Digital Mapping of Ecological Land Units using a Nationally Scalable Modeling Framework

Jonathan J. Maynard*
USDA-ARS
Jornada Experimental Range
Las Cruces, NM 88003

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

Shawn W. Salley
Brandon T. Bestelmeyer
USDA-ARS
Jornada Experimental Range
Las Cruces, NM 88003

Michael C. Duniway
US Geological Survey
Southwest Biological Science Center
Moab, UT 84532

Curtis J. Talbot
Joel R. Brown
USDA-NRCS National Ecological Site
Team Jornada Experimental Range
Las Cruces, NM 88003

Ecological site descriptions (ESDs) and associated state-and-transition models (STMs) provide a nationally consistent classification and information system for defining ecological land units for management applications in the United States. Current spatial representations of ESDs, however, occur via soil mapping and are therefore confined to the spatial resolution used to map soils within a survey area. Land management decisions occur across a range of spatial scales and therefore require ecological information that spans similar scales. Digital mapping provides an approach for optimizing the spatial scale of modeling products to best serve decision makers and have the greatest impact in addressing land management concerns. Here, we present a spatial modeling framework for mapping ecological sites using machine learning algorithms, soil survey field observations, soil survey geographic databases, ecological site data, and a suite of remote sensing-based spatial covariates (e.g., hyper-temporal remote sensing, terrain attributes, climate data, land-cover, lithology). Based on the theoretical association between ecological sites and landscape biophysical properties, we hypothesized that the spatial distribution of ecological sites could be predicted using readily available geospatial data. This modeling approach was tested at two study areas within the western United States, representing 6.1 million ha on the Colorado Plateau and 7.5 million ha within the Chihuahuan Desert. Results show our approach was effective in mapping grouped ecological site classes (ESGs), with 10-fold cross-validation accuracies of 70% in the Colorado Plateau based on 1405 point observations across eight expertly-defined ESG classes and 79% in the Chihuahuan Desert based on 2589 point observations across nine expertly-defined ESG classes. Model accuracies were also evaluated using external-validation datasets; resulting in 56 and 44% correct classification for the Colorado Plateau and Chihuahuan Desert, respectively. National coverage of the training and covariate data used in this study provides opportunities for a consistent national-scale mapping effort of ecological sites.

Abbreviations: ESD, ecological site description; ESG, ecological site group; MAPVI, model-agnostic version of the permutation-based variable importance; MLRA, major land resource area; PR-AUC, precision-recall curve; SMU, soil map unit; STM, state and transition model; SSURGO, Soil Survey Geographic Database.

Core Ideas

- **Digital modeling framework for mapping ecological land units was developed.**
- **Digital maps eliminate spatial ambiguity of multicomponent soil map units.**
- **Spatial scale of modeling products can be optimized to meet management needs.**
- **National coverage of the training and covariate data provides ability to scale.**

The implementation of management practices to maintain or enhance the condition of a landscape requires a knowledge of patterns in ecological potential and the responses of soils, vegetation, and wildlife to management actions. Traditionally, this knowledge has been generated through long-term observations of how the land responds to management practices or disturbance processes, and further augmented by expert knowledge and scientific research (Knapp and Fernandez-Gimenez, 2009; Caudle et al., 2013; Karl and Talbot, 2016). In recent years, however, the cumulative effects of soil degradation, persistent vegetation change, invasive species, and a changing climate are forcing land managers to seek new tools and approaches to more efficiently direct the allocation of management

efforts (Bestelmeyer et al., 2016). In the United States, decades of research and land manager experience has been synthesized in a system of ecological site descriptions (ESDs), each featuring a state and transition model (STM) that describes dynamics within an ecological site. Ecological sites offer a framework for understanding and managing landscapes based on the concepts of land potential and an evolving understanding of ecological resilience and threshold-type responses of many ecosystems to perturbations. The ecological site classification system partitions landscapes into ecological land units (ecological sites) that share a similar range of biophysical properties (e.g., soil, climate, and potential vegetation) which leads to similar responses to management activities and disturbance processes (Brown, 2010). STMs describe the responses of vegetation and soil surface processes to natural drivers and management actions. The formal adoption of the ESD/STM framework by US federal land management agencies (Caudle et al., 2013) has resulted in a national system for defining, mapping, and monitoring rangelands and forests.

Ecological site concepts are developed and mapped across a wide range of spatial scales to address different management questions and concerns. At the broadest spatial scales, physiographic climate and land use zones (e.g., ecoregions, Cleland et al., 1997; MLRAs, McMahon et al., 2004; TEUI, Winthers et al., 2005) define the spatial extent of a single ecological site mapping effort due to regional scale controls on soil-vegetation dynamics and feasible land use options (Salley et al., 2016b). At intermediate spatial scales, soil-geomorphic systems (SGSs) subdivide climate zones into discrete land areas with a characteristic spatial arrangement of ecological sites that experience similar ecological and soil-forming processes (Monger and Bestelmeyer, 2006). At the finest spatial scales, variations in soil physical and chemical properties along with local topographic context define individual ecological sites (Salley et al., 2016a). The spatial and/ or thematic scale at which ecological site concepts are developed largely determines their efficacy in addressing specific management questions. For example, ecological sites developed at fine scales (1:12,000 to 1:24,000) may be effective at targeting critical habitat for species of concern (Karl and Herrick, 2010), but may be less efficient at managing for invasive species (e.g., cheatgrass) that are regulated by broader scale ecological controls (Chambers et al., 2016). Thus, ecological site concepts can be usefully developed and mapped at a variety of scales and with flexible criteria for different management objectives (Karl and Herrick, 2010; Maynard and Karl, 2017).

Currently, the spatial representations of ecological sites occur through assigning ecological sites to Soil Survey Geographic Database (SSURGO) soil map unit (SMU) components (i.e., soil types) based on their observed co-occurrence across a landscape. Thus, the spatial accuracy of ecological site maps is determined by the accuracy of soil map delineations and the quality of the grouping of soil components making up an ecological site. Criteria for the spatial delineation of SSURGO soil map units, however, is often different than those used to separate landscapes according to ecological potential, and SSURGO map units may circumscribe more than one ecological site (Ireland and Drohan, 2015; Nauman and Duniway, 2016; Maynard and Karl, 2017). Consequently, the current ecological site mapping framework, consisting of the forced linkage between SMU components and ecological site concepts, has been criticized within the scientific community for not accurately representing important ecosystem dynamics and thus limiting the spatial accuracy of ecological site maps used for land management (Ireland and Drohan, 2015; Nauman and Duniway, 2016; Maynard and Karl, 2017).

Ongoing advancements in geospatial technologies are providing the tools needed to move beyond conventional soil survey maps. Recent work has demonstrated the potential of digital mapping techniques in predicting ecological site distribution (Maynard and Karl, 2017) and ecological site states (Nauman et al., 2015; Poitras et al., 2018). Through leveraging remote sensing products and machine learning models, it is now possible to create high spatial resolution maps of ecological sites. Based on the theoretical association between ecological sites and many landscape biophysical properties, we hypothesized that the spatial distribution of ecological sites could be predicted using readily available geospatial data that represent the biophysical concept distinguishing the sites. Furthermore, the type of variables and roles they play in predictive models can help quantitatively define the biophysical factors that distinguish ecological sites and allow for iterative refinement of site concepts.

In many areas of the western United States, ecological site concepts have been developed at a very fine thematic resolution, reflecting subtle differences in soil properties and associated vegetation characteristics. In some cases, this was done to address specific management objectives, while in others it reflects a mirroring of the level of detail used to develop soil series and soil series phase (i.e., soil component) concepts within a soil survey area. In two regions of the western United States (Colorado Plateau and Chihuahuan Desert), growing interest among stakeholders for ecological site information relevant to landscape-level patterns and transition processes was addressed by the development of new concepts for mapping ecological sites with a coarser thematic resolution (Bestelmeyer et al., 2016; Duniway et al., 2016). This resulted in approximately 150 ecological sites being aggregated into eight ecological site groups in the Colorado Plateau, and 42 ecological sites aggregated into nine ecological site groups in the Chihuahuan Desert. The development of ecological site groups (ESGs) represents a higher order classification of ecological potential that focuses on differences in plant functional groups that control ecosystem dynamics. This results in land units that can produce more interpretable maps that are well matched to landscape-level decision-making (Steele et al., 2012; Bestelmeyer et al., 2016; Duniway et al., 2016).

Here we present a consistent, nationally scalable framework for predicting ecological sites at multiple spatial scales using a national point database, remotely sensed geospatial data layers, and machine learning algorithms. The main objective of this study was to evaluate the accuracy of our modeling framework for predicting the spatial distribution of the ESGs developed in the
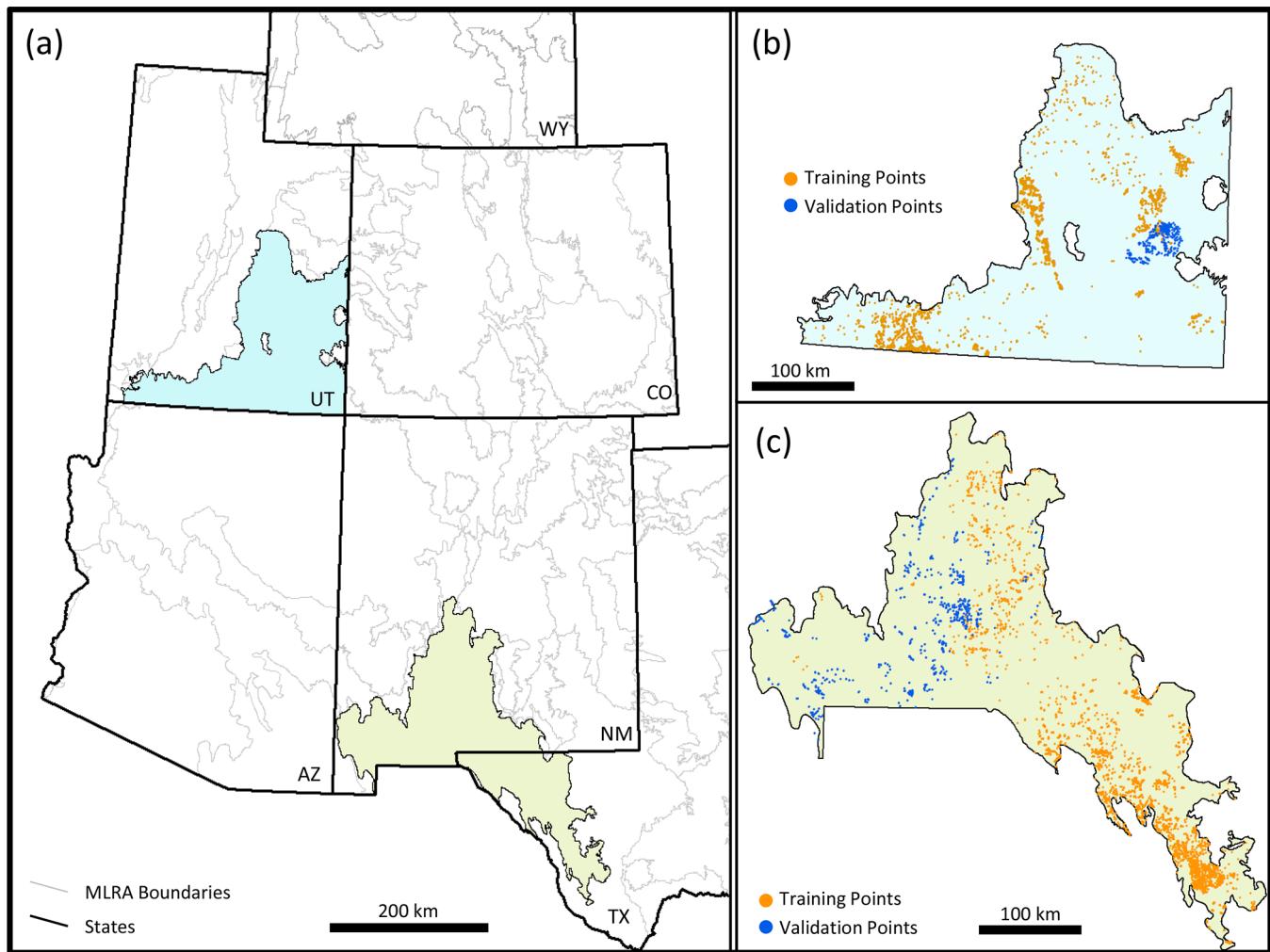
Fig. 1. Location of study areas. (a) Location of Colorado Plateau study area in south-eastern UT and Chihuahuan Desert study area in southern NM and western TX; location of training (orange) and validation (blue) points for (b) Colorado Plateau and (c) Chihuahuan Desert study areas.

Colorado Plateau and Chihuahuan Desert study areas (Fig. 1). Specific objectives were to: (i) evaluate which environmental covariates were the most important in predicting ESGs, (ii) test the consistency of ESG concepts by evaluating the individual error of ecological sites within each ESG, and (iii) evaluate the potential application of our modeling framework for the consistent mapping of ecological sites nationally.

MATERIALS AND METHODS
Study Areas

Two study areas were used to evaluate our ecological site modeling framework: the Utah portion of the Colorado Plateau (within MLRA 35) and the western half of the US extent of the Chihuahuan Desert (within MLRA 42). Study area selection was based on the following criteria: (i) high quality ecological site and SSURGO data, (ii) established ESG concepts, and (iii) ecological site field data for model validation. Study area boundaries were delineated using the EPA level 4 ecoregions that intersected portions of USDA-NRCS MLRA 35 for the Colorado Plateau and MLRA 42 for the Chihuahuan Desert. Both study areas represent arid to semiarid grass–shrubland ecosystems that differ with respect to climate, geology, soils, vegetation, and

topographic complexity. For more detailed descriptions of the Colorado Plateau and Chihuahuan study areas, see Duniway et al. (2016) and Bestelmeyer et al. (2016), respectively.

Modeling Framework

The spatial modeling framework for predicting ESGs consists of five steps: (1) develop an ESG point dataset for training and cross-validation, (2) preprocess ESG covariates for tiling and point overlay, (3) covariate feature selection and model development, (4) spatial predictions using tiled raster stacks, and (5) model validation and uncertainty analysis. All modeling steps are presented in Fig. 2 and were implemented using Open-Source software, including: SAGA GIS, GDAL, and the R environment for statistical computing (Conrad et al., 2015; R Development Core Team, 2015; GDAL/OGR Contributors, 2018).

ESG Point Dataset

The development of an ESG point dataset first required that ESG concepts be developed and correlated to SSURGO SMU components. The development of ESG concepts in our study areas involved the establishment of workgroups comprised of scientists and managers with knowledge of existing ESDs in each study

Soil Science Society of America Journal
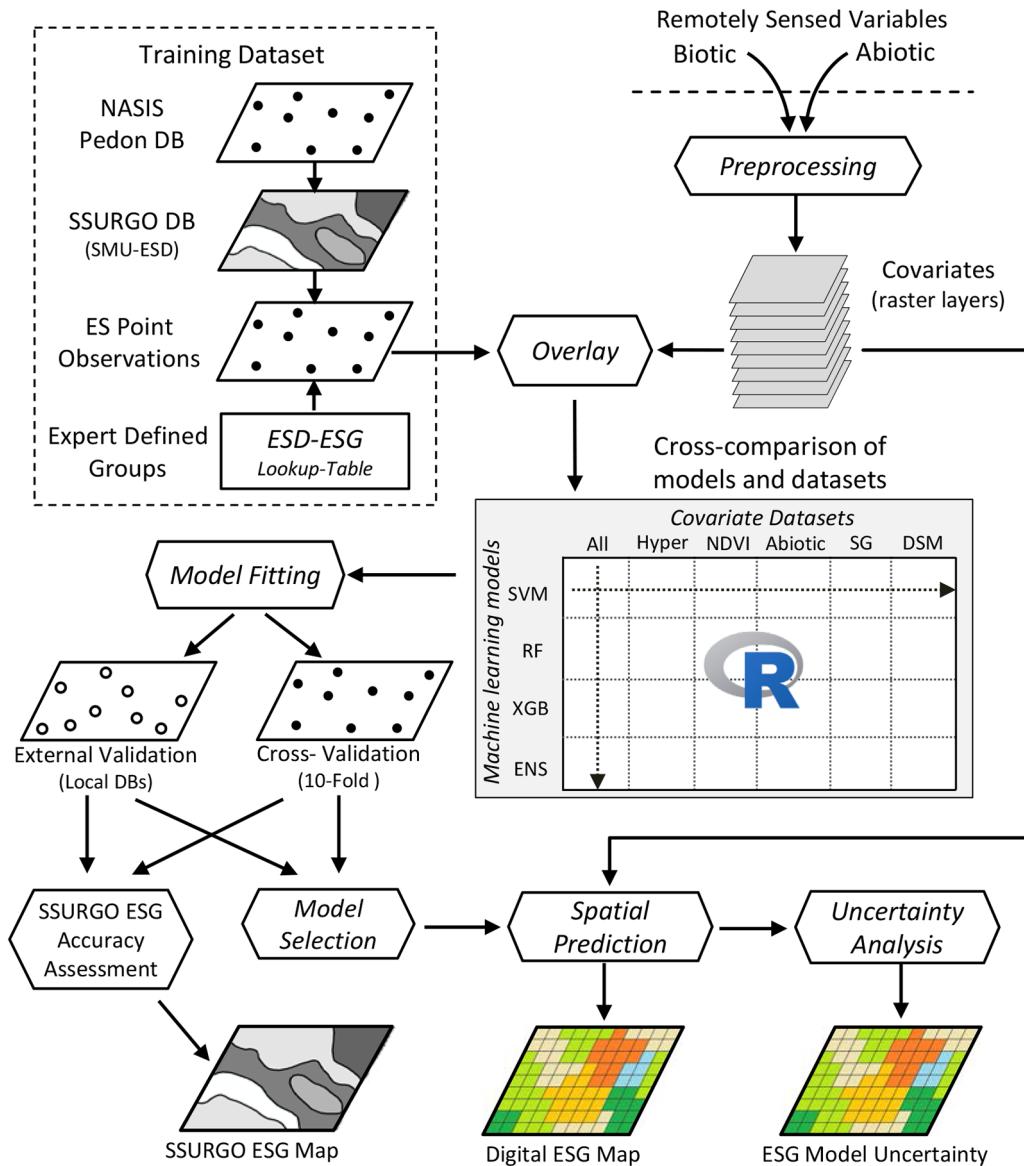
Fig. 2. Modeling framework for the spatial prediction of ecological land units.

area. Through leveraging the current inventory of ESDs correlated to SSURGO soil map units, relevant geospatial data (e.g., terrain and remote sensing indices), and expert knowledge; these groups were able to develop 8 ESGs from approximately 150 ESDs in the Colorado Plateau (Duniway et al., 2016) and 9 ESGs from 45 ESDs in the Chihuahuan Desert (Bestelmeyer et al., 2016) (Table 1). Next, we compiled a subset of soil observations from the National Soil Information System (NASIS) database maintained by the USDA—NRCS, extracting all point observations that had been assigned nearest soil series designation. This resulted in 4643 points in the Colorado Plateau (Fig. 1b) and 4295 points in the Chihuahuan Desert (Fig. 1c) (270,927 points were identified for the conterminous United States). At each NASIS point within our study areas a spatial query was performed, extracting the ecological site class for the first matched component of the same soil series identified from the intersected or adjacent neighboring SMUs in the SSURGO database. The ESD associated with the matched component was extracted and assigned to its corresponding

NASIS point. This spatial matching procedure resulted in 1405 points (30% match rate) for the Colorado Plateau and 2589 points (61% match rate) for the Chihuahuan Desert. With the exception of one Order 4 soil survey in the Colorado Plateau (San Juan County, Utah, Navajo Indian Reservation), soils in our study areas were mapped at the Order 3 scale (i.e., 1:20,000 to 1:63,360). This resulted in mean map unit areas of approximately 375 acres (1.3 to 5440 acres for the 1st and 99th percentiles) and 575 acres (1.7 to 9390 acres for the 1st and 99th percentiles), for the Colorado Plateau and Chihuahuan Desert, respectively. While the mapping scale may have influenced our match rates in some areas, the consistency of Order 3 surveys likely minimized significant scaling effects. The lower match rate on the Colorado Plateau was due in large part to portions of SSURGO being incomplete, but in the process of publication. We then used expertly defined look-up tables relating all matched ESDs to the ESG concepts developed from the two workgroups to create a final ESG point dataset. In the Chihuahuan Desert study area, the Saline and Playa/lakebeds ecological site

Table 1. Description of ecological site group (ESG) characteristics for the Colorado Plateau and Chihuahuan Desert study areas.†

ESG	Code	Soil-landform setting and vegetation communities
		Colorado Plateau
Bottoms and Flats	ESG1	Occurs in flat, low-lying areas with ephemeral washes and streams. Soil texture, depth, and chemistry vary widely. Dominated by shrubs and mixture of perennial cool-season/warm-season grasses.
Outcrops and Slopes	ESG2	Bedrock controlled landforms with vegetation relegated to pockets, very shallow soil, or fissures. Often steep. Pinyon-Juniper woodlands, with various shrubs interspersed. Mostly exposed bedrock.
Saline Hills and Badlands	ESG3	Highly salt limited, erosion features common, often sloping. Ephedra and Mat Saltbush dominated, with associated salt-tolerant species.
Saline Uplands and Flats	ESG4	Salt limitations are less apparent than in hills and badlands because of mixing of non-saline/nongypsic parent material (often sandstone). Shadscale and Galleta communities.
Shallow Shrublands and Woodlands	ESG5	Soils are shallow to bedrock (~ < 50 cm) and often have high coarse fragment content (~ very gravelly and coarser). Blackbrush shrublands and Pinyon-Juniper woodlands.
Sandy Grasslands and Shrublands	ESG6	Deep aeolian and alluvial generally sandy deposits, range in soil development. Grasslands with some scattered shrubs (primarily Fourwing Saltbrush, but with some Sand Sage, Blackbrush, and Ephedra on sandier sites).
Finer Shrublands	ESG7	Deep aeolian and alluvial deposits, sandy loam to clay loam textures, varying levels of soil development. Mixed shrub-grasslands, Blackbrush at the lower elevations transitioning to mostly Sagebrush at higher elevations.
Deep Rocky	ESG8	Loamy soils that are > 50 cm deep and have > 35% rock fragments by volume. Wide variety of dominant shrubs and trees, including Blackbrush, Big Sagebrush, and Juniper.
		Chihuahuan Desert
Sandy	ESG1	Basin floors and fan piedmonts; sandy surface, increased clay/carbonates in subsurface, usually sandy loam to sandy clay loam. Perennial grassland, mostly Black Grama and Dropseeds, invasion and dominance by Mesquite.
Deep sand	ESG2	Dunes, sand sheets, mantling fan piedmonts, alluvial flats, and floodplains. Sand or loamy sand textures. Mixed grass and shrub species, especially Dropseeds, Sand Sagebrush, Broom Dalea, Mesquite, and Creosotebush.
Loamy-Clayey	ESG3	Basin floors and fan piedmonts; sandy, loamy or clayey surface and loam, clay loam, or clay subsurface. Perennial grassland, mostly Tobosa, invasion by Tarbush, Mesquite, Creosotebush unless soils are very clayey.
Gravelly and Calcic	ESG4	Alluvial fans, fan piedmonts, and terraces; gravelly surface and subsurface. May have a petrocalcic horizon, but otherwise deep. Shrub savanna featuring Creosotebush and other shrubs and succulents, with Black Grama, Bush Muhly, or Tobosa.
Bedrock and Colluvium	ESG5	Hills, desert mountain slopes, flanks, and bases. Shallow to bedrock or colluvium. Large variations in soil texture and depth. Shrub savanna or shrubland depending on soil texture and depth, abundant succulents and high diversity.
Gypsic	ESG6	Basin floors, relict lakebeds, playas, gypsiferous dunes, and fan piedmonts. Includes gypsic and hypergypsic soils. Alkali Sacaton and Saltbush on gypsic soils and gypsophilous plants, including Gypsum Grama and Coldenia on hypergypsic soils.
Bottomlands	ESG7	Basin floors, floodplains, or low lying landscape positions within uplands, intermittently flooded, may be saline. Often cultivated. Alkali Sacaton/Giant Sacaton grassland in flooded areas. Tobosa in upland swales. Invasion by Mesquite and invasive Tamarisk.

† Table adapted from Duniway et al. (2016) and Bestelmeyer et al. (2016).

groups were not adequately represented by our ESG point dataset due to their spatial rarity, and consequently excluded from our analysis. After completing these filtering and matching steps, the ESG class distribution for both of our datasets was relatively uniform with all classes having adequate representation (Table 2).

ESG Covariates

We evaluated a range of geospatial datasets commonly used in digital soil mapping due to the central role of soils in determining ecological potential and modulating plant community response to change drivers. All spatial predictions in this study were modeled at 250-m resolution and covariates either sourced or resampled to conform to this resolution. Only covariates with complete coverage across the US were used to ensure a scalable framework. To determine which environmental covariates have the strongest correlation to ESG concepts, we evaluated 1846 covariates (1570 from hyper-temporal imagery) from the following data types:

- DEM-derived terrain attributes: Thirty terrain attributes (see Supplemental Table S1 for detailed list) commonly used in environmental modeling were derived from the 30-m National Elevation Dataset (NED) (Gesch et al., 2018).

- Hyper-temporal MODIS indices: MODIS 16-d time-series were acquired from 18 Feb. 2000 to 6 Mar. 2017 (393 temporal observations) for normalized difference vegetation index (NDVI; MOD13Q1), MODIS Nadir Reflectance Band 7 (mid-infrared, MIR; MCD43A4), and day- and night-time land surface temperature (LST-day, LST-night; MOD11A2).

- Aggregated MODIS indices: Long-term monthly mean and standard deviation for MODIS NDVI (MOD13Q1), MIR (Band 7; MCD43A4), and LST (daytime and nighttime; MOD11A2).

- Climate variables: WorldClim long-term (i.e., 1970–2000) monthly mean temperature and precipitation and nineteen long-term bioclimatic variables.

- SoilGrids250m: 86 SoilGrids250m soil property layers (see Supplemental Table S1 for complete list) at seven soil depths (0, 5, 15, 30, 60, 100, and 200 cm) were downloaded from SoilGrids250 web services (ftp://ftp.soilgrids.org/data/recent/).

- Other environmental covariates: lithological classes from the global lithological map (GLiM) geodatabase

Table 2. Sample size distribution and areal extent for the ecological site group (ESG) classes in the Colorado Plateau and Chihuahuan Desert study areas.

	ES1	ES2	ES3	ES4	ES5	ES6	ES7	ES8
				Colorado Plateau				
Sample plots (N)	138	123	344	357	111	163	99	70
Area (Mha)†	0.385	0.744	1.367	0.582	0.383	0.660	0.123	0.176
Area per sample (ha)	2796	6048	3973	1631	3448	4046	1242	2519
				Chihuahuan Desert				
Sample plots (N)	269	133	774	655	297	151	310	–
Area (Mha)†	1.004	0.422	1.414	1.922	1.346	0.397	0.593	–
Area per sample (ha)	3731	3176	1827	2935	4532	2631	1912	–

† ESG areas calculated based on SSURGO dominant condition; N, number of observations; Mha, mega-hectare (ha × 10^6).

(Hartmann and Moosdorf, 2012), MODIS IGBP land cover classes (MCD12Q1), average soil and sedimentary-deposit thickness (m) (Pelletier et al., 2016), and USGS aeroradiometric grids (Hill et al., 2009).

A detailed description of each covariate evaluated in this study is presented in Supplemental Table S1. In this study we included several modeled covariate layers, many of which were derived using the primary covariates in this study (e.g., LST, MIR, climate, lithology, etc.). While these modeled layers likely exhibit some collinearity with our primary covariates, they can also contain information that explains a unique portion of the variance within our dependent variable. Consequently, in models containing both modeled and primary covariates we relied on our feature selection routine (described below) to reduce the level of multicollinearity. Terrain attributes were calculated using SAGA GIS software (Conrad et al., 2015) and resampled to 250-m resolution using mean resampling in the 'raster' package for R (Hijmans, 2019). Aggregated MODIS indices were processed in Google Earth Engine, which included quality assurance processing, temporal averaging, resampling to 250-m resolution using bilinear interpolation when required, and downloading finished imagery. Hyper-temporal MODIS indices were processed using the 'MODIS' package for R following the steps described by Maynard et al. (2016).

Model Building

To determine the relative strength of each dataset in predicting ESGs, we evaluated the performance of four machine learning models: support vector machines (SVM), random forest (RF), extreme gradient boosting (XGB), and an ensemble (ENS) of these models based on the average of their prediction probabilities. We focused on machine learning models rather than a variety of geospatial modeling approaches (e.g., geostatistics) for two main reasons. First, a growing body of literature has shown that machine learning techniques outperform traditional geostatistical or spatial regression models (Niculescu-Mizil and Caruana, 2005; Oliveira et al., 2012; Brungard et al., 2015; Heung et al., 2016). This is particularly true for the modeling of multi-class categorical variables. Second, the spatial distribution of our training data is highly uneven (Fig. 1), which is problematic for geostatistical techniques which require sufficient spatial autocorrelation to accurately perform.

For the construction and optimization of each model, all data preprocessing, feature selection, and hyper-parameter tuning

were performed within each cross-validation fold, thus preventing any data leakage which could result in overoptimistic performance results. Data preprocessing, which included centering and scaling of all numeric covariates, was performed prior to modeling (Kuhn and Johnson, 2013). Feature selection was performed using the "information.gain" filtering method from the 'FSelector' package for R (Romanski and Kotthoff, 2018). This method calculates an entropy-based information gain between each covariate and target variable. Non-informative covariates were removed by applying a threshold where only covariates whose importance exceeded that threshold were retained. A tuning procedure using threefold cross-validation was used to determine optimal threshold values for each model. Hyperparameter tuning was performed using the 'mlrHyperopt' R package (Richter et al., 2017). The search space for each hyperparameter was selected based on published values and optimized using a random search with 50 iterations using threefold cross-validation. SVM, RF, and XGB models were implemented using the 'e1071', 'randomForest', and 'xgboost' packages in R, respectively; and all modeling was performed using the 'mlr' package in R (Bischl et al., 2016).

To decipher which covariates were most important in accounting for unique variance within our ESG models, we ran a series of separate models using different combinations of covariate groups. Our goal was to partition our large covariate matrix into subgroups that represented unique ecological and pedological variability within our study areas. These subgroups included:

- Complete covariate set (All; $n = 1846$): terrain attributes, lithology, climatic variables, USGS aeroradiometric grids, soil thickness, SoilGrids250m, and both hyper-temporal and long-term monthly MODIS250m NDVI, MIR, daytime LST and nighttime LST.

- Hyper-temporal MODIS250m (Hyper; $n = 1570$): hyper-temporal MODIS250m NDVI, MIR, daytime LST and nighttime LST.

- Hyper-temporal NDVI (NDVI; $n = 393$): hyper-temporal MODIS250m NDVI

- Abiotic covariates (Abiotic; $n = 170$): terrain, lithology, climatic variables, USGS aeroradiometric grids, soil thickness, SoilGrids

- SoilGrids250m (SG; $n = 86$): 88 SoilGrids250m soil property raster layers

- Digital soil mapping (DSM; $n = 266$): terrain attributes, lithology, climatic variables, USGS aeroradiometric grids, soil thickness, SoilGrids250m, and long-term monthly NDVI, MIR, daytime LST and nighttime LST.

Model Accuracy

Cross-Validation Accuracies

Model accuracy for all datasets and model types was assessed using 10-fold cross-validation. Model predictions and corresponding observations from each cross-validation fold were compiled and used to calculate several performance metrics, including: overall map accuracy, quantity disagreement, allocation disagreement, producer's accuracy, and user's accuracy. Overall model performance was evaluated using overall map accuracy, quantity disagreement, and allocation disagreement, while class-wise model performance was evaluated using producer's accuracy, user's accuracy, quantity disagreement, and allocation disagreement.

Overall map accuracy is the proportion of observation points at which the model predicts the correct ESG class. Model error can be attributed to two different sources of randomness; the random distribution of the quantity of each class and the random spatial allocation of the classes. Pontius et al. (2011) presented two measures for quantifying these sources of model disagreement, termed quantity disagreement and allocation disagreement (Pontius et al., 2011; Warrens, 2015). Quantity disagreement represents the amount of difference between the validation and prediction data that is due to a less than perfect match in the proportion of classes. Allocation disagreement represents the amount of difference between the validation and prediction data that is due to the less than optimal match in the spatial allocation of classes. Quantity disagreement (QD) is calculated as:

$$QD = \frac{1}{2} \sum_{i=1}^{c} \left| p_{i+} - p_{+i} \right| \qquad [1]$$

where p_{i+} and p_{+i} represent the row and column totals of the error matrix for the ith class for c number of classes. Values for quantity disagreement can range from 0 to 1, where a value of 0 represents perfect agreement in the proportion of coverage for each class between the validation and prediction data. Allocation disagreement (AD) is calculated as:

$$AD = \left[\sum_{i=1}^{c} \min\left(p_{i+}, p_{+i} \right) \right] - C \qquad [2]$$

where C is the overall agreement or correct classification. Values for allocation disagreement can range from 0 to 1, where a value of 0 represents perfect agreement in the spatial allocations for each class between the validation and prediction data. The producer's accuracy, also known as recall, measures the proportion of correctly identified observations from a class relative to all observations of that class. Values of producer's accuracy range from 0 to 1, with values of 1 indicating a perfect match. The user's accuracy, also known as precision, is the proportion of correctly identified observations from a class relative to all predictions of that class. Values of user's accuracy range from 0 to 1, with values of 1 indicating a model with perfect precision. The pre-

dicted probabilities of each ESG class occurrence were used to calculate the area under the precision-recall curve (PR-AUC). The PR-AUC summarizes the trade-offs between producer's accuracy and user's accuracy for a model using different probability thresholds. Values of PR-AUC close to 1 indicated high model performance, while values <0.5 indicate poor performance. We implemented a multi-class calculation of PR-AUC using the 'multiROC' R package (Wei and Wang, 2018). Cross-validation folds were micro-aggregated (Forman and Scholz, 2010), while PR-AUC for each ESG class and the total model were macro-averaged (i.e., one-vs-rest). We calculated nonparametric bootstrap 95% confidence intervals for each PR-AUC using the basic bootstrap method and 100 replicates.

Finalized machine learning models were used to estimate the probability of occurrence for each ESG class at each 250-m grid cell covering our study areas. The ESG class with the largest probability at each grid cell was assigned the predicted class. Estimates of model uncertainty for our spatial predictions were generated using the Scaled Shannon Entropy Index (SSEI) (Shannon, 1948; Kempen, 2011; Hengl et al., 2017):

$$SSEI_s(x) = -\sum_{k=1}^{K} p_k(x) * \log_K \left[p_k(x) \right] \qquad [3]$$

where K is the number of possible classes, $\log_K$ is the logarithm to base K and p_k is the probability of class k. Values of SSEI can range from 0 to 1, with 0 indicating no uncertainty (one class has p_k equal to 1 and all remaining are 0), and 1 indicating maximum uncertainty (all classes have equal probabilities) (Kempen, 2011). SSEI is an internal accuracy measure derived from model probabilities and therefore should not be confused with classification accuracy assessment or validation. In other words, it provides an indication of how certain the model is in its predictions regardless of whether those predictions are correct. Classifier calibration was used to estimate the agreement between the predicted probabilities of each class relative to the rate at which that class occurs. This was calculated by grouping data points with similar predicted probabilities for each class and plotting these against the observed frequency for each class. Classifier calibration was calculated using the 'mlr' R package.

External-Validation Accuracies

Independent ESG datasets were obtained for both study areas to perform external model validation. For the Colorado Plateau, an independent dataset of 356 points was obtained for external validation where soil and vegetation surveys had been conducted and ecological site and ecological site group designations assigned. The network of points, previously described in Miller et al. (2011) and Bowker et al. (2012), spans an area of ~1500 km^2 along the southern edge of Canyonlands National Park where a wide range in elevation, climate, vegetation, and soils occur (Fig. 1b). External-validation data for the Chihuahuan Desert was obtained from several sources, including, (i) sites established for the project 'Restore New Mexico' (Coffman et al., 2014) and (ii) a collection of field data from the

Jornada Experimental Range as described in Maynard and Karl (2017) (Fig. 1c). These locations all contained validated ecological site concepts based on a full characterization of the soil at each plot.

Model-Agnostic Interpretability Methods

Most machine learning algorithms are inherently complex, producing 'black box' models that prevent a direct explanation of how covariates contribute toward model predictions or affect model performance (Casalicchio et al., 2018; Fisher et al., 2018; Greenwell et al., 2018). While some machine learning algorithms provide model specific variable importance calculations, these measures are often not comparable across model types or can give seemingly incoherent results when several different models produce a strong fit to the data (Fisher et al., 2018). To correct for these deficiencies, model-agnostic interpretability methods have been developed (Goldstein et al., 2015; Fisher et al., 2018). Permutation-based variable importance was first introduced by Breiman (2001) in the Random Forests algorithm. Building on this framework, Fisher et al. (2018) recently proposed a model-agnostic version of the permutation-based variable importance (MAPVI) measure, but applied at the aggregate model-level and for a general loss function. In this approach, a covariate's importance in a model is measured by calculating the increase in prediction error after randomly permuting its values. Covariates with large increases in error after permutation are considered "important" because the model relied on them for the prediction. Conversely, if the model error remains unchanged after permutation, the covariate is considered "unimportant" because it was ignored for the prediction. We used the MAPVI method to determine which covariates were most important in predicting ESG classes for the different machine learning models. The MAPVI algorithm was implemented using the 'iml' R package (Molnar et al., 2018).

SSURGO Ecological Site Maps

Currently, use of ecological site data in a spatial context requires some form of generalization to resolve the one-to-many relationship between SSURGO SMUs and soil components. In the correlation between ecological site concepts and SSURGO SMU components, multiple SMU components concepts (e.g., variations of soil series, phases, or other higher taxonomic concepts) are often correlated to a single ESD. Thus, multiple soil components within a SMU can be correlated to the same ESD. Consequently, we employed the dominant condition generalization approach, where the ecological site was selected by the most frequent condition (i.e., percent area occurrence) within a map unit. SSURGO ecological site and SMU component data were downloaded from SoilDataAccess (Accessed 12 Oct. 2017), merged based on the representative component percentage for each SMU, and aggregated by dominant condition. In rare cases (<0.5%) where multiple ecological site concepts resulted in equal representative percentages within a SMU, ESD selection was based on consulting local soil survey office and analysis of correlated minor components. The ESD-ESG look-up tables developed from the Ecological Site Grouping workshops were then used to assign the appropriate ESG group to each SMU ecological site class. The accuracy of the SSURGO dominant-condition ESG maps were evaluated using both the NASIS-SSURGO training dataset and the external-validation datasets.

RESULTS

Model results from our ESG mapping framework support our initial hypothesis that ESGs can be accurately predicted using readily available geospatial data. Cross-validation and external-validation accuracies for all combinations of model types and covariate datasets are presented in Fig. 3. Cross-validation accuracies ranged from 64 to 70% for the Colorado Plateau and from 72 to 79% for the Chihuahuan Desert across all datasets and model types (Fig. 3a-b).

External-validation accuracies ranged from 37 to 56% for the Colorado Plateau and from 34 to 44% for the Chihuahuan Desert across all datasets and model types (Fig. 3c-d). Our analysis of multiple datasets and model types revealed that the range in accuracy between datasets was much higher than between models, highlighting the importance of covariate selection in developing accurate ESG models (Fig. 3). In all but one case the ensemble (ENS) model produced the highest cross-validation accuracies at both study sites (Fig. 3a-b). In general, the range in model accuracy across the four models for a given dataset was small (<6%) and often the top performing model was only one or two percentage points more accurate than the next model. In our external validation, the top performing model alternated between the ENS and RF models across our datasets (Fig. 3c-d). In terms of covariate datasets, the complete dataset (All) generally produced the highest accuracies both across sites and validation approaches (Fig. 3). The DSM dataset produced the second highest accuracy and NDVI the lowest at both sites in our cross-validation analysis. However, in the external-validation NDVI produced one of the highest accuracies and the abiotic dataset the lowest for the Colorado Plateau; whereas for the Chihuahuan Desert NDVI accuracy remained low while the abiotic datasets continued to have high accuracies. While the complete covariate dataset (All) generally produced the highest total model accuracies, it contains 1846 covariates which significantly increased computational time relative to other datasets, like DSM ($n = 268$), which produced only slightly lower accuracies. Given the need to produce scalable models that are both accurate and parsimonious, we focused additional analysis on the DSM dataset.

Individual ESG cross-validation accuracies for the Colorado Plateau and Chihuahuan Desert are presented in Supplemental Fig. S1 and S2, respectively. At both study sites, the effect of covariate dataset and model type on prediction accuracies varied slightly between ESG classes, indicating potential differences in the importance of biotic vs. abiotic variables for predicting different ESGs. A more detailed assessment of individual ESG prediction accuracies was performed on results from the ENS model of the DSM dataset (Tables 3 and 4). The ENS-DSM model produced both high producer's and user's accuracy for the majority of ESG classes. However, a few ESG classes were appreciably lower relative to the other classes at both study sites. In the Colorado Plateau ESG6 (Outcrops and Slopes) had the lowest producer's accuracy (0.48), while in the
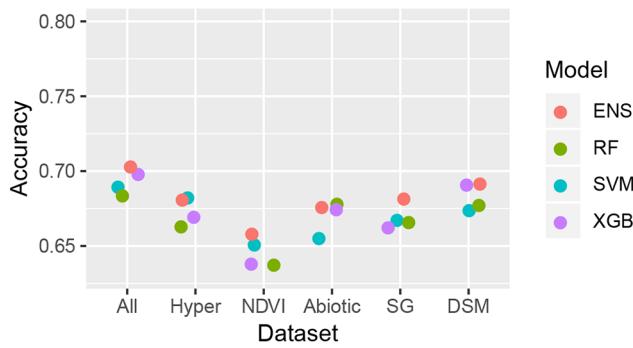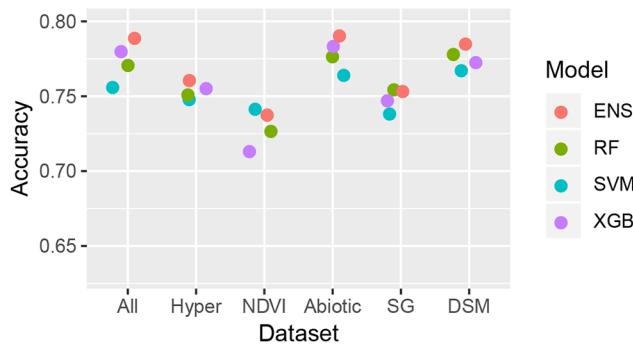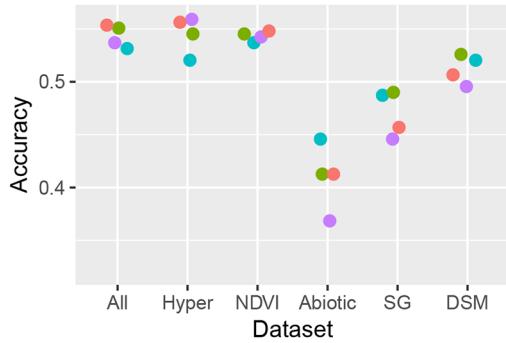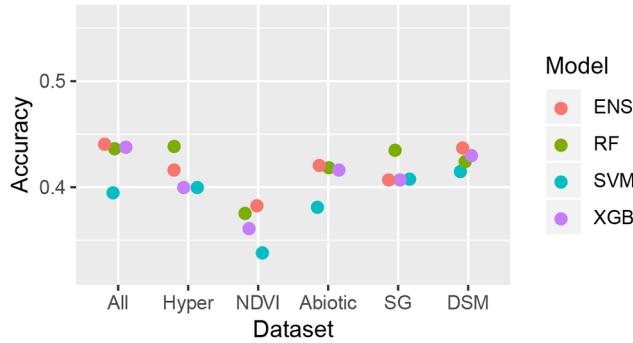
Fig. 3. Comparison of cross-validation model accuracy between covariate datasets (*x* axis) and machine learning models (colored circles) for (a) the Colorado Plateau and (b) Chihuahuan Desert study areas. Comparison of external-validation accuracies for (c) the Colorado Plateau and (d) Chihuahuan Desert study areas.

Chihuahuan Desert ESG7 (Bottomland) had the lowest producer's accuracy (0.56). In general, values of producer's accuracy and user's accuracy closely followed each other both in terms of their absolute values and relative differences. Analysis of the area under the precision-recall curve for each ESG class and the overall models showed a similar trend to values of producer's accuracy and user's accuracy and had narrow 95% confidence intervals around each PR-AUC value (Tables 3 and 4). At both study sites, analysis of model disagreement revealed that the quantity disagreement was low across all classes, while allocation disagreement tended to be higher, particularly for classes with high sample sizes (Tables 3 and 4). For example, in the Colorado Plateau ESG3 (Shallow Shrublands and Woodlands) has one of the largest sample sizes and highest allocation disagreement (0.17). This high allocation disagreement for ESG3 can be seen where six out of the seven other classes has ESG3 as their most dominant misclassified class (Table 3). A similar pattern was observed in the Chihuahuan Desert for ESG3 (Loamy-Clayey) and ESG4 (Gravely and Calcic) (Table 4).

In both study sites, SSURGO derived ESG maps revealed that large areas currently do not have spatial information on ESG classes (Fig. 4). These unmapped areas occur where SSURGO has yet to be completed or where ESD correlations have yet to be established (gridded in Fig. 4). Due to this incomplete spatial coverage, only a subset of our internal- and external-validation datasets could be used to query SSURGO delineated ESG classes. To ensure an accurate comparison between the accuracy of

SSURGO ESG delineations and our modeled ESG distributions, we subset our training–validation dataset to match our SSURGO dataset and recalculated model accuracy statistics. Comparisons of map accuracy between the SSURGO ESG map and our ESG model predictions from the ENS model using the DSM dataset revealed strong similarities (Table 5). In the Colorado Plateau, the DSM-ENS model had slightly higher cross-validation accuracy (0.70 vs. 0.66) and slightly lower external-validation accuracy (0.51 vs. 0.66) relative to SSURGO. In the Chihuahuan Desert, DSM-ENS and SSURGO accuracies were nearly identical with cross-validation accuracies of 0.82 and 0.83, and external-validation accuracies of 0.45 and 0.43 for DSM-ENS and SSURGO, respectively. However, in both study areas the external-validation data only covered smaller subsets of the overall study areas and may not be fully representative of all areas.

Spatial Predictions of ESGs

Spatial predictions of ESG classes from our four models using the DSM dataset are presented in Fig. 5 and 6 for the Colorado Plateau and Chihuahuan Desert, respectively. In general, the spatial distribution of ESGs conforms to our expectations based on exert knowledge of these study areas. In the Colorado Plateau, the RF and XGB predictions (both tree based models) displayed similar distributions, while the SVM model produced a noticeably different spatial distribution (Fig. 5). Spatial differences in prediction surfaces were less pronounced between models for

Table 3. Colorado Plateau class-wise accuracy statistics for the ensemble (ENS) model of the digital soil mapping (DSM) dataset.†

Class	N	PA	UA	PR-AUC‡	QD	AD	First class	Second class
ES1	138	0.84	0.83	0.93 ± 0.02	0.00	0.03	ES1 (84%)	ES3 (09%)
ES2	123	0.73	0.70	0.74 ± 0.09	0.00	0.05	ES2 (73%)	ES3 (11%)
ES3	344	0.65	0.62	0.68 ± 0.06	0.01	0.17	ES3 (65%)	ES4 (13%)
ES4	357	0.78	0.70	0.83 ± 0.03	0.03	0.11	ES4 (78%)	ES3 (11%)
ES5	111	0.61	0.71	0.72 ± 0.07	0.01	0.04	ES5 (61%)	ES3 (19%)
ES6	163	0.48	0.57	0.52 ± 0.09	0.02	0.08	ES6 (48%)	ES3 (26%)
ES7	99	0.64	0.76	0.76 ± 0.08	0.01	0.03	ES7 (64%)	ES4 (21%)
ES8	70	0.83	0.84	0.89 ± 0.07	0.00	0.02	ES99 (83%)	ES3 (04%)
Overall	1405	0.70	–	0.76 ± 0.03	0.04	0.27	–	–

† *N*, number of observations; PA, producer's accuracy; UA, user's accuracy; PR-AUC, multiclass area under the precision-recall curve; QD, quantity disagreement; AD, allocation disagreement; First and Second, most probable groups and percentage of ESG observations predicted to that group.

‡ Multiclass implementation of the area under the precision-recall curve (i.e., PA vs. UA) plus/minus the bootstrapped 95% confidence interval.

the Chihuahuan Desert (Fig. 6). The averaging effect of the ENS model can be seen at both sites. In the Colorado Plateau, despite large gaps in SSURGO coverage, a fairly good correspondence between SSURGO and model result can be seen (Fig. 4 and 5). One noticeable difference is in the distribution of ES6 (Outcrops and Slopes), where our model results show significantly less coverage relative to SSURGO. ESG6 had the lowest producer's accuracy and a high rate of misclassification as ESG3 (Shallow Shrublands and Woodlands). Large areas mapped as ESG6 in the SSURGO map were predicted as ESG3 in our predicted maps, highlighting the need for further refinement with respect to these ESG classes. Correspondence between SSURGO and spatial predictions in the Chihuahuan Desert were high, particularly in the eastern half of the study area where most of our training–validation points occurred (Fig. 4 and 6).

It should be noted that the NASIS point dataset used in this study was collected using a purposive sampling design (i.e., rapid field transects) which can introduce potential bias in our model estimates and uncertainty as to how well our sample of points represents the larger population (Brus et al., 2011). Spatial assessments of prediction uncertainty were made with the scaled Shannon Entropy index (Fig. 7). Mapped areas with low uncertainty (low Shannon Entropy) correspond to areas with a higher density of training data. This is particularly noticeable in the Chihuahuan Desert site where the relatively dense and even distribution of points in the eastern portion of the study area resulted in very low levels of uncertainty. In contrast, very few observations in the western portion of the study area resulted

in higher levels of model uncertainty. Classifier calibration plots provide additional insight into the accuracy of modeled prediction probabilities. A classification model is considered 'calibrated' when the predicted probability of a class matches the expected frequency of that class, in which case predicted probabilities can be directly interpreted as a confidence level. Thus, for a perfectly calibrated model each class would plot along the 1:1 line (dotted black line in Fig. 7c-d), and deviations from this would indicate that the model is either over or under estimating the probability of occurrence for a particular class. The classifier calibration plot for the Colorado Plateau shows a general trend of slightly underestimating probabilities at moderate-to-low probability values, and overestimating probabilities at high probability values. This was particularly true for ESG5 and ESG6, which also had the lowest producer's accuracies (Fig. 7c; Table 3). In contrast, the classifier calibration plot for the Chihuahuan Desert was better calibrated with most groups aligning with the 1:1 line (Fig. 7d)

The uneven spatial distribution of points across our sites may diminish our ability to quantify the range of covariate space that characterizes each ESG. Thus, despite adequate representation of each ESG class in terms of point observations, it seems likely that ESGs found within regions of uncharacterized geographic space will inhabit slightly different regions of covariate space relative to the covariate space characterized for that ESG from the training–validation dataset. Extrapolation of our models to these uncharacterized regions produces higher levels of uncertainty, as shown in our maps of Shannon Entropy, and higher

Table 4. Chihuahuan Desert class-wise accuracy statistics for the ensemble (ENS) model of the digital soil mapping (DSM) dataset. †

Class	N	PA	UA	PR-AUC‡	QD	AD	First group	Second group
ES1	269	0.82	0.81	0.88 ± 0.03	0.00	0.04	ES1 (82%)	ESG2 (05%)
ES2	133	0.79	0.86	0.88 ± 0.05	0.00	0.01	ES2 (79%)	ESG1 (13%)
ES3	774	0.81	0.78	0.88 ± 0.02	0.01	0.11	ES3 (81%)	ESG4 (09%)
ES4	655	0.82	0.77	0.86 ± 0.03	0.02	0.09	ES4 (82%)	ESG3 (09%)
ES5	297	0.79	0.83	0.88 ± 0.04	0.01	0.04	ES5 (79%)	ESG4 (16%)
ES6	151	0.88	0.80	0.89 ± 0.04	0.01	0.01	ES6 (88%)	ESG3 (05%)
ES7	310	0.56	0.71	0.73 ± 0.05	0.03	0.05	ES7 (56%)	ESG3 (28%)
Overall	2589	0.79	–	0.86 ± 0.02	0.04	0.18	–	–

† *N*, number of observations; PA, producer's accuracy; UA, user's accuracy; PR-AUC, multiclass area under the precision-recall curve; QD, quantity disagreement; AD, allocation disagreement; First and Second, most probable groups and percentage of ESG observations predicted to that group.

‡ Multiclass implementation of the area under the precision-recall curve (i.e., PA vs. UA) plus/minus the bootstrapped 95% confidence interval.

(a) Colorado Plateau SSURGO ESG

ESG
- Saline hills and badlands
- Saline uplands and flats
- Shallow shrublands and woodlands
- Sandy grasslands and shrublands
- Finer shrublands
- Outcrops and slopes
- Bottoms and flats
- Deep rocky

(b) Chihauhaun Desert SSURGO ESG

ESG
- Sandy
- Deep sand
- Loamy−clayey
- Gravelly and calcic
- Bedrock and colluvium
- Gypsic
- Bottomland
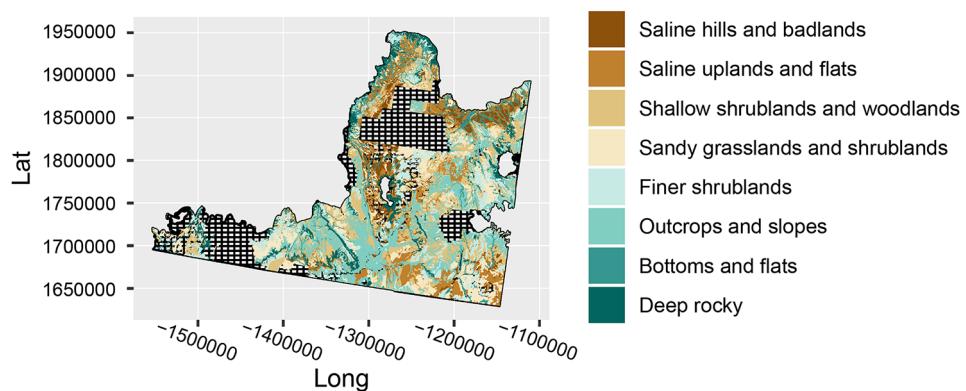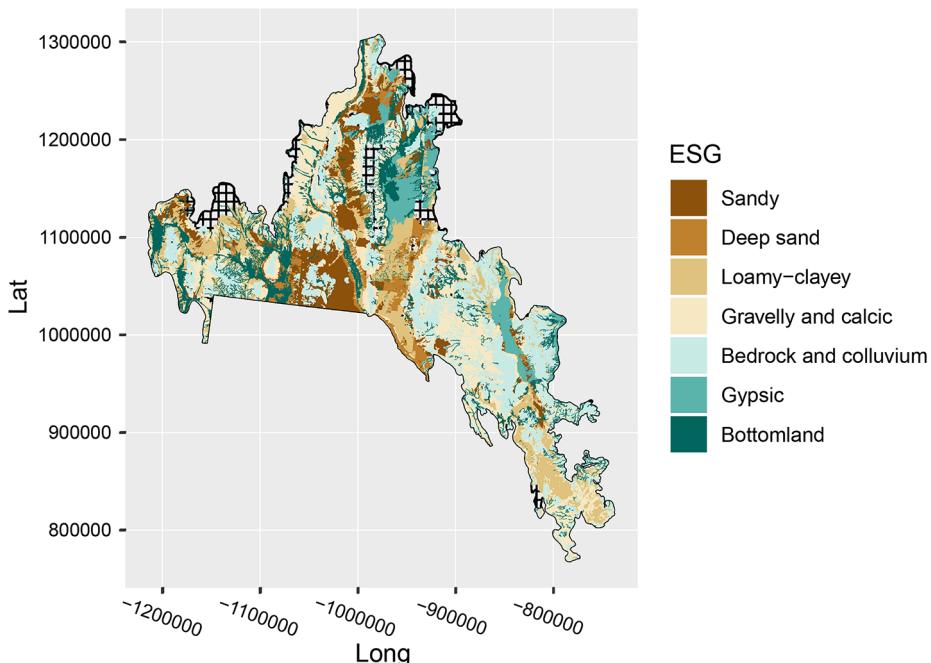
Fig. 4. SSURGO dominant condition maps of ecological site groups in (a) Colorado Plateau, and (b) Chihuahuan Desert study areas.

rates of misclassification, as demonstrated by low external-validation results for the Chihuahuan Desert.

Covariate Importance

The top 10 predictors from the model-agnostic permutation-based variable importance (MAPVI) calculation of the SVM, RF, and XGB models on the DSM dataset are presented in Fig. 8. At both study sites, covariate importance dropped significantly after the first covariate. This was most pronounced for the RF and XGB models (Fig. 8). In the Colorado Plateau, the top-ranking covariate was different between the three models, with MRRTF as most important for SVM (terrain attribute), NDVI-12-M for RF (MODIS imagery), and TAXOUSDA for XGB (SoilGrids250m). While the top ten covariates largely differed between the Colorado Plateau models, there were some similarities in terms of covariate types. Terrain attributes, MODIS imagery, and SoilGrids250m layers were all important

to different extents in each model. However, some notable differences included the importance of USGS aeroradiometric grids in RF and the importance of climate variables in XGB. For the Chihuahuan Desert, TAXOUSDA was the top-ranking covariate in all three models (Fig. 8). The importance of subsequent

Table 5. Cross-validation and external-validation accuracies from the SSURGO dominant-condition ecological site group (ESG) map and model predictions from the ensemble (ENS) model of the digital soil mapping (DSM) dataset at both study sites.

Site	Source	Training-validation†	External-validation
Colorado Plateau	SSURGO	0.66	0.66
Colorado Plateau	DSM-ENS	0.70	0.51
Chihuahuan Desert	SSURGO	0.83	0.43
Chihuahuan Desert	DSM-ENS	0.82	0.45

† Training-validation accuracy for SSURGO consisted of comparing the SSURGO ESG dominant condition at each training point relative to its actual ESG class. Training-validation accuracy for the DSM-ENS model was the 10-fold cross-validation accuracy of the training dataset.
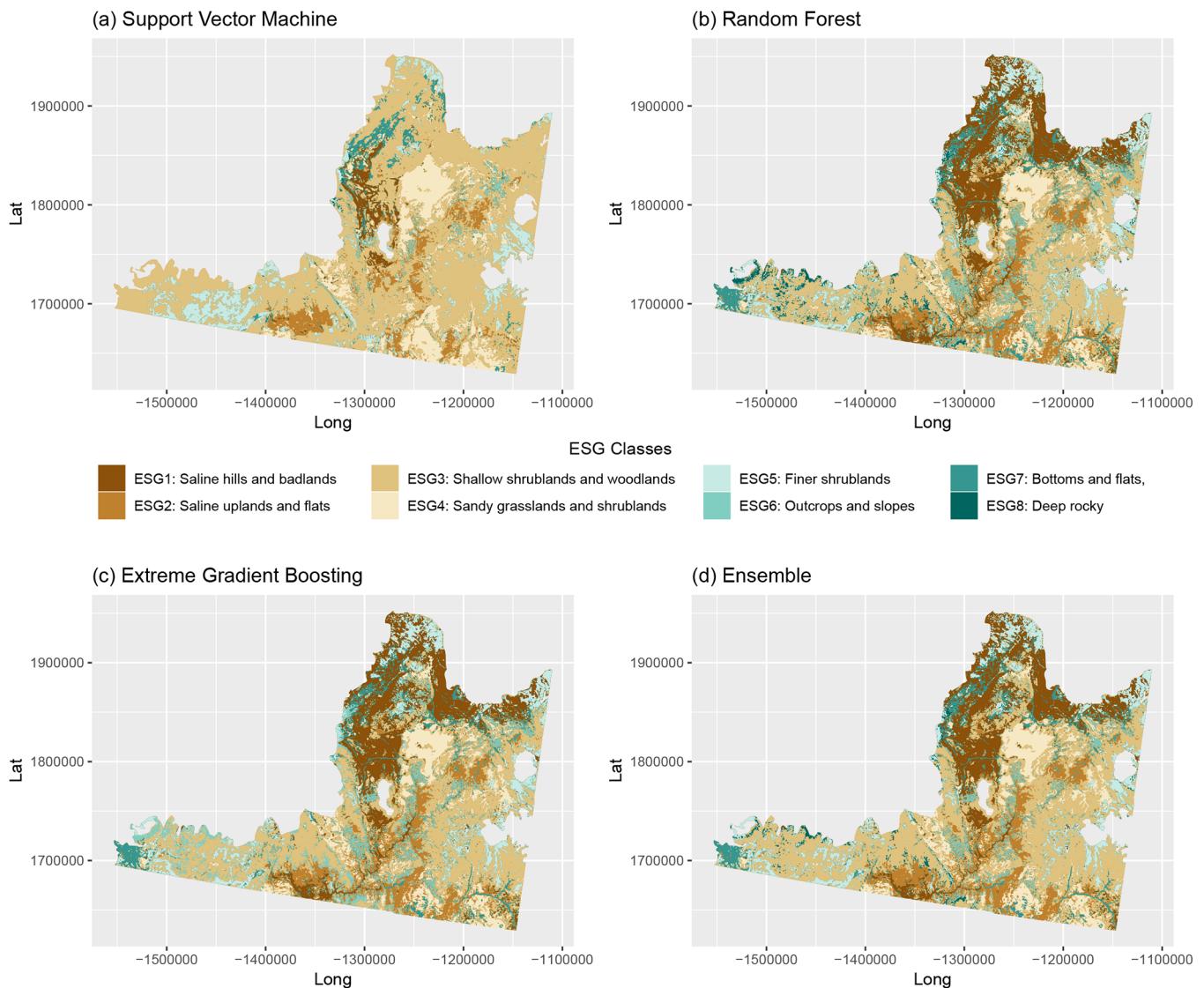
Fig. 5. Spatial prediction maps of the Colorado Plateau study area for, (a) support vector machine, (b) random forest, (c) extreme gradient boosting, and (d) an ensemble of all models using the digital soil mapping (DSM) dataset.

covariates differed widely between the three models, where terrain attributes, SoilGrids250m layers, soil thickness, and climate variables assumed different levels of importance.

ESG Misclassification

Our ability to predict ESGs is dependent on the establishment of ecologically distinct thematic groupings that can be detected and differentiated using available covariate datasets. Two sources of error can occur that results in low prediction accuracies: (1) conceptually overlapping thematic classes that result in significant overlap in covariate space, and (2) deficiencies in our covariate data preventing the identification of ecologically meaningful differences between ESG classes. Here we present examples from both study areas of thematic overlap and covariate limitations that may result in higher rates of misclassification within ESGs.

In the Colorado Plateau, ESG3 (Shallow Shrublands and Woodlands) has one of the lower classification accuracies (65%) and the largest sample size from our point dataset (Table 3). ESG3 represents the aggregation of 12 established ecological sites, spanning three precipitation/climate zones (ranging from 15 cm to 40 cm in MAP), with soil profiles that are shallow to bedrock and soil textures ranging from loam to sand. Of the 12 ecological sites, the Desert Shallow Sandy Loam (Blackbrush) site had the highest classification error (68%), with most observations being misclassified as ESG2 (Saline Uplands and Flats) (Fig. 9a). Analysis of ESG2 misclassifications showed that five out of seven ecological sites (Fig. 9b, purple text) had some percentage misclassification as ESG3. The most dominant ecological site for ESG2 from our point dataset is the Desert Shallow Sandy Loam (Shadscale), which has similar characteristics to the Desert Shallow Sandy Loam (Blackbrush), with the main difference being the Shadscale is found in more saline soils. Soil salinity is an important property used to distinguish between ESG2 and ESG3, thus much of the misclassification between these two groups could possibly be corrected if a soil salinity covariate was available. Furthermore, the highest number of misclassifications within ESG3 was from the Semidesert Shallow Sandy Loam (Utah Juniper, Blackbrush) ecological site which also comprises the largest number of point ob-
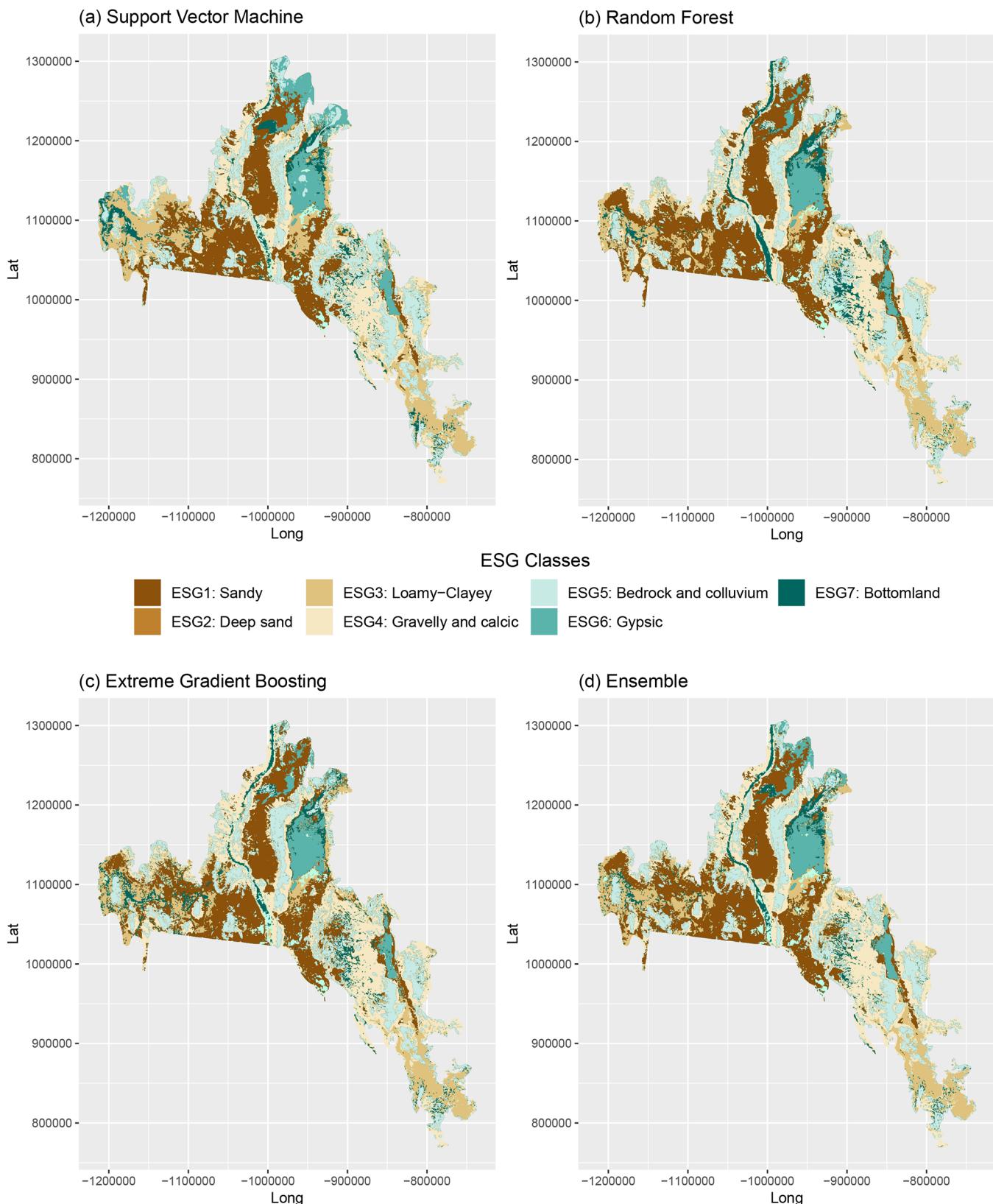
Fig. 6. Spatial prediction maps of the Chihuahuan Desert study area for, (a) support vector machine, (b) random forest, (c) extreme gradient boosting, and (d) an ensemble of all models using the digital soil mapping (DSM) dataset.

(a) Colorado Plateau SSEI
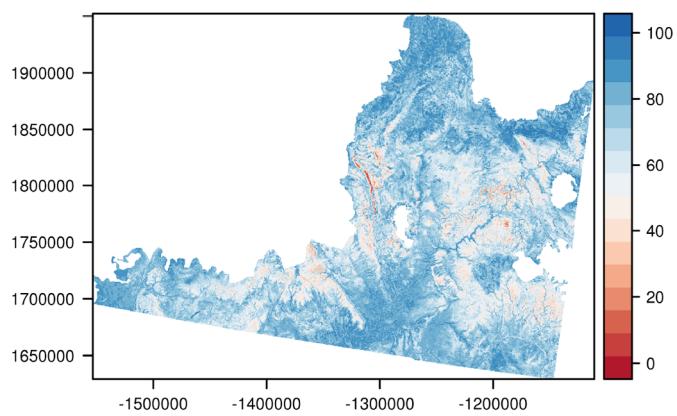
(b) Chihuahuan Desert SSEI
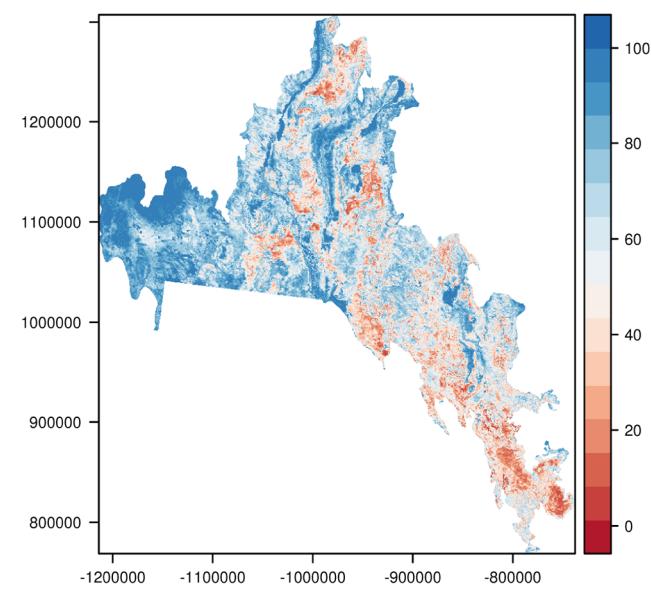
(c) Colorado Plateau Classifier Calibration
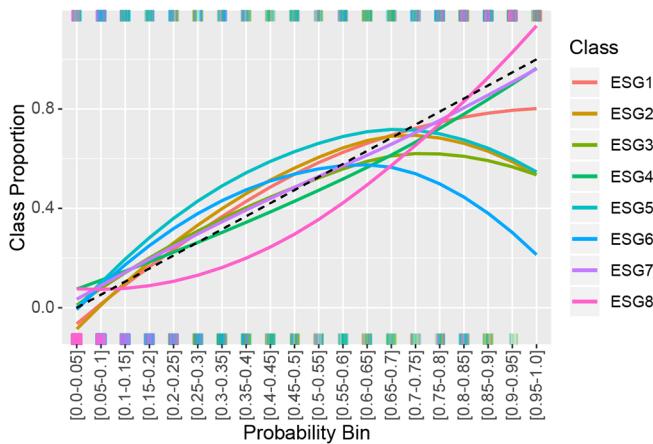
**(d) Chihuahaun Desert Classifier Calibration**
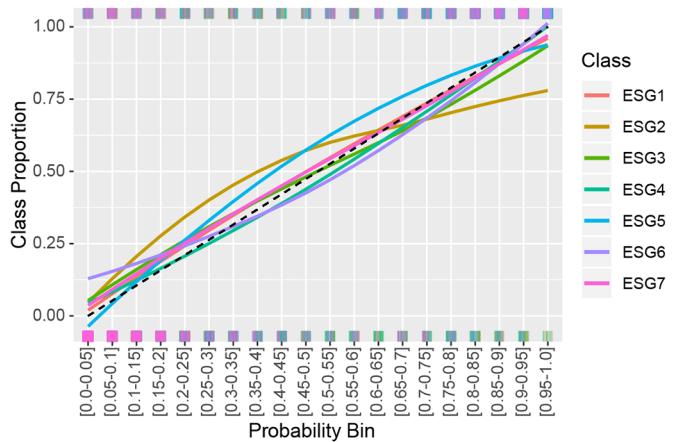
Fig. 7. Scaled Shannon's Entropy index for (a) the Colorado Plateau, and (b) Chihuahuan Desert study sites. Classifier calibration plot for (c) the Colorado Plateau and (d) Chihuahuan Desert study areas

servations for ESG3. The Semidesert Shallow Sandy Loam (Utah Juniper, Blackbrush) was dominantly misclassified as ESG4 (Sandy Grasslands and Shrublands), which on closer examination reveals that 47% of ESG4 misclassifications were misclassified as ESG3. There are many complex areas in this region where shallow soils with more blackbrush and juniper vegetation have subtle gradients into deeper areas of soils where grasses become more dominant possibly creating confusion at the interface between the two sites. Thus, similarities between ecological sites across ESGs prevent the establishment of crisp thematic or taxonomic separation. Consequently, misclassifications can result when ESGs overlap in thematic space, as well as potential deficiencies in our ability to characterize ecological differences with our current covariate dataset and at 250-m resolution.

In the Chihuahuan Desert, ESG7 (Bottomland) had the lowest class-wise accuracy (58%) and was predominantly mis-

classified as ESG3 (Loamy-Clayey, 28%). ESG7 represents the aggregation of 12 established ecological sites, including draws, flats, swales, bottomlands, and meadow sites that are characterized by fine textured soils (e.g., loamy) that may include some salt accumulation. Of these 12 ecological sites, the Loamy Swale (Mixed Prairie) site was responsible for 65% of the classification error, with the majority (i.e., 84%) being misclassified as ESG3 (Loamy-Clayey) (Fig. 10a). Analysis of ESG3 misclassifications showed that six out of nine ecological sites (Fig. 10b, purple text) had some percentage misclassification as ESG7, indicating that these ecological sites share some similarities with the ecological sites in ESG7.

The most dominant ecological site for ESG3 from our point dataset is the Loamy Slope (Mixed Prairie), which has similar characteristics to the Loamy Swale (Mixed Prairie) site which may partly explain its high misclassification rate.
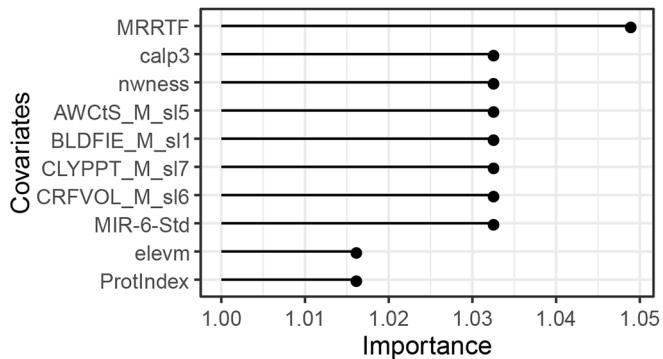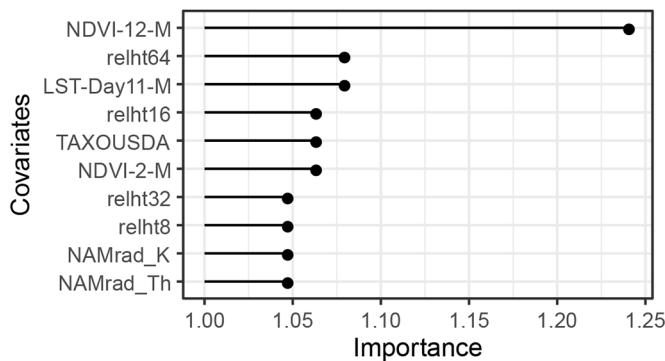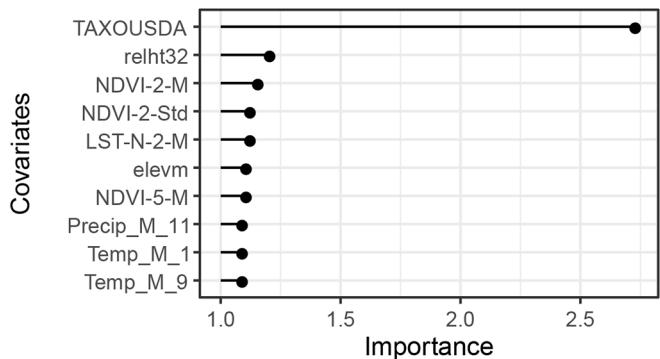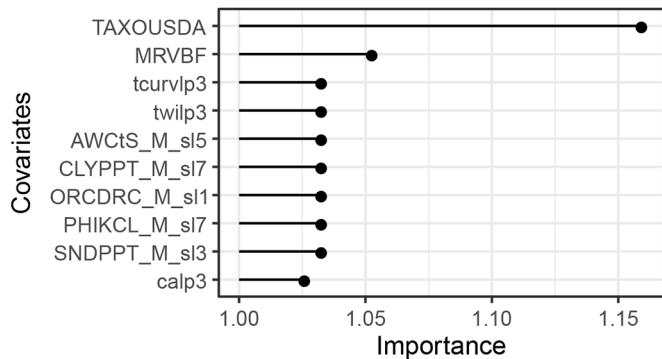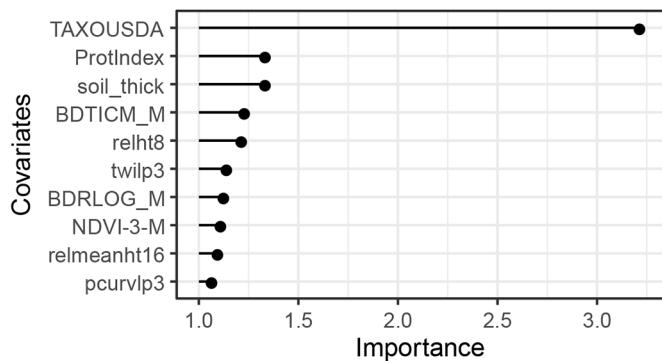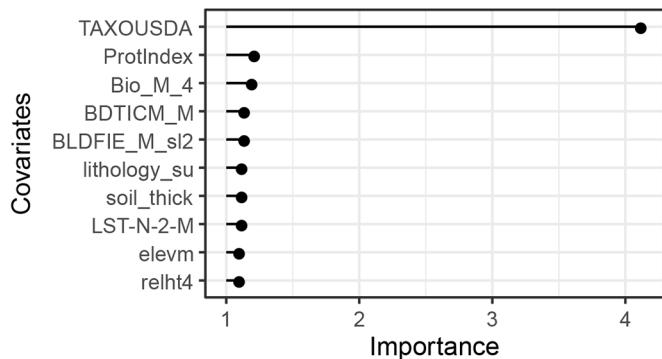
Fig. 8. Model agnostic permutation-based variable importance measurements from the support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB) models of (a-c) the Colorado Plateau and (d-f) the Chihuahuan Desert.

DISCUSSION
Building a Nationally Scalable Ecological Site Modeling Framework

In this study we developed and tested a modeling framework for mapping ecological sites, in this case via ESGs. Our objective was to develop a nationally consistent framework, leveraging readily available remote sensing-based raster data layers and a national ecological site point dataset developed from NASIS and SSURGO. Our results show that our modeling framework was effective at predicting ESG classes, with high model accuracies (70–79% cross-validated accuracy and 44–56% external-validated accuracy) and predicted maps aligned with our expectations of ESG spatial distributions. Through analyzing and comparing multiple machine learning models and covariate datasets, we determined that covariate selection was more important than model selection for modeling of ESGs. The lower importance of model type was not surprising given that we evaluated three of the strongest machine learning algorithms used for environmental modeling (Kuhn and Johnson, 2013). Since ecological site classification is based on the relationship between soils, vegetation and climate; it is logical that covariates typically used to approximate factors of soil formation, like those within our DSM dataset (e.g., climate, lithology, terrain attributes, NDVI), were the most important for predicting ESG classes. It is clear from the similarity in model performance across the different datasets that there is a high degree of correlation between covariates and shared contribution toward explaining model variance. The known interrelationships between biotic and abiotic factors
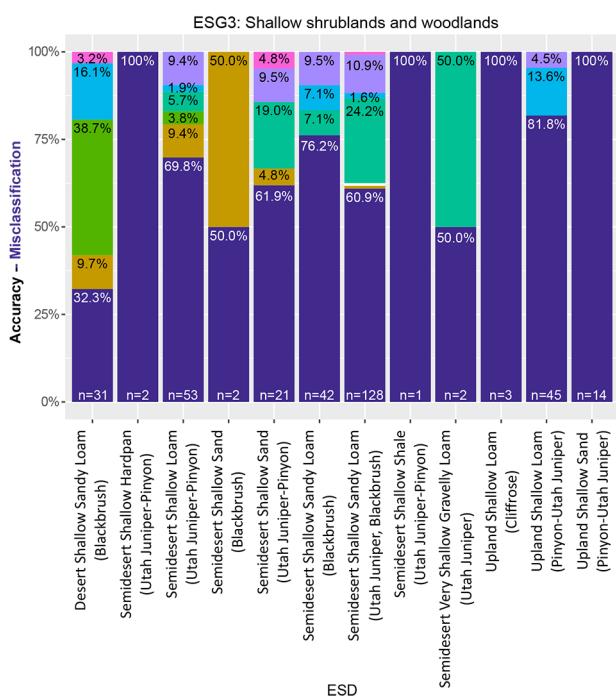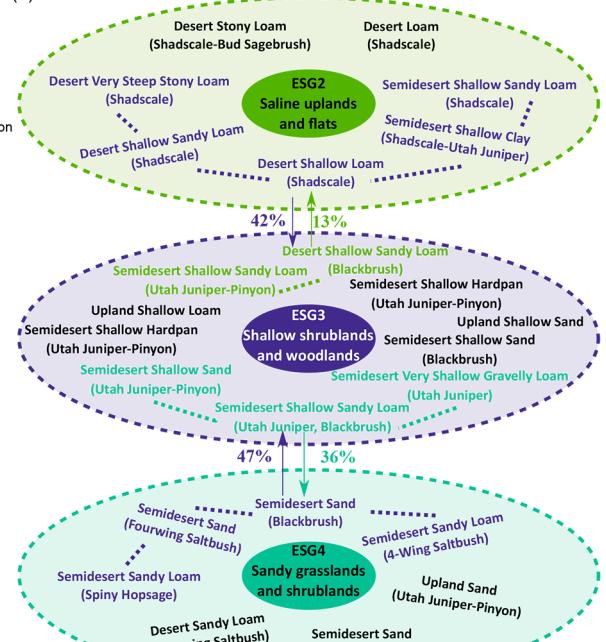
Fig. 9. Analysis of Colorado Plateau model misclassification for ecological site group (ESG) 3 (Shallow Shrublands and Woodlands): (a) relative percent accuracy/misclassification of ESG3 broken out by ecological site description (ESD) and the model predicted ESG (_n_ = number of ESG3 observations within each ESD class); (b) relationship of misclassifications between ESG3 and both ESG2 (Saline Uplands and Flats) and ESG4 (Sandy Grasslands and Shrublands), showing the ESDs with misclassified observations (bold colored text) and the relative percent of total misclassified points attributed to each group.

influencing soil and ecosystem development explains the strong similarity in model accuracy observed between our contrasting datasets (Behrens et al., 2014; Miller et al., 2015).

Ecological site concepts are developed at a regional scale (i.e., MLRA or Land Resource Unit) due to differences in soil and climatic controls on vegetation composition, distribution, and resilience. Differences in covariate importance between our study areas confirm the importance of these local-to-regional controls on ecosystem structure and function, and the need to develop models that characterize and quantify these differences. While both study sites displayed strong relationships to both biotic and abiotic covariates, underlying soil themes used often to distinguish ecological sites are strong predictors across all models, especially TAXOUSDA (SoilGrids250m USDA soil taxonomy suborders) (Fig. 3; Fig. 8c-f; Supplemental Fig. S2). USDA soil taxonomy suborders are representative of unique soil forming environments and encapsulate information on a wide range of soil properties (Soil Survey Staff, 2014). Suborders were also the fifth most important variable in the MLRA 35 RF model (Fig. 8b), and a suite of basic soil properties are heavily utilized in the MLRA 35 SVM model. All models also include at least two topographical covariates in the top ten with elevation, relative elevation, topographic wetness, and protection (similar to exposure) indices being commonly utilized in prediction decisions. Overall, the Chihuahuan Desert had low overall and classwise accuracies for the remotely sensed vegetation index datasets and largely favored soil and topographical variables.

The Colorado Plateau RF and XGB models showed a stronger response to NDVI (Fig. 8b-c), also shown in both the external validation (Fig. 3c) and in some of the individual ESG producer's accuracies (Supplemental Fig. S1; i.e., ESG1 and ESG3). The MAPVIP analysis further confirmed the importance of NDVI in the Colorado Plateau, where it was the top covariate in the RF model and ranked several times within the top 10 covariates for both RF and XGB models. The greater influence of NDVI in the Colorado Plateau may be because of the extreme difference in vegetation cover and greenness between ESGs in this region. For example, both the Saline Hills and Badlands as well as Outcrops and Slopes can have very low total foliar cover overall and thus would be easily distinguishable from other sites using NDVI. However, the NDVI data on its own had the lowest cross-validation accuracy (Fig. 3a), which may indicate the need to also incorporate soil and topographic parameters. The importance of soil and topography was also reported in recent work on the Colorado Plateau showing that 30-m maps of taxonomic soil particle size class and local topography can largely distinguish ecological sites (Nauman and Duniway, 2016) and similar finer scaled soil taxonomic data and topographic indices should be utilized in future work. The balance of soil, topography and vegetation index variable importance also reflect recommendations from recent work mapping ecological sites from hyper-temporal remotely sensed vegetation indices where over-reliance on spectral data can cause confusion between ecological states (e.g., disturbance) and ecological sites (Maynard and Karl, 2017).

Current spatial application of ESD by land managers and other users is dependent on SSURGO data to generate maps. We
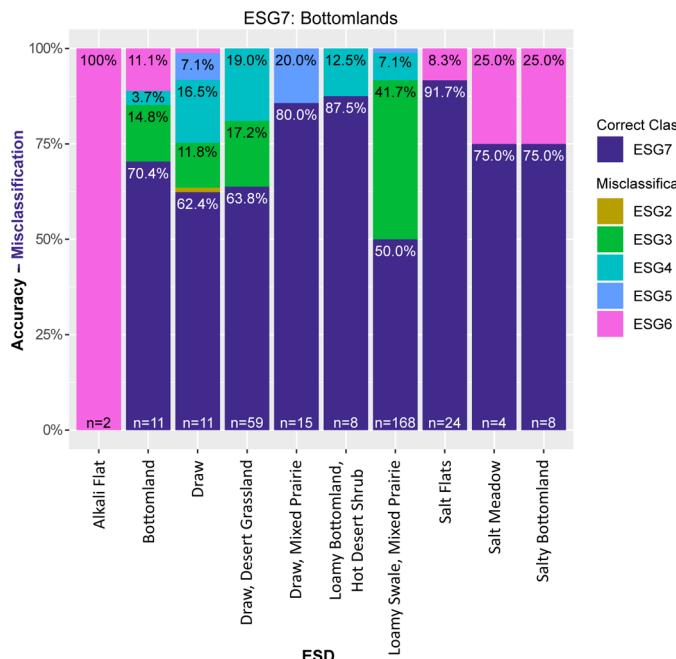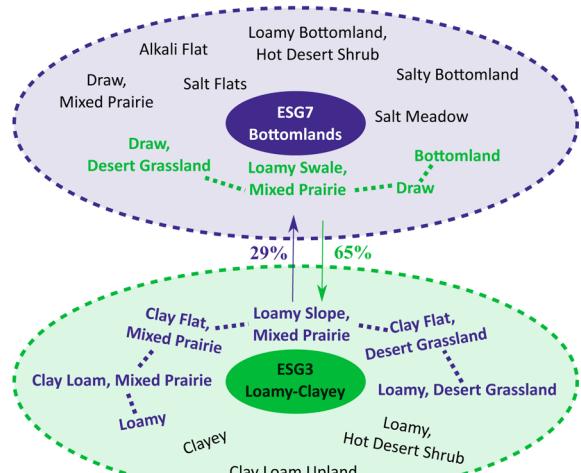
Fig. 10. Analysis of Chihuahuan Desert model misclassification for ecological site group (ESG) 7 (Bottomlands): (a) relative percent accuracy/ misclassification of ESG7 broken out by ecological site description (ESD) and the model predicted ESG (*n* = number of ESG7 observations within each ESD class); (b) relationship of misclassifications between ESG7 and ESG3 (Loamy-Clayey), showing the ESDs with misclassified observations (bold colored text) and the relative percent of total misclassified points attributed to each group.

have demonstrated that our modeling framework can produce predictive ESG maps that have a similar accuracy to SSURGO ESG maps, but in many cases higher precision for spatially rare ESGs. Our approach has the added benefit of predicting ESG distribution within areas that currently lack SSURGO ecological site data. Furthermore, through the ongoing refinement of our modeling framework we have the potential to achieve accuracies that surpass what is currently available from SSURGO.

Accounting for Sources of Model Error

Traditional soil and ecological class maps are based on the 'double crisp' model where groups are supposed to be crisply delineated in both thematic and geographic space (Burrough et al., 1997). One of the initial objectives of this study was to evaluate the different factors contributing to both thematic and spatial inaccuracies. We also differentiate a third source of inaccuracy that is not directly related to either thematic or spatial issues but originates from factors within the model itself (e.g., sample size distribution). Thus, ecological site prediction inaccuracy can be broadly attributed to three general categories; (i) thematic factors, (ii) spatial factors, and (iii) model-based factors.

Ecological site group concepts are created with the intent to minimize the number of groups while simultaneously minimizing within group variation and maximizing between group variations with respect to ecosystem properties and dynamics (e.g., resilience, response to disturbance; Bestelmeyer et al., 2016). Each ecological site group represents a modal concept that encompasses a range of variability in soil properties and vegetation composition. Devising discrete ecological groups, however, can be dif-

ficult when landscapes are characterized by broad gradients in biotic and abiotic properties that mediate ecosystem resilience and response to disturbance. ESG thematic error commonly occurs within these transitional environments when two or more classes exhibit high inter-class similarities with respect to the ecosystem properties and/or processes that define group concepts and the covariates that approximate them. Our analysis of ESG misclassification examined several examples where one or more ecological site within an ESG exhibited a higher similarity in characteristics to a different ESG resulting in high rates of misclassification for those ecological sites. These results suggest that in certain cases grouping concepts should be reevaluated; adjusting which ecological sites should be assigned within each ESG concept.

In spatial modeling, two aspects of scale can influence model accuracy: spatial resolution and spatial extent. All analysis in this study was conducted at a set spatial resolution of 250 m. This was chosen to balance the need for a spatial resolution high enough to detect the properties and processes relevant to ESG distribution across our study areas, as well as to manage the computational requirements needed for any future implementation of our approach at a regional or national scale. Continual advancements in computing power are making it feasible to implement spatial modeling frameworks such as ours, at even finer spatial scales (Ramcharan et al., 2018). In areas with high topographic complexity, our 250-m spatial resolution was too coarse to adequately characterize variability in ecosystem types. This was clearly demonstrated in the area surrounding our external-validation points in the Colorado Plateau study area, where our 250-m ENS-DSM predictions appear coarse relative to the SSURGO derived ESG map which

Fig. 11. Evaluation of model scaling effects in the Colorado Plateau study area. (a) SSURGO ecological site group (ESG) map of area surrounding external validation points; (b) digital soil mapping- ensemble (DSM-ENS) model results of area surrounding external validation points, (c, d) enlarged subregion from maps a and b illustrating the effect of the spatial resolution of covariate data in characterizing landscape variability.

was created at a finer spatial resolution (Fig. 11). Issues relating to spatial extent, include the uneven spatial distribution of training data. An uneven spatial distribution of training data can result in spatial extrapolation into areas not well defined in terms of the model covariate space. This was seen in both study areas where the prediction uncertainty is highest in areas where training data is limited (Fig. 7). Our results (Fig. 7 c-d) and a number of different digital soil mapping studies have shown that model prediction uncertainties are closely related to validation accuracy and can be 'calibrated' or rescaled by validation data to produce validation uncertainty maps that can more effectively direct new field sampling efforts aimed at reducing model uncertainty (Häring et al., 2012; Nauman and Thompson, 2014; Nauman et al., 2014; Ramcharan et al., 2018). Consequently, future modeling efforts should focus on translating model uncertainties into validated uncertainties, and field sampling efforts should target areas of mapped high uncertainty to help improve model accuracy.

The spatial distribution of points, as well as the spatial resolution of our data, both influences our ability to detect the boundaries between ESGs. While abrupt ESG boundaries can occur due to clearly observable physiographic features such as changes in geomorphology and lithology, gradual transitions or gradients between distinct soil forming environments and ecosystem types commonly occur (Burrough et al., 1997). There are several model-based factors that influence the ability to predict ESG classes, particularly in these transitional areas with high uncertainty. The first relates to deficiencies in our model covariate space for differentiating ESGs. From our example in the Colorado Plateau, we saw that the confusion between the ESG3 Desert Shallow Sandy Loam (Blackbrush) and ESG2 Desert Shallow Sandy Loam (Shadescale) was most likely due to an inability to detect differences in soil salinity. Examples like this illustrate how expert knowledge of ecosystem dynamics can help inform possible causes of model error. Machine learning algorithms allows for probabilistic outputs that can help to quantify the higher levels of uncertainty associated with classes that share similar regions of covariate space. This quantification of model uncertainty can help direct the acquisition of new covariate data

(e.g., soil salinity map) that has the potential to minimize the high uncertainty within these areas.

Through leveraging NASIS and SSURGO databases we were able to generate an extensive ESG point dataset at both study sites. However, low match rates (30% Colorado Plateau; 61% Chihuahuan Desert) between NASIS and SSURGO ecological site designations limited the number of possible points for model building and validation. These match rates will continue to increase as missing portions of SSURGO are surveyed and published, and with possible improvements to our current matching algorithm. Our sample size distribution was relatively balanced across our ESG classes at both sites; thus, we did not see a pattern of lower producer's accuracy in underrepresented classes. Increasing the number of observations, however, may help resolve some of the confusion between similar ESG classes, as a more complete characterization of the covariate space for each ESG is established. The utilization of field transect observations from NASIS in concert with final parameter attribution from SSURGO has great potential for modeling many different soil and ecological parameters. By combining the spatial density of field observations in NASIS (mostly soil taxonomic data) with the rich descriptive content in SSURGO, much more location-specific soils data can be leveraged into modeling efforts, essentially allowing for a de-constructing of SSURGO and remodeling with machine learning for spatial representation.

Future Research Efforts

While our ESG model accuracies were high and predicted maps aligned with our expectation of ESG distribution, there are still ways to improve on our modeling framework. One important improvement would be to improve the spatial resolution of our models and predicted surfaces. As our results demonstrate, in highly heterogeneous landscapes 250-m resolution predicted surfaces may fail to accurately delineate spatially rare or irregularly shaped ESGs (Fig. 11). Preliminary analysis of the Colorado Plateau has shown that improving the spatial resolution to 30 m resulted in a marked increase in both cross-validated and externally validated accuracies (unpubl. data, 2018). The topographic complexity of a landscape influences the spatial scale at which an ecosystem property or process can be detected and modeled (Maynard and Johnson, 2014). The Colorado Plateau study area exhibits a high degree of topographic complexity which has a strong influence on the spatial structure, potential composition and temporal dynamics of plant communities (Duniway et al., 2016). The ability to model these plant community dynamics is clearly dependent on identifying and adjusting the spatial scale of our covariate data to match the scale of the dominant soil-landscape processes. The ability to identify and adjust the spatial modeling scale will allow for the development and mapping of ecological site concepts that support and inform a variety of different land management objectives.

Machine learning models are effective at detecting patterns within complex datasets and provide a relatively automated approach to model fitting that does not require imposed relationships based on expert knowledge from soil scientists or ecologists

(Hengl et al., 2017, 2018). On the one hand machine learning approaches can remove or diminish many of the obstacles that have impeded the development and mapping of ecological sites. On the other hand, machine learning models have been criticized as 'black boxes' that run the risk of detecting relationships and predicting results that are not aligned with reality. The recent emergence of model agnostic interpretability methods offers new ways to shed light on the complexity of machine learning models, allowing greater interpretation and refinement of model covariates based on our knowledge of soil forming processes. Our use of a model agnostic variable importance measure allowed us to compare which covariates were most important across our machine learning models and thus infer dominant factors and processes influencing ESG differentiation.

We believe that machine learning approaches have tremendous potential but must be guided and refined by the wealth of expert knowledge that has been generated over the years. There are several areas where the combination of expert knowledge and data-mining and/or modeling techniques can provide significant benefit toward the development and refinement of ecological site mapping efforts. The first involves developing data-driven approaches to ecological site concept development and aggregation strategies for ecological site grouping. For the Colorado Plateau, Duniway et al. (2016) describe some of the preliminary data mining work they performed for evaluating ecological site variability and the development of ESG concepts. This information was then used to help guide the development of ESGs by a workgroup of scientists and land managers. Currently, large portions of the US still lack ecological site concepts or, where provisional concepts have been developed, lack a linkage to detailed soil mapping. Consequently, new approaches need to be developed to expedite the initial development of ecological site concepts, leveraging data-mining techniques and the wealth of existing soil and environmental data. The second area involves the mapping of ecological sites which is the focus of this study. We have already demonstrated and discussed many of the benefits of digital ESG mapping (e.g., extrapolation to unmapped areas, prediction of uncertainty). However, one significant benefit not yet addressed is the extensible nature of our modeling approach. While traditional soil and ecological maps encapsulate a tremendous wealth of expert knowledge, they largely exist as static products. This is due to the difficulties of transferring and extracting the necessary information needed to update maps as the resources change or to increase the spatial resolution as demands for more detailed land resources information rises. In contrast, machine learning methods provide an extensible framework that allows for continual updating and improvement as new data and analytical approaches become available. This will allow for the accuracy and precision of our mapping products to continue to improve, and for updated products to be generated at a much faster pace relative to traditional mapping approaches.

CONCLUSIONS

Spatial representations of ecological sites and ecological site groups using digital mapping techniques offers clear advantages over the current SSURGO soil-site correlation approach in terms of informing and directing land management actions across multiple spatial scales. In this study we presented a consistent, nationally scalable modeling framework for mapping ecological sites using a national point database, remotely sensed geospatial data layers, and machine learning algorithms. While our results are presented within the context of the current NRCS SSURGO spatial framework, our mapping framework is independent of existing mapping systems. Given the availability of geospatial covariate data and point data characterizing ecological site concepts, this approach has the potential of being applied anywhere in the world. This offers several advantages over existing mapping systems that produce static data products, most importantly the ability for models and data products to continually evolve and improve as new data sources and modeling techniques emerge. The ability to explicitly evaluate and adjust the spatial and thematic resolution of our modeled results will allow for the creation and delineation of ecological sites that provide more accurate representations of targeted processes. Results from this study demonstrate that our modeling framework can produce predictive ESG maps that have a similar accuracy to SSURGO ESG maps, but in many cases higher precision for spatially rare ESGs. Furthermore, digital mapping techniques provide the ability to predict ecological site distributions within areas currently unmapped in SSURGO. Static data products built largely on expert knowledge cannot meet the increasing demands for data products capable of addressing a range of new and evolving land management concerns at a variety of spatial scales. The development of open-source extensible modeling frameworks is not a replacement of expert knowledge but rather present new ways of encapsulating that knowledge into a system that is more flexible to user needs and capable of evolving as those needs change through time.

ACKNOWLEDGMENTS

SUPPLEMENTAL MATERIAL

Supplemental material is available with the online version of this article. The supplement document contains Table S1, Spatial raster covariate datasets use to model the spatial distribution of ESG classes; Fig. S1, Model class-wise producer's accuracies for the Colorado Plateau; and Fig. S2, Model class-wise producer's accuracies for the Chihuahuan Desert.

REFERENCES

Behrens, T., K. Schmidt, L. Ramirez-Lopez, J. Gallant, X. Zhu, and T. Scholten. 2014. Hyper-scale digital soil mapping and soil formation analysis. Geoderma 213:578–588. doi:10.1016/j.geoderma.2013.07.031

Bestelmeyer, B.T., J.C. Williamson, C.J. Talbot, G.W. Cates, M.C. Duniway, and J.R. Brown. 2016. Improving the effectiveness of ecological site descriptions: General state-and-transition models and the ecosystem dynamics interpretive tool (EDIT). Rangelands 38:329–335. doi:10.1016/j.rala.2016.10.001

Bischl, B., M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, and Z.M. Jones. 2016. mlr: Machine learning in R. J. Mach. Learn. Res. 17:1–5.

Bowker, M.A., M.E. Miller, R.T. Belote, and T. Belote. 2012. Assessment of rangeland ecosystem conditions, Salt Creek Watershed and Dugout Ranch. Southeastern Utah USGS Survey Open-File Rep. USGS, Washington, DC.

Breiman, L. 2001. Random forests. Mach. Learn. 45:5–32. doi:10.1023/A:1010933404324

Brown, J. 2010. Ecological sites: Their history, status, and future. Rangelands 32:5–8. doi:10.2111/Rangelands-D-10-00089.1

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

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

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

Casalicchio, G., C. Molnar, and B. Bischl. 2018. Visualizing the feature importance for black box models. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, New York. p. 655–670.

Caudle, D., F. Worth, and F. Service. 2013. Interagency ecological site handbook for rangelands. US Dep. of the Interior, Bureau of Land Management, Washington, DC.

Chambers, J.C., J.D. Maestas, D.A. Pyke, C.S. Boyd, M. Pellant, and A. Wuenschel. 2016. Using resilience and resistance concepts to manage persistent threats to sagebrush ecosystems and greater sage-grouse. Rangel. Ecol. Manage. 70:149-164.

Cleland, D.T., P.E. Avers, W.H. McNab, M.E. Jensen, R.G. Bailey, T. King, and W.E. Russell. 1997. National hierarchical framework of ecological units. In: M.S. Boyce, and A. Haney, editors, Ecosystem management applications for sustainable forest and wildlife resources. Yale Univ. Press, New Haven, CT. p. 181–200.

Coffman, J.M., B.T. Bestelmeyer, J.F. Kelly, T.F. Wright, and R.L. Schooley. 2014. Restoration practices have positive effects on breeding bird species of concern in the Chihuahuan Desert. Restor. Ecol. 22:336–344. doi:10.1111/rec.12081

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

Duniway, M.C., T.W. Nauman, J.K. Johanson, S. Green, M.E. Miller, J.C. Williamson, and B.T. Bestelmeyer. 2016. Generalizing ecological site concepts of the Colorado Plateau for landscape-level applications. Rangelands 38:342–349. doi:10.1016/j.rala.2016.10.010

Fisher, A., C. Rudin, and F. Dominici. 2018. Model class reliance: Variable importance measures for any machine learning model class, from the "Rashomon" perspective. arXiv preprint no. arXiv1801.01489.

Forman, G., and M. Scholz. 2010. Apples-to-apples in cross-validation studies. ACM SIGKDD Explor. Newsl. 12:49. doi:10.1145/1882471.1882479

GDAL/OGR Contributors. 2018. GDAL/OGR Geospatial data abstraction software library. Open Source Geospatial Foundation. http://gdal.org (verified 14 Apr. 2019).

Gesch, D.B., G.A. Evans, M.J. Oimoen, and S. Arundel. 2018. The National Elevation Dataset. In: American Society for Photogrammetry and Remote Sensing. Vol. 68. ASPRS, Earth Resources Observation and Science Center. p. 83–110.

Goldstein, A., A. Kapelner, J. Bleich, and E. Pitkin. 2015. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. J. Comput. Graph. Stat. 24:44–65. doi:10.1080/10618600.2014.907095

Greenwell, B.M., B.C. Boehmke, and A.J. McCarthy. 2018. A simple and effective model-based variable importance measure. arXiv preprint no. arXiv1805.04755.

Häring, T., E. Dietz, S. Osenstetter, T. Koschitzki, and B. Schröder. 2012. Spatial disaggregation of complex soil map units: A decision-tree based approach in Bavarian forest soils. Geoderma 185–186:37–47. doi:10.1016/j.geoderma.2012.04.001

Hartmann, J., and N. Moosdorf. 2012. The new global lithological map database GLiM: A representation of rock properties at the Earth surface. Geochem. Geophys. Geosys. 13. doi:10.1029/2012GC004370

Hengl, T., J. Mendes de Jesus, G.B.M. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, et al. 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12:e0169748.

Hengl, T., M.G. Walsh, J. Sanderman, I. Wheeler, S.P. Harrison, and I.C. Prentice.

2018. Global mapping of potential natural vegetation: An assessment of machine learning algorithms for estimating land potential. PeerJ 6:e5457. doi:10.7717/peerj.5457

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

Hijmans, R.J. 2019. raster: Geographic data analysis and modeling. R package version 2.8-19. R Core Development Team, Vienna.

Hill, P.L., R.P. Kucks, and D. Ravat. 2009. Aeromagnetic and aeroradiometric data for the conterminous United States and Alaska from the National Uranium Resource Evaluation (NURE) Program. US Dep. of Energy, Washington, DC.

Ireland, A.W., and P.J. Drohan. 2015. Rapid delineation of preliminary ecological sites applied to forested northern Appalachian landscapes. Soil Sci. Soc. Am. J. 79:185–192. doi:10.2136/sssaj2014.06.0271

Karl, J., and J. Herrick. 2010. Monitoring and assessment based on ecological sites. Rangelands 32:60–64. doi:10.2111/Rangelands-D-10-00082.1

Karl, J.W., and C.J. Talbot. 2016. The role of data and inference in the development and application of ecological site concepts and state-and-transition models. Rangelands 38:322–328. doi:10.1016/j.rala.2016.10.009

Kempen, B. 2011. Updating soil information with digital soil mapping. Ph.D. diss., Wageningen Univ., the Netherlands.

Knapp, C.N., and M.E. Fernandez-Gimenez. 2009. Understanding change: Integrating rancher knowledge into state-and-transition models. Rangeland Ecol. Manage. 62:510–521. doi:10.2111/08-176.1

Kuhn, M., and K. Johnson. 2013. Applied predictive modeling. Springer, the Netherlands. doi:10.1007/978-1-4614-6849-3

Maynard, J.J., and M.G. Johnson. 2014. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. Geoderma 230-231:29–40.

Maynard, J.J., and J.W. Karl. 2017. A hyper-temporal remote sensing protocol for high-resolution mapping of ecological sites. PLoS One 12. doi:10.1371/journal.pone.0175201

Maynard, J.J., J.W. Karl, and D.M. Browning. 2016. Effect of spatial image support in detecting long-term vegetation change from satellite time-series. Landsc. Ecol. 31:2045-2062. doi:10.1007/s10980-016-0381-y

McMahon, G., E.B. Wiken, and D.A. Gauthier. 2004. Toward a scientifically rigorous basis for developing mapped ecological regions. Environ. Manage. 34:S111.

Miller, M.E., R.T. Belote, M.A. Bowker, and S.L. Garman. 2011. Alternative states of a semiarid grassland ecosystem: Implications for ecosystem services. Ecosphere 2:1–18. doi:10.1890/ES11-00027.1

Miller, B.A., S. Koszinski, M. Wehrhan, and M. Sommer. 2015. Impact of multi-scale predictor selection for modeling soil properties. Geoderma 239-240:97-106.

Molnar, C., B. Bischl, and G. Casalicchio. 2018. iml: An R package for interpretable machine learning. JOSS 3:786.

Monger, H.C., and B.T. Bestelmeyer. 2006. The soil-geomorphic template and biotic change in arid and semi-arid ecosystems. J. Arid Environ. 65:207–218.

Nauman, T.W., and M.C. Duniway. 2016. The automated reference toolset (ART): An ecological potential matching algorithm based on soil particle size in the control section and neighborhood geomorphic variability. Soil Sci. Soc. Am. J. 80:3–6. doi:10.2136/sssaj2016.05.0151

Nauman, T.W., and J.A. Thompson. 2014. Semi-automated disaggregation of conventional soil maps using knowledge driven data mining and classification trees. Geoderma 213:385–399. doi:10.1016/j.geoderma.2013.08.024

Nauman, T.W., J.a. Thompson, and C. Rasmussen. 2014. Semi-automated disaggregation of a conventional soil map using knowledge driven data mining and random forests in the Sonoran Desert, USA. Photogramm. Eng. Remote

Sens. 80:353–366. doi:10.14358/PERS.80.4.353

Nauman, T.W., J.A. Thompson, J. Teets, T. Dilliplane, J.W. Bell, S.J. Connolly, H.J. Liebermann, and K. Yoast. 2015. Pedoecological modeling to guide forest restoration using ecological site descriptions. Soil Sci. Soc. Am. J. 79:1406. doi:10.2136/sssaj2015.02.0062

Niculescu-Mizil, A., and R. Caruana. 2005. Predicting good probabilities with supervised learning. In: Proc. 22nd Int. Conf. Mach. Learn. Bohn, Germany. p. 625–632. doi:10.1145/1102351.1102430

Oliveira, S., F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J.M.C. Pereira. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest. For. Ecol. Manage. 275:117–129. doi:10.1016/j.foreco.2012.03.003

Pelletier, J.D., P.D. Broxton, P. Hazenberg, P.A.X. Zeng, G.N. Troch, Z.C. Williams, M.A. Brunke, and D. Gochis. 2016. Global 1-km gridded thickness of soil, regolith, and sedimentary deposit layers. ORNL DAAC, Oak Ridge, TN.

Poitras, T.B., M.L. Villarreal, E.K. Waller, T.W. Nauman, M.E. Miller, and M.C. Duniway. 2018. Identifying optimal remotely-sensed variables for ecosystem monitoring in Colorado Plateau drylands. J. Arid Environ. 153:76–87. doi:10.1016/j.jaridenv.2017.12.008

Pontius, R.G., M. Millones, J. Pontius, R. Gilmore, M. Millones, R.G. Pontius, and M. Millones. 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. Int. J. Remote Sens. 32:4407–4429. doi:10.1080/01431161.2011.552923

R Development Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ramcharan, A., T. Hengl, T. Nauman, C. Brungard, S. Waltman, S. Wills, and J. Thompson. 2018. Soil property and class maps of the conterminous US at 100 meter spatial resolution based on a compilation of national soil point observations and machine learning. Soil Sci. Soc. Am. J. 82:186–201. doi:10.2136/sssaj2017.04.0122

Richter, J., J. Rahnenführer, and M. Lang. 2017. mlrHyperopt: Effortless and collaborative hyperparameter optimization experiments. In: The R user conference, useR! 2017 July 4-7 2017. Brussels, Belgium. p.78.

Romanski, P., and L. Kotthoff. 2018. FSelector: Selecting attributes. R package version 0.31. R Core Development Team, Vienna.

Salley, S.W., H. Curtis Monger, and J.R. Brown. 2016a. Completing the land resource hierarchy. Rangelands 38:313–317. doi:10.1016/j.rala.2016.10.003

Salley, S.W., C.J. Talbot, and J.R. Brown. 2016b. The Natural Resources Conservation Service land resource hierarchy and ecological sites. Soil Sci. Soc. Am. J. 80:1. doi:10.2136/sssaj2015.05.0305

Shannon, C.E. 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27:379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Soil Survey Staff. 2014. Keys to soil taxonomy. USDA-NRCS, Washington, DC.

Steele, C.M., B.T. Bestelmeyer, L.M. Burkett, P.L. Smith, and S. Yanoff. 2012. Spatially explicit representation of state-and-transition models. Rangeland Ecol. Manag. 65:213–222. doi:10.2111/REM-D-11-00047.1

Warrens, M.J. 2015. Properties of the quantity disagreement and the allocation disagreement. Int. J. Remote Sens. 36:1439–1446. doi:10.1080/01431161.2015.1011794

Wei, R., and J. Wang. 2018. multiROC: Calculating and visualizing ROC and PR curves across multi-class classifications. R package version 1.1.1. R Core Development Team, Vienna.

Winthers, E., D. Fallon, J. Haglund, T. DeMeo, G. Nowacki, D. Tart, M. Ferwerda, G. Robertson, A. Gallegos, and A. Rorick. 2005. Terrestrial ecological unit inventory technical guide. Gen. Tech. Rep. WO-GTR-68. USDA Forest Service, Washington, DC.