the regions as well as inherent pedodiversity. We tentatively suggest soil systems (also termed soilscapes) be used to define fundamental modeling domains.
4. Ensembles of regionally-specific models were approximately as accurate as global models, which we attribute to the particular methodology used to construct the regional ensembles. Other methods of data fusion could have increased model accuracy relative to the global model.
5. Both region-specific models and ensembles of region-specific models resulted in less uncertainty than the global model.

In conclusion, we find that modeling by region has promise for DSM and recommend further investigation into regional modeling because it appears to decrease prediction uncertainty as well as to provide insight into variation in accuracy and uncertainty. We suggest that perhaps too little attention has been placed on study area selection. This seemingly innocuous decision usually made at the beginning of a DSM project is important because it determines the modeling domain. We suggest that an appropriate modeling domain might be a soil system. At the very least, we suggest that DSM be performed by ecophysiographic region rather than by an arbitrarily or politically defined project area.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Contact the corresponding author, Dr. Colby Brungard who is responsible for coordinating the release of any data associated with this product, if you would like access to the resulting spatial predictions. This work was supported by the Utah Division of Wildlife Resources grant #160832, the Bureau of Land Management, US Geological Survey Ecosystems Mission Area, and the USDA National Institute of Food and Agriculture. USDA: Mention of a trade name, proprietary product, or vendor is for information only and does not guarantee or warrant the product by the US Government and does not imply its approval to the exclusion of other products or vendors that may also be suitable. The USDA is an equal opportunity provider and employer. USGS: Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. : Confusion Matrices

Table A1. Global model confusion matrix.

	Reference						
Prediction	BR	VS	S	MD	D	VD	User's Accuracy
BR	68	12	9	4	2	6	0.67
VS	7	46	22	14	5	10	0.44
S	8	37	86	26	6	27	0.45
MD	3	7	30	79	22	36	0.45
D	1	1	3	14	76	19	0.67
VD	12	19	80	119	79	729	0.70
Producer's Accuracy	0.69	0.38	0.37	0.31	0.40	0.88	

Table A2. Global model per-class sensitivity, specificity, and balanced accuracy. Sensitivity is the true positive rate (observations of that class that were correctly classified as belonging to that class). Sensitivity is the true negative rate (observations not in that class that were correctly classified as not belonging to that class). Balanced accuracy is the average of sensitivity and specificity.

	BR	VS	S	MD	D	VD
Sensitivity	0.69	0.38	0.37	0.31	0.40	0.88
Specificity	0.98	0.96	0.93	0.93	0.98	0.66
Balanced Accuracy	0.83	0.67	0.65	0.62	0.69	0.77

Table A3. MLRA regional + global prediction ensemble confusion matrix.

	Reference						
Prediction	BR	VS	S	MD	D	VD	User's Accuracy
BR	72	16	12	6	3	19	0.56
VS	1	39	26	13	4	5	0.44
S	12	35	88	32	5	36	0.42
MD	0	5	28	68	17	17	0.50
D	2	1	5	11	80	23	0.66
VD	12	25	69	123	81	721	0.70
Producer's Accuracy	0.73	0.32	0.39	0.27	0.42	0.88	

Table A4. MLRA regional + global prediction ensemble per-class sensitivity, specificity, and balanced accuracy. Sensitivity is the true positive rate (observations of that class that were correctly classified as belonging to that class). Specificity is the true negative rate (observations not in that class that were correctly classified as not belonging to that class). Balanced accuracy is the average of sensitivity and specificity.

	BR	VS	S	MD	D	VD
Sensitivity	0.73	0.32	0.39	0.27	0.42	0.88
Specificity	0.97	0.97	0.92	0.95	0.97	0.65
Balanced Accuracy	0.85	0.65	0.65	0.61	0.70	0.77

Table A5. Ecoregion regional + global prediction ensemble confusion matrix

	Reference						
Prediction	BR	VS	S	MD	D	VD	User's Accuracy
BR	62	6	8	4	2	9	0.68
VS	0	46	20	17	4	11	0.47
S	12	38	99	37	9	53	0.40
MD	0	5	28	62	19	9	0.50
D	8	2	11	13	84	58	0.48
VD	17	24	62	120	72	681	0.70
Producer's Accuracy	0.63	0.38	0.43	0.25	0.44	0.83	

Table A6. Ecoregion regional + global prediction ensemble per-class sensitivity, specificity, and balanced accuracy. Sensitivity is the true positive rate (observations of that class that were correctly classified as belonging to that class). Specificity is the true negative rate (observations not in that class that were correctly classified as not belonging to that class). Balanced accuracy is the average of sensitivity and specificity.

	BR	VS	S	MD	D	VD
Sensitivity	0.63	0.38	0.43	0.25	0.44	0.83
Specificity	0.98	0.97	0.90	0.96	0.94	0.67
Balanced Accuracy	0.80	0.67	0.67	0.60	0.69	0.75

Table A7. Landform regional + global prediction ensemble confusion matrix

Prediction	Reference BR	VS	S	MD	D	VD	User's Accuracy
BR	65	11	8	4	3	7	0.66
VS	2	46	16	12	4	8	0.52
S	10	33	93	29	6	23	0.48
MD	1	4	23	80	19	28	0.52
D	1	1	5	12	77	24	0.64
VD	20	26	83	116	81	731	0.69
Producer's Accuracy	0.66	0.38	0.41	0.32	0.41	0.89	

Table A8. Landform regional + global prediction ensemble per-class sensitivity, specificity, and balanced accuracy. Sensitivity is the true positive rate (observations of that class that were correctly classified as belonging to that class). Specificity is the true negative rate (observations not in that class that were correctly classified as not belonging to that class). Balanced accuracy is the average of sensitivity and specificity.

	BR	VS	S	MD	D	VD
Sensitivity	0.66	0.38	0.41	0.32	0.41	0.89
Specificity	0.98	0.97	0.93	0.95	0.97	0.63
Balanced Accuracy	0.82	0.68	0.67	0.63	0.69	0.76

References

Adhikari, K., Hartemink, A.E., Minasny, B., Kheir, R.B., Greve, M.B., Greve, M.H., 2014. Digital mapping of soil organic carbon contents and stocks in Denmark. PLOSone 9, e105519. https://doi.org/10.1371/journal.pone.0105519.

Adhikari, K., Kheir, R.B., Greve, M.B., Bocher, P.K., Malone, B.P., Minasny, B., MecBratney, A.B., Greve, M.H., 2013. High-Resolution 3-D Mapping of Soil Texture in Denmark. Soil Sci. Soc. Am. J. 860–876 https://doi.org/10.2136/sssaj2012.0275.

Ault, T.R., Mankin, J.S., Cook, B.I., Smerdon, J.E., 2016. Relative impacts of mitigation, temperature, and precipitation on 21st-century megadrought risk in the American Southwest. Sci. Adv. 2, e1600873.

Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261, 110–123. https://doi.org/10.1016/J.GEODERMA.2015.07.006.

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

Belcher, J.W., Keddy, P.A., Twolan-Strutt, L., 1995. Root and Shoot Competition Intensity Along a Soil Depth Gradient. J. Ecol. 83, 673–682.

Bernard-Verdier, M., Navas, M.L., Vellend, M., Violle, C., Fayolle, A., Garnier, E., 2012. Community assembly along a soil depth gradient: Contrasting patterns of plant trait convergence and divergence in a Mediterranean rangeland. J. Ecol. 100, 1422–1433. https://doi.org/10.1111/1365-2745.12003.

Bivand, R., Keitt, T., Rowlingson, B., 2019. rgdal: Bindings for the "Geospatial" Data Abstraction Library. https://cran.r-project.org/web/packages/rgdal/index.html. Accessed 12.21.2020.

Bivand, R., Rundel, C., 2019. rgeos: Interface to Geometry Engine -. Open Source ('GEOS').

Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. Applied spatial data analysis with {R}, second ed. Springer, NY.

Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Nield, S.J., Saunders, A.M., Stum, A.K., 2008. Landsat Spectral Data for Digital Soil Mapping. Media 193–202. https://doi.org/10.1007/978-1-4020-8592-5_16.

Brenning, A., 2012. Spatial Cross-validation and bootstrap for the assessement of prediction rules in remote sensing: The R package SPERROREST. 2012 IEEE Int. Geosci. Remote Sens. Symp. 5372–5375. https://doi.org/10.1109/IGARSS.2012.6352393.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards Jr., T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239–240, 68–83. https://doi.org/10.1016/j.geoderma.2014.09.019.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62, 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond DSM. Geoderma 142, 69–79. https://doi.org/10.1016/j.geoderma.2007.08.015.

Caubet, M., Román Dobarco, M., Arrouays, D., Minasny, B., Saby, N.P.A., 2019. Merging country, continental and global predictions of soil texture: Lessons from ensemble modelling in France. Geoderma 337, 99–110. https://doi.org/10.1016/j.geoderma.2018.09.007.

Chander, G., Markham, B.L., Helder, D.L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sens. Environ. 113, 893–903. https://doi.org/10.1016/j.rse.2009.01.007.

Chen, S., Leatitia, V., Martin, M.P., Walter, C., Lacoste, M., Richer-de-forges, A.C., Saby, N.P.A., Loiseau, T., Hu, B., Arrouays, D., 2019. Probability mapping of soil thickness by random survival forest at a national scale. Geoderma 344, 184–194. https://doi.org/10.1016/j.geoderma.2019.03.016.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Copeland, S.M., Bradford, J., Duniway, M.C., Schuster, R., 2017. Potential impacts of overlapping land-use and climate in a sensitive dryland: a case study of the Colorado Plateau, USA. Ecosphere 8, e01823. https://doi.org/10.1002/ecs2.1823.

Deane-Mayer, Z.A., Knowles, J.E., 2019. caretEnsemble: Ensembles of Caret Models.

Duniway, M.C., Bestelmeyer, B.T., Tugel, A.J., 2010. Soil Processes and Properties That Distinguish Ecological Sites and States. Rangelands 32, 159–160.

FAO, ITPS, 2017. Global Soil Organic Carbon Map (GSOCmap) Technical Report. Rome.

Fick, S.E., Belnap, J., Duniway, M.C., Wash, B., 2020. Grazing-Induced Changes to Biological Soil Crust Cover Mediate Hillslope Erosion in Long-Term Exclosure Experiment. Rangel. Ecol. Manag. 73, 61–72. https://doi.org/10.1016/j.rama.2019.08.007.

Fuhlendorf, S.D., Smeins, F.E., 1998. The influence of soil depth on plant species response to grazing within a semi-arid savanna. Plant Ecol. 138, 89–96. https://doi.org/10.1023/A:1009704723526.

Goldstein, H.L., Reynolds, R.L., Reheis, M.C., Yount, J.C., Neff, J.C., 2008. Compositional trends in aeolian dust along a transect across the southwestern United States. J. Geophys. Res. Earth Surf. 113, 1–15. https://doi.org/10.1029/2007JF000751.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine : Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Guevara, M., Olmedo, G.F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G.E., Arroyo-Cruz, C.E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C.O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J.A., Ibelles Navarro, A.R., Loayza, V., Manueles, A.M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Ramos, I.A., Rey Brina, J.C., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Rosales Ibarra, A., Rosales Riveiro, K.A., Schulz, G.A., Spence, A., Vasques, G.M., Vargas, R.R., Vargas, R., 2018. No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America. SOIL 4, 173–193. https://doi.org/10.5194/soil-4-173-2018.

Guo, Y., Gong, P., Amundson, R., 2003. Pedodiversity in the United States of America. Geoderma 117, 99–115. https://doi.org/10.1016/S0016-7061(03)00137-X.

Hastie, T., Tibshirani, R., Friedman, J.H., 2001. The elements of statistical learning: data mining, inference, and prediction : with 200 full-color illustrations. Springer, New York.

Hengl, T., de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12, 1–40. https://doi.org/10.1371/journal.pone.0169748.

Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Gonzalez, M.R., 2014. SoilGrids1km — Global Soil Information Based on Automated Mapping. PLoS ONE 9, 1–17. https://doi.org/10.1371/journal.pone.0105992.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Tamene, L., Tondoh, J.E., 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. PLoS ONE 10, e0125814. https://doi.org/10.1371/journal.pone.0125814.

Hewitt, A.E., Barringer, J.R.F., Forrester, G.J., McNeill, S.J., 2010. Soilscapes Basis for Digital Soil Mapping in New Zealand. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping Bridging Research, Environmental Application, and Operation. Springer, Netherlands, pp. 297–307. https://doi.org/10.1016/S0016-7061(03)00223-4.

Hijmans, R.J., 2019. raster: Geographic Data Analysis and Modeling. https://cran.r-project.org/web/packages/raster/index.html. Accessed 12.21.2020.

Hill, P.L., Kucks, R.P., Ravat, D., 2009. Aeromagnetic and Aeroradiometric Data for the Conterminous United States and Alaska from the National Uranium Resource Evaluation (NURE) Program of the U.S. Department of Energy, Open-File Report. 10.3133/ofr20091129.

Hole, F.D., 1978. An approach to landscape analysis with emphasis on soils. Geoderma 21, 1–23. https://doi.org/10.1016/0016-7061(78)90002-2.

Iwahashi, J., Kamiya, I., Matsuoka, M., Yamazaki, D., 2018. Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification, Progress in Earth and Planetary Science. Progress in Earth and Planetary Science. 10.1186/s40645-017-0157-2.

Kassambara, A., 2019. ggpubr: "ggplot2" Based Publication Ready Plots. https://www.tidyverse.org/. Accessed 12.21.2020.

Kempen, B., Brus, D.J., Heuvelink, G.B.M., Stoorvogel, J.J., 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151, 311–326. https://doi.org/10.1016/j.geoderma.2009.04.023.

Kuhn, M., 2019. caret: Classification and Regression Training. https://topepo.github.io/caret/. Accessed 12.21.2020.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York, NY.

Lagacherie, P., Robbez-Masson, J.M., Nguyen-The, N., Barthès, J.P., 2001. Mapping of reference area representativity using a mathematical soilscape distance. Geoderma 101, 105–118. https://doi.org/10.1016/S0016-7061(00)00101-4.

Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B., 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. Geoderma 232–234, 34–44. https://doi.org/10.1016/j.geoderma.2014.04.033.

McBratney, A., Mendonça Santos, M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

McBratney, A.B., Hart, G.A., McGarry, D., 1991. The use of region partitioning to improve the representation of geo statistically mapped soil attributes. J. Soil Sci. 42, 513–532. https://doi.org/10.1111/j.1365-2389.1991.tb00427.x.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environ. Model. Softw. 101, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001.

Miller, M.E., Belote, R.T., Bowker, M.A., Garman, S.L., 2011. Alternative states of a semiarid grassland ecosystem: implications for ecosystem services. Ecosphere 2, art55. https://doi.org/10.1890/ES11-00027.1.

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma 263, 16–34. https://doi.org/10.1016/J.GEODERMA.2015.08.035.

Munson, Seth M., Belnap, J., Okin, G.S., 2011a. Responses of wind erosion to climate-induced vegetation changes on the Colorado Plateau. Proc. Natl. Acad. Sci. U. S. A. 108, 3854–3859. https://doi.org/10.1073/pnas.1014947108.

Munson, S.M., Belnap, J., Schelz, C.D., Moran, M., Carolin, T.W., 2011b. On the brink of change: plant responses to climate on the Colorado Plateau. Ecosphere 2, 1–15. https://doi.org/10.1890/ES11-00059.1.

National Cooperative Soil Survey, 2019. National Cooperative Soil Characterization Database. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/research/?cid=nrcs142p2_053543. Accessed 12.21.2020.

Nauman, T.W., Duniway, M.C., 2020. A hybrid approach for predictive soil property mapping using conventional soil survey data. Soil Sci. Soc. Am. J. 84, 1170–1194.

Nauman, T.W., Duniway, M.C., 2016. The Automated Reference Toolset: A Soil-Geomorphic Ecological Potential Matching Algorithm. Soil Sci. Soc. Am. J. 80, 1317. https://doi.org/10.2136/sssaj2016.05.0151.

Nauman, T.W., Duniway, M.C., Webb, N., Belnap, J., 2018. Elevated aeolian sediment transport on the Colorado Plateau, USA: The role of grazing, vehicle disturbance, and increasing aridity. Earth Surf. Process. Landforms 43, 2897–2914.
Nauman, T.W., Ely, C.P., Duniway, M.C., 2019. Salinity Yield Modeling of the Upper Colorado River Basin Using 30 - m Resolution Soil Maps and Random Forests Water Resources Research. Water Resour. Res. 55, 1–20. https://doi.org/10.1029/2018WR024054.
Neff, J.C., Reynolds, R.L., Belnap, J., Lamothe, P., 2008. Multi-decadal impacts of grazing on soil physical and biogeochemical properties in southeast Utah. Ecol. Appl. 15, 87–95.
Neuwirth, E., 2014. RColorBrewer: ColorBrewer Palettes. https://cran.r-project.org/web/packages/RColorBrewer/index.html. Accessed 12.21.2020.
Nussear, K.E., Esque, T.C., Inman, R.D., Gass, L., Thomas, K. a, Wallace, C.S. a, Blainey, J. B., Miller, D.M., Webb, R.H., 2009. Modeling habitat of the desert tortoise (Gopherus agassizii) in the Mojave and parts of the Sonoran Deserts of California, Nevada, Utah, and Arizona, US Geological Survey open-file report.
Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the Conterminous United States : Evolution of a Hierarchical Spatial Framework. Environ. Manage. 54, 1249–1266. https://doi.org/10.1007/s00267-014-0364-1.
Padarian, J., Minasny, B., Mcbratney, A.B., 2017. Chile and the Chilean soil grid : A contribution to GlobalSoilMap. Geoderma Reg. 9, 17–28. https://doi.org/10.1016/j.geodrs.2016.12.001.
Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. R J. 10, 439–446. 10.32614/RJ-2018-009. https://cran.r-project.org/web/packages/sf/index.html. Acessed 12.21.2020.
Pelletier, J.D., Broxton, P.D., Hazenberg, P., Zeng, X., Troch, P.A., Niu, G.-Y., Williams, Z., Brunke, M.A., Gochis, D., 2016. A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. J. Adv. Model. Earth Syst. 8, 41–65. https://doi.org/10.1002/2015MS000526.
Peng, Y., Xiong, X., Adhikari, K., Knadel, M., Grunwald, S., Greve, M.H., 2015. Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra. PLoS ONE 10, e0142295. https://doi.org/10.1371/journal.pone.0142295.
PRISM Climate Group, 2010. 30-yr climate normals. https://prism.oregonstate.edu/. Accessed 12.21.2020.
Probst, P., 2018. measures: Performance Measures for Statistical Learning. https://cran.r-project.org/web/packages/measures/index.html. Accessed 12.21.2020.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. https://www.r-project.org/. Accessed 12.21.2020.
Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. Soil Sci. Soc. Am. J. 82 https://doi.org/10.2136/sssaj2017.04.0122.
Rasaei, Z., Bogaert, P., 2019. Bayesian data fusion for combining maps of predicted soil classes : A case study using legacy soil profiles and DEM covariates in Iran. Catena 182, 104138. https://doi.org/10.1016/j.catena.2019.104138.
Román Dobarco, M., Arrouays, D., Lagacherie, P., Ciampalini, R., Saby, N.P.A., 2017. Prediction of topsoil texture for Region Centre (France) applying model ensemble methods. Geoderma 298, 67–77. https://doi.org/10.1016/J.GEODERMA.2017.03.015.
Ross, C.W., Grunwald, S., Vogel, J.G., Markewitz, D., Jokela, E.J., Martin, T.A., Bracho, R., Bacon, A.R., Brungard, C.W., Xiong, X., 2020. Accounting for two-billion tons of stabilized soil carbon. Sci. Total Environ. 703, e134615 https://doi.org/10.1016/j.scitotenv.2019.134615.
RStudio Team, 2018. RStudio: Integrated Development Environment for R. https://rstudio.com/. Accessed 12.21.2020.
Salley, S., Monger, H.C., Brown, J., 2016. Completing the Land Resource Hierarchy 38, 313–317. 10.1016/j.rala.2016.10.003.
Salley, S.W., Talbot, C.J., Brown, J.R., 2016b. The Natural Resources Conservation Service Land Resource Hierarchy and Ecological Sites. Soil Sci. Soc. Am. J. 80, 1–9. https://doi.org/10.2136/sssaj2015.05.0305.
Schauberger, P., Walker, A., 2019. openxlsx: Read, Write and Edit xlsx Files. https://cran.r-project.org/web/packages/openxlsx/openxlsx.pdf. Accessed 12.21.2020.
Schmidt, K., Behrens, T., Friedrich, K., Scholten, T., 2010. A method to generate soilscapes from soil maps. J. Plant Nutr. Soil Sci. 173, 163–172. https://doi.org/10.1002/jpln.200800208.
Schoeneberger, P.J., Wysocki, D.A., Benham, E.C., Soil Survey Staff, 2012. Field Book for Describing and Sampling Soils, Version 3.0. Natural Resources Conservation Service. National Soil Survey Center, Lincoln, NE.
Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A., 2019. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecol. Modell. 406, 109–120. https://doi.org/10.1016/j.ecolmodel.2019.06.002.
Skovlin, J., Roecker, S., 2019. soilDB: Soil Database Interface. https://cran.r-project.org/web/packages/soilDB/index.html. Access 12.21.2020.
Soil Survey Staff, 2019. Soil Survey Geographic (SSURGO) Database. Available online at https://websoilsurvey.nrcs.usda.gov/. Accessed 12.21.2020.
Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More Data or a Better Model? Figuring Out What Matters Most for the Spatial Prediction of Soil Carbon. Soil Sci. Soc. Am. J. 81, 1413–1426. 10.2136/sssaj2016.11.0376.
South, A., 2017a. rnaturalearth: World Map Data from Natural Earth. https://cran.r-project.org/web/packages/rnaturalearth/README.html. Accessed 12.21.2020.
South, A., 2017b. rnaturalearthdata: World Vector Map Data from Natural Earth Used in "rnaturalearth." https://www.naturalearthdata.com/. Accessed 12.21.2020.
Taghizadeh-Mehrjardi, R., Minasny, B., Toomanian, N., Zeraatpisheh, M., Amirian-Chakan, A., Triantafilis, J., 2019. Digital Mapping of Soil Classes Using Ensemble of Models in Isfahan Region. Iran. Soil Syst. 3, 37. https://doi.org/10.3390/soilsystems3020037.
Taghizadeh-mehrjardi, R., Schmidt, K., Amirian-chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., Scholten, T., 2020a. Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space. Remote Sens. 12, 1095. https://doi.org/10.3390/rs12071095.
Taghizadeh-Mehrjardi, R., Schmidt, K., Eftekhari, K., Behrens, T., Jamshidi, M., Davatgaar, N., Toomanian, N., Scholten, T., 2020b. Synthetic resampling strategies and machine learning for digital soil mapping in Iran. Eur. J. Soil Sci. 71, 352–368. https://doi.org/10.1111/ejss.12893.
Thompson, J.A., Kienast-brown, S., Avello, T.D., Philippe, J., Brungard, C., 2020. Soils 2026 and digital soil mapping – A foundation for the future of soils information in the United States. Geoderma Reg. 22, e00294 https://doi.org/10.1016/j.geodrs.2020.e00294.
Thompson, J.A., Kolka, R.K., 2005. Soil Carbon Storage Estimation in a Forested Watershed using Quantitative Soil-Landscape Modeling. Soil Sci. Soc. Am. J. 69, 1086–1093. https://doi.org/10.2136/sssaj2004.0322.
Udall, B., Overpeck, J., 2017. The twenty-first century Colorado River hot drought andimplications for the future. Water Resour. Res. 53, 2404–2418. https://doi.org/10.1002/2016WR019638.
USDA-NRCS, 2006. Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. Handbook 296. U.S. Department of Agriculture, Washington.
USDA-NRCS, USGS, EPA, 2019. Watershed Boundary Dataset. https://www.usgs.gov/core-science-systems/ngp/national-hydrography/watershed-boundary-dataset?qt-science_support_page_related_con=4#qt-science_support_page_related_con. Accessed 12.21.2020.
Viscarra-Rossel, R., Webster, R., Bui, E.N., Baldock, J.A., 2014. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. Glob. Chang. Biol. 2953–2970 https://doi.org/10.1111/gcb.12569.
Walvoort, D.J.J., Brus, D.J., de Gruijter, J.J., 2010. An R package for spatial coverage sampling and random sampling from compact geographical strata by k-means. Comput. Geosci. 36, 1261–1267.
Wickham, H., 2019a. forcats: Tools for Working with Categorical Variables (Factors). https://www.tidyverse.org/. Accessed 12.21.2020.
Wickham, H., 2019b. stringr: Simple, Consistent Wrappers for Common String Operations. https://www.tidyverse.org/. Accessed 12.21.2020.
Wickham, H., 2017. tidyverse: Easily Install and Load the "Tidyverse." https://www.tidyverse.org/. Accessed 12.21.2020.
Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.
Wickham, H., 2007. Reshaping Data with the reshape Package. J. Stat. Softw. 21, 1–20.