

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

USGS Northern Prairie Wildlife Research Center

US Geological Survey

2008

Comparing Local vs. Global Visible and Near-Infrared (VisNIR) Diffuse Reflectance Spectroscopy (DRS) Calibrations for the Prediction of Soil Clay, Organic C and Inorganic C

Joel B. Sankey
Idaho State University

David J. Brown
Washington State University, Pullman

Melisa L. Bernard
USGS Northern Prairie Wildlife Research Center

Rick L. Lawrence
Montana State University, rickl@montana.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/usgsnpwrc>

 Part of the [Other International and Area Studies Commons](#)

Sankey, Joel B.; Brown, David J.; Bernard, Melisa L.; and Lawrence, Rick L., "Comparing Local vs. Global Visible and Near-Infrared (VisNIR) Diffuse Reflectance Spectroscopy (DRS) Calibrations for the Prediction of Soil Clay, Organic C and Inorganic C" (2008). *USGS Northern Prairie Wildlife Research Center*. 27. <https://digitalcommons.unl.edu/usgsnpwrc/27>

This Article is brought to you for free and open access by the US Geological Survey at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in USGS Northern Prairie Wildlife Research Center by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C

Joel B. Sankey^{a,*}, David J. Brown^{b,1}, Melisa L. Bernard^{c,2}, Rick L. Lawrence^{d,3}

^a 322 E. Front St. Suite 240, Idaho State University, Boise, ID, 83702-7359, United States

^b PO Box 646420, Washington State University, Pullman, WA 99164-6420, United States

^c 8711 37th St. SE, Northern Prairie Wildlife Research Center, Jamestown, ND 58401, United States

^d PO Box 173490, Montana State University, Bozeman, MT 59717-3490, United States

ARTICLE INFO

Article history:

Received 30 January 2008

Received in revised form 26 August 2008

Accepted 23 September 2008

Available online 1 November 2008

Keywords:

Diffuse reflectance spectroscopy

VisNIR

Soil organic carbon

Proximal soil sensing

Boosted regression trees

ABSTRACT

Local, field-scale, VisNIR-DRS soil calibrations generally yield the most accurate predictions but require a substantial number of local calibration samples at every application site. Global to regional calibrations are more economically efficient, but don't provide sufficient accuracy for many applications. In this study, we quantified the value of augmenting a large global spectral library with relatively few local calibration samples for VisNIR-DRS predictions of soil clay content (clay), organic carbon content (SOC), and inorganic carbon content (IC). VisNIR models were constructed with boosted regression trees employing global, local+global, and local spectral data, using local samples from two low-relief, sedimentary bedrock controlled, semiarid grassland sites, and one granitic, montane, subalpine forest site, in Montana, USA. The local+global calibration yielded the most accurate SOC predictions for all three sites [Standard Error of Prediction (SEP)= 3.8, 6.7, and 26.2 g kg⁻¹]. This was similarly true for clay (SEP=95.3 and 102.5 g kg⁻¹) and IC (SEP=5.5 and 6.0 g kg⁻¹) predictions at the two semiarid grassland sites. A purely local calibration produced the best validation results for soil clay content at the subalpine forest site (SEP=49.2 g kg⁻¹), which also had the largest number of local calibration samples (N=210). Using only samples from calcareous soils in the global spectral library combined with local samples produced the best SOC and IC results at the more arid of the two semiarid sites. Global samples alone never achieved more accurate predictions than the best local+global calibrations. For the temperate soils used in this study, the augmentation of a large global spectral library with relatively few local samples generally improved the prediction of soil clay, SOC, and IC relative to global or local samples alone.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) has been employed to predict soil properties for specific localities with samples and spectral libraries representing local, regional, and global soil variability (Brown et al., 2006). However, there have been few attempts to explicitly merge spectral data from local, regional, and global scales to improve prediction of soil properties for local areas. Local libraries refer to field-scale data sets. Regional libraries are from a greater geographic extent than local libraries, representing coherent physio-climatic zones. Global libraries include the world's major soil taxa, with samples from multiple continents (Brown et al., 2006). Brown, (2007)

successfully combined tropical soil samples from Uganda with a global soil-spectral library to predict clay content, soil organic carbon, and clay mineralogy for Ugandan soils. To our knowledge, local soil-spectral data sets have not been combined with global VisNIR spectral libraries to estimate soil characteristics in less weathered, and therefore potentially more spectrally complex, mid-latitude, temperate soil environments.

For VisNIR-DRS (0.35–2.5 μm electromagnetic range⁴), soil organic matter, carbonates, iron oxides, and clay minerals have distinguishable spectral signatures (Hunt, 1989; Henderson et al., 1992; Clark, 1999). These spectral signatures are largely due to overtones and combinations of mid-infrared absorption features associated with C–O, C–H, O–H, metal–OH, and H₂O bonds (Hunt, 1989; Clark, 1999).

* Corresponding author. Tel.: +1 208 220 6571.

E-mail addresses: sankjoel@isu.edu (J.B. Sankey), david_brown@wsu.edu (D.J. Brown), mborino@usgs.gov (M.L. Bernard), rickl@montana.edu (R.L. Lawrence).

¹ Tel.: +1 509 335 1859.

² Tel.: +1 701 253 5538.

³ Tel.: +1 406 994 5409.

⁴ Established terms for spectral ranges include: (1) visible (Vis), 0.4–0.7 μm and (2) near-infrared (NIR), 0.7–3.0 μm. However, the remote sensing has adopted terms unique to that community, including: (1) visible/near-infrared (VNIR), 0.4–1.0 μm and (2) short-wave infrared (SWIR), 1.0–2.5 μm (Clark, 1999). To reduce confusion, we employ VisNIR (a combination of visible and NIR) to capture the spectral range commonly used in soil spectroscopy, 0.35–2.5 μm.

Iron-bearing minerals (primary and secondary) as well as soil organic matter have electronic absorptions in the visible and short-wave-length near-infrared range (400–1200 nm), giving rise to distinctive soil coloring (Henderson et al., 1992; Clark, 1999). Many soil properties related to these fundamental VisNIR absorptions (e.g. cation exchange capacity, organic C, inorganic C, and exchangeable cations) have been successfully modeled using VisNIR-DRS (Dunn et al., 2002; McCarty et al., 2002; Shepherd and Walsh, 2002; Brown et al., 2005, 2006).

Estimates of clay content (e.g. Chang et al., 2001; Shepherd and Walsh, 2002; Sorensen and Dalsgaard, 2005; Brown et al., 2006), soil organic carbon (SOC) (e.g. McCarty et al., 2002; Shepherd and Walsh, 2002; Brown et al., 2005, 2006), and soil inorganic carbon (IC) (e.g. McCarty et al., 2002; Brown et al., 2005) – based on soil spectra from specific local, regional, or global data sets – have been validated with a substantial range of prediction errors. A comparison of results highlights a general trend of decreasing accuracy for local to regional to global clay, SOC, and IC calibrations. A cross-validation Root Mean Square Deviation (RMSD) of 19.1 g kg⁻¹ clay, for example, was reported for a local calibration study (Viscarra Rossel et al., 2006), cross-validation and validation RMSD's from 20 g kg⁻¹ to 75 g kg⁻¹ clay have been reported based on regional data sets (Chang et al., 2001; Shepherd and Walsh, 2002; Sorensen and Dalsgaard, 2005), and a validation RMSD of 95 g kg⁻¹ clay was obtained using a global spectral library (Brown et al., 2006). For SOC, Viscarra Rossel et al. (2006) obtained a validation RMSD of 1.8 g kg⁻¹ for a local field-scale calibration, as compared to published regional and global validation RMSD's of 3.1 g kg⁻¹ (Shepherd and Walsh, 2002), 4.3 g kg⁻¹ (Islam et al., 2003), and 9.0 g kg⁻¹ (Brown et al., 2006), respectively.

In between local and regional calibrations, Brown et al. (2005) and McCarty et al. (2002) have reported validation SOC RMSD's of 1.26 and

5.5 g kg⁻¹, respectively, for what were essentially multiple local calibrations combined into one regional result. In both of these studies, results were worse for geographically independent validation of regional calibrations: 1.0 to 3.5 g kg⁻¹ for north central Montana 6-fold cross-validation (Brown et al., 2005) and 7.9 g kg⁻¹ for a single site in Nebraska using 13 other US Great Plains sites for calibration (McCarty et al., 2002). Similarly, Brown et al. (2005) reported a soil IC validation RMSD of 1.6 g kg⁻¹ for six combined local calibrations but geographically independent validation RMSD's of 1.5 to 5.0 g kg⁻¹ for six sites in north central Montana. McCarty et al. (2002) obtained a combined IC validation RMSD of 3.1 g kg⁻¹ for 14 Great Plains sites, with 4.4 g kg⁻¹ for the independent Nebraska validation. By comparison, Brown et al. (2006) reported a soil IC validation RMSD of 6.2 g kg⁻¹ for their global spectral library.

A number of researchers have promoted the construction of large regional and/or global spectral libraries of soil samples to increase the efficiency of DRS characterization (Shepherd and Walsh, 2002; McCarty et al., 2002; Brown, 2007). Questions remain, however, as to the utility of global libraries, and the relative importance of local spectral samples in the calibration of DRS models based on global spectral libraries, particularly in mid-latitude, temperate environments.

In this study, we employed a global soil-spectral library and local calibration samples from three study sites to model clay content, SOC, and IC from first derivative VisNIR spectral reflectance. Our main objective was to determine whether a larger global soil-spectral library augmented with local calibration samples yields improved soil property predictions relative to purely global or local calibrations. A secondary objective was to evaluate (i) selection of a particularly relevant fraction of the global library (only those profiles with detectable carbonates) or (ii) reducing the “weight” of the global samples relative to local calibration samples to improve soil property predictions.

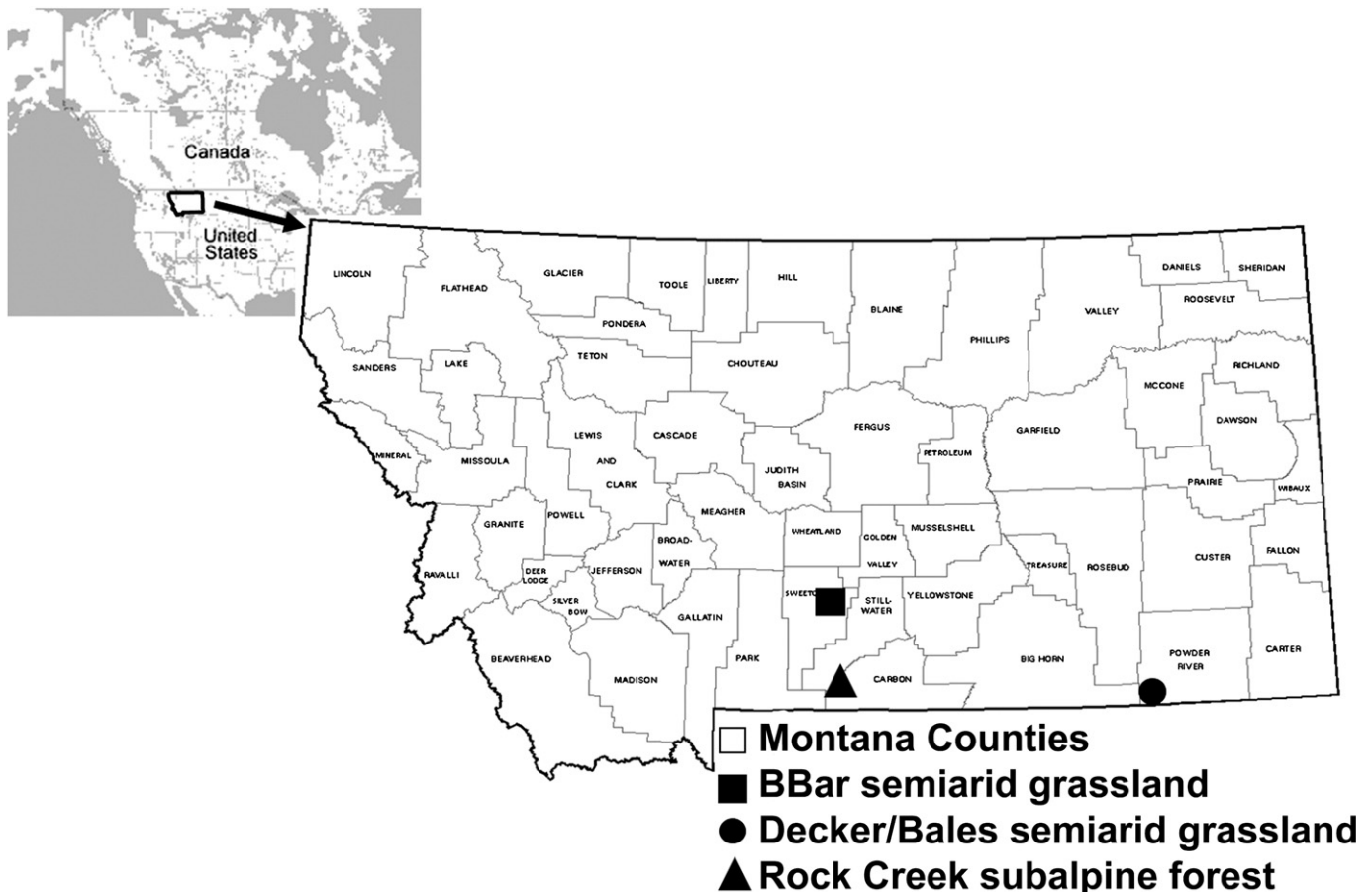


Fig. 1. Study site locations in Montana, USA.

2. Study sites

Samples were collected at three Montana study sites (Figs. 1 and 2): the Rock Creek watershed (subalpine forest), and the Decker/Bales and BBar ranches (semiarid grasslands). The Rock Creek watershed lies within the Custer National Forest in south-central Montana, located in the Beartooth Mountains approximately 8 km southwest of Red Lodge, Montana in Carbon County. The area is within the Greater Yellowstone Ecosystem, with a portion of the watershed within the Absaroka–Beartooth Wilderness. The terrain is rugged with elevations ranging from 1600 to 3800 m with 0 to 100% slopes, rock outcrops and talus slopes. The area is underlain by Archean granitic gneiss and Pleistocene glacial deposits in the valley bottoms (VanGosen et al., 2000). Soils include: Typic and Lithic subgroups of Eutrocryepts, Dystrrocryepts, and Cryochrepts on mountain slopes and plateaus; Typic and Lithic subgroups of Haplocryalfs on forested hillslopes, and Typic Cryofluvents in valley bottoms. Forest vegetation of lodgepole pine (*Pinus contorta* Doug.), Douglas-fir (*Pseudotsuga menziesii* Mirb.), sub-alpine fir (*Abies lasiocarpa* Hook.), Engelmann spruce (*Picea engelmannii* Parry.), and whitebark pine (*Pinus albicaulis* Engelm.) covers most of the landscape. Mean annual precipitation is 50–60 cm at lower elevations, increases to 130 cm at higher elevations, and snow remains at elevations over 2400 m until mid to late summer (Forest Service Staff, 1986).

The BBar and Decker/Bales ranches are both located in Montana's non-glaciated plains. BBar is in the westernmost extent of the plains, in northern Sweet Grass County. Decker/Bales is in the eastern plains, in southwestern Powder River County, in southeastern Montana. The landscape at both sites is characterized by dissected sedimentary layers that form a low relief, fluvially incised landscape, though the hills are more rolling at BBar with isolated surfaces of alluvium and outwash from the Crazy Mountains. Elevation ranges from 1070 m to 1290 m at Decker/Bales, and 1340 m to 1450 m at BBar. Vegetation at both ranches includes grassland communities of western wheatgrass (*Agropyron smithii* Rydb.), needle and thread (*Stipa comata* Trin. & Rupr.), and blue grama (*Bouteloua gracilis* Willd. ex Kunth) (Montagne et al., 1982). Soils at Decker/Bales include loamy, calcareous Typic Ustorthents formed in siltstones, clayey, calcareous Typic Ustorthents formed in shales, fine to coarse-loamy Typic Haplustalfs formed in slope alluvium, loamy-skeletal Lithic Haplustalfs formed in scoria beds, and fine Typic Natrustalfs often associated with prairie dog communities (Veseth and Montagne, 1980; Montagne et al., 1982). Soils found at BBar range from fine Typic Argiustolls on backslopes, footslopes, and toeslopes, to loamy-skeletal Lithic Ustorthents on summit and shoulder positions, as well as fine Typic Natrustalfs on toeslopes and valley floor positions, and fine and fine-loamy Typic Torrifluvents in drainageways (Veseth and Montagne, 1980; Montagne et al., 1982). The Decker/Bales ranch receives approximately 30 cm of mean annual precipitation, with soil temperature and moisture regimes at the Mesic/Frigid and Ustic/Aridic boundaries, respectively (Soil Survey Staff, 1971). BBar is slightly wetter and colder with 35 cm of mean annual precipitation, and classified as Frigid and Ustic (Soil Survey Staff, 2004).

3. Methods

3.1. Soil sampling and laboratory analysis

At the semiarid grassland sites, we collected 106 samples from 37 profiles in May, 2005 as part of a soil water modeling study (Sankey,

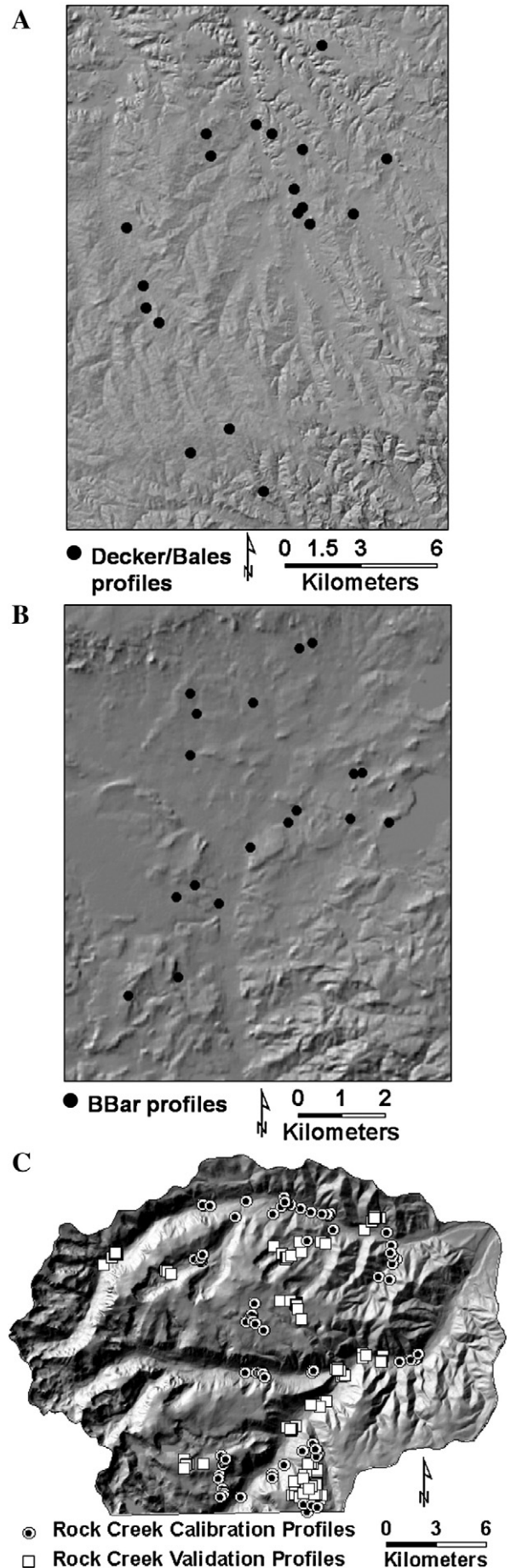


Fig. 2. Sample profile locations [(A) Decker/Bales (semiarid grassland), (B) BBar (semiarid grassland), (C) Rock Creek (subalpine forest)]. Locations are presented on a hillshade (illuminated from NW, with 1x vertical exaggeration) created from a digital elevation model for each site. The hillshades provide a qualitative representation of topographic variability at the sites.

2005). Profile locations (Fig. 2) were generated for both ranches with a stratified random approach that included at least one location per soil survey map unit. Soil profiles were sampled in 10-cm depth increments with a 5.59 cm diameter bucket auger, and samples from the 0–10 cm, 30–40 cm, and 70–80 cm depths were selected for lab characterization. Samples from the semi-arid grassland sites were oven-dried after collection (Soil Survey Staff, 1996).

For the subalpine forest site, we collected 444 samples at 0–10 and 10–20 cm depth intervals from 225 profiles during the summers of 2004 and 2005 as part of a U.S. Forest Service terrestrial ecological unit inventory (TEUI) and soil-landscape modeling study (Bernard, in preparation). Sample profiles were located using a Y-shaped cluster design, with 30 clusters, and up to 13 profiles per cluster (Fig. 2). The cluster centers were generated with a stratified random approach that included at least one cluster per local landscape unit. Samples from the subalpine forest site were air-dried after collection (Soil Survey Staff, 1996).

The global soil-spectral library with 4184 total samples was constructed by scanning previously characterized samples held in the US Natural Resources Conservation Services (NRCS) Soil Survey Center archives in Lincoln, NE, USA (Brown et al., 2006). The global library did not include any samples from the three local sites used in this study. The global library was constructed with spectra from air-dried samples, however, both air and oven-dried samples were scanned in the construction of the global spectral library and no difference in spectral characteristics was detected between the drying methods (Brown et al., 2006). No more than one sample per pedon was included in the global library, and samples were selected to maximize soil geographic and property diversity.

After drying, the fine earth fractions (<2 mm) of all samples were separated from coarse fragments (>2 mm) by hand grinding and sieving. These fractions were then scanned with an ASD “Fieldspec Pro FR” spectroradiometer (Analytical Spectral Devices, Boulder, CO) using the same procedure and instrument previously described for construction of the global soil-spectral library (Brown et al., 2006). The spectroradiometer had a spectral range of 350–2500 nm, a 2-nm sampling resolution, and a spectral resolution of 3 nm and 10 nm from 350–1000 nm and 1000–2500 nm, respectively. Soil samples were placed in an optical quality, borosilicate petri dish (Duraplan®). A Spectralon® panel was used for white referencing. The Fieldspec Pro FR illuminates samples and the white referencing panel with a white light from below and scans them from below. Ten internally averaged scans were recorded for each interrogation. Two scans were collected for each sample with a 90 degree rotation between scans. Reflectance and first derivatives of replicate scan spectra were compared and

samples were rescanned when possible errors were detected. For each sample, replicate spectra were averaged and smoothed, with 1st derivative values of reflectance extracted at 10 nm intervals from 360 to 2490 nm (Brown et al., 2006).

We measured dry combustion total carbon (TC) for fine-ground samples from all three local study sites using a LECO C/N/S 2000 analyzer (LECO Corporation, St Joseph, MI, USA). Inorganic carbon (IC) was determined using the modified pressure calcimeter method (Sherrod, 2002) for the semiarid grassland sites, with IC assumed to be absent from the subalpine forest site. Soil organic carbon (SOC) was calculated as the difference between TC and IC. Particle size analysis was performed by the pipette method with NaOCl used for SOM removal (Gee and Bauder, 1986; Soil Survey Staff, 1996). For the global soil-spectral library (Brown et al., 2006): (i) clay content was determined via the pipette method ($N=4184$); (ii) IC was determined by HCl treatment and manometer ($N=4184$) (Soil Survey Staff, 1996); (iii) for a subset of samples SOC was determined using the modified Walkley–Black method ($N=3794$); and (iv) for a subset of samples TC was determined using dry combustion with SOC computed as TC-IC ($N=1548$).

We estimated the standard error of the laboratory (SEL) for each method used to characterize soil samples from the local study sites using the following equation:

$$SEL = \sqrt{\frac{\sum(Y_1 - Y_2)^2}{2N}}$$

where (Y_1 , Y_2) are duplicate reference analysis and N is the number of replicate pairs. Forty two samples were randomly selected from the subalpine forest site for replicate analysis of particle size characterization. Twenty samples were randomly selected from the subalpine forest site for replicate analysis of SOC characterization. Five samples were randomly selected from the semiarid grassland sites for replicate analysis of IC characterization.

3.2. Calibration and validation schemes

Models were constructed for all sites with several calibration data combinations (Tables 1 and 2, and Fig. 3). For the semiarid grassland sites these were: 1) only global (NRCS) samples, 2) only global “calcareous” samples (samples were not necessarily calcareous, but were taken from *pedons* with detected calcium carbonates), 3) both

Table 1
Characteristics of spectral data sets, validation, calibration, and modeling schemes employed in this study

Soil datasets	Calibration and validation subsets	Modeling schemes
Subalpine forest (Rock Creek) • Clay ($N=444$) • SOC ($N=225$)	• 1/2 of sampling clusters randomly withheld for independent validation • Calibration $N=210$ (clay) and 107 (SOC) • Validation $N=234$ (clay) and 118 (SOC)	• Local only • Local+global • Global only
Semiarid grassland BBar • clay, SOC, IC ($N=52$) Decker/Bales • Clay, SOC, IC ($N=54$)	• Profiles randomly assigned to 10 strata • 5 strata per site • 10-fold cross-validation (no independent validation)	• Local only • Local+global • Global only • Local+calcareous global • Calcareous global
Global-All • Clay ($N=4184$) • IC ($N=4184$) • SOC Walkley–Black ($N=3794$) • SOC dry combust. ($N=1548$)	• SOC Walkley–Black for subalpine forest • SOC dry combustion for semi-arid grasslands	
Global-CaCO ₃ • Clay ($N=1548$) • SOC dry combust. ($N=554$) • IC ($N=1,548,377$)		

Table 2
Summary of soil properties for data sets used in this study

Data set	Soil property	N	Lab method	Min.	Max.	Median	Mean
Global-all							
	Clay (g kg ⁻¹)	4184	Pipette	1.0	930.0	402.0	256.6
	SOC (g kg ⁻¹)	3794	Walkley–Black	0.0	536.8	4.7	12.1
	SOC (g kg ⁻¹)	1548	Dry combustion	0.0	30.6	7.0	18.1
	IC (g kg ⁻¹)	4184	HCl & manometer	0.0	128.8	0.0	5.8
Global-CaCO₃							
	Clay (g kg ⁻¹)	1548	Pipette	4.0	930.0	241.5	269.7
	SOC (g kg ⁻¹)	554	Dry combustion	0.0	182.0	5.6	12.4
	IC (g kg ⁻¹)	1548	HCl & manometer	0.0	128.0	6.4	15.6
Local-subalpine forest (Rock Creek)							
	Clay (g kg ⁻¹)	444	Pipette	0.0	411.1	111.5	121.4
	SOC (g kg ⁻¹)	225	Dry combustion	3.4	335.0	34.3	53.9
Local-semiarid grassland (BBar)							
	Clay (g kg ⁻¹)	52	Pipette	66.4	585.5	295.4	316.7
	SOC (g kg ⁻¹)	52	Dry combustion	0.0	107.6	4.4	10.8
	IC (g kg ⁻¹)	52	Mod. Pres. calcimeter	0.0	29.5	3.9	5.1
Local-semiarid grassland (Decker/Bales)							
	Clay (g kg ⁻¹)	54	Pipette	92.3	647.3	286.8	319.5
	SOC (g kg ⁻¹)	54	Dry combustion	0.0	64.9	7.9	11.6
	IC (g kg ⁻¹)	54	Mod. pres. calcimeter	0.0	47.0	6.4	8.3

global and local samples, 4) both global calcareous samples and local samples, and 5) only local samples. Modeling for the subalpine forest site included calibration data combinations 1, 3, and 5 above and did not employ the calcareous subset of NRCS profiles.

The local semiarid grassland samples were randomly stratified by profile into 5 strata per site (10 strata total). Using these strata, we then employed 10-fold cross-validation modeling and computed cross-validation results for the two sites separately. Independent validation was not performed for the semiarid grassland sites because of the relatively small number of samples from each site. At the subalpine forest site, sampling clusters were stratified by landscape position (upper, middle and lower) then randomly assigned in equal proportions by strata to calibration and validation pools. Validation samples were

withheld from modeling and used for independent validation at the subalpine forest site.

Different subsets of the global dataset were used in carbon calibration models for the semiarid grassland vs. sub-alpine forest sites. Soil organic carbon calibration included just the samples characterized by dry combustion for the semiarid grassland sites, and just those characterized by Walkley–Black for the subalpine forest site. All global IC samples were used for semiarid grassland calibrations while IC was not modeled for the subalpine forest soils.

3.3. VisNIR-DRS modeling

Prior to modeling, we tested the pair-wise correlation of soil attributes at each study site. Soil attributes were not strongly correlated (all Pearson’s correlation coefficients < 0.45). TreeNet® software (Salford Systems, San Diego, CA, USA) was used to model clay content (g kg⁻¹), SOC (g kg⁻¹), and IC (g kg⁻¹) with boosted regression tree (BRT) models. Boosted regression tree model development with TreeNet begins with a single regression tree, from which multiple regression trees are iteratively fit in an attempt to minimize residuals. A weighting scheme is employed at each iteration which gives highest weight to observations with the largest residuals in the previous iteration. TreeNet® constructs BRT models by repeating this process with random subsamples of the training data and then averaging the resultant predictions for a final output. Boosted regression tree modeling has been shown to be an appropriate and useful tool in the prediction of soil characteristics using a global spectral library of soil samples (Brown et al., 2006). For the semiarid grassland sites a maximum of 1000 trees was specified, with minimum and maximum number of nodes per tree, 10 and 12, respectively – parameters arrived at heuristically in a previous study (Brown et al., 2006). The parameters arrived at heuristically and used for the subalpine forest site were 600 trees and 6 nodes per tree for developing clay content BRT models, and 1000 trees and 6 nodes per tree for predicting SOC. Models were developed for three random realizations of all calibration data combinations and the prediction results were averaged.

We used partial least squares regression (PLSR) in addition to BRT models, for purely local calibrations at all three sites. Partial least squares models were developed using Unscrambler® 8.0.5 software package (CAMO Technologies, Inc., Woodbridge, NJ, USA). It was anticipated that

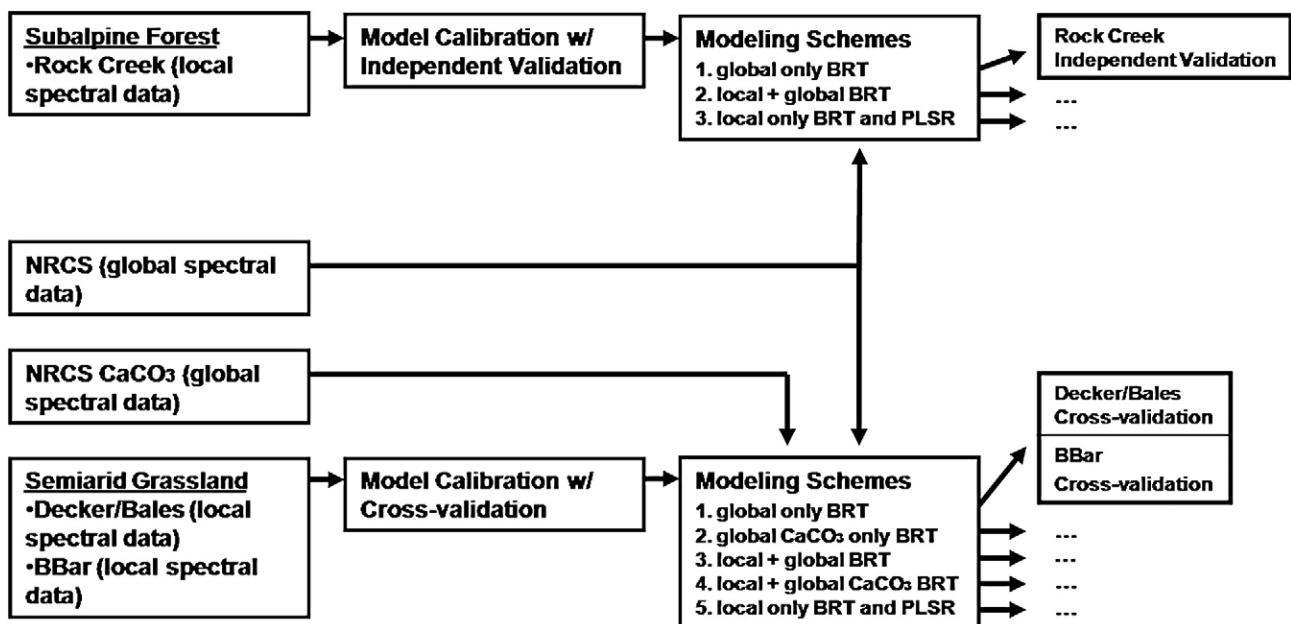


Fig. 3. Overview of general modeling approach.

Table 3
Validation results for clay content (g kg⁻¹) predictions

	Global ^a & local ^b	Global-CaCO ₃ ^c & local	Global	Global-CaCO ₃	Local
<i>Bbar (semiarid grassland)</i>					
N (Validation)	52	52	52	52	52
SEP (g kg ⁻¹ clay)	95.4	95.3	96.2	103.0	108.5
RPD	1.10	1.10	1.09	1.02	0.97
SB (%)	0	0	1	11	6
NU (%)	67	87	74	74	92
LC (%)	33	13	26	15	2
r ²	0.21	0.20	0.19	0.16	0.02
<i>Decker/Bales (semiarid grassland)</i>					
N (Validation)	54	54	54	54	54
SEP (g kg ⁻¹ clay)	102.5	107.0	123.1	110.4	137.6
RPD	1.36	1.30	1.13	1.26	1.01
SB (%)	4	2	36	3	4
NU (%)	40	52	29	51	93
LC (%)	55	46	35	45	3
r ²	0.49	0.43	0.51	0.40	0.09
<i>Subalpine forest (Rock Creek)</i>					
N (Validation)	234	–	234	–	234
SEP (g kg ⁻¹ clay)	56.3	–	65.1	–	49.2
RPD	1.2	–	1.1	–	1.4
SB (%)	0	–	0	–	0
NU (%)	41	–	40	–	70
LC (%)	59	–	60	–	30
r ²	0.38	–	0.24	–	0.52

^a Global refers to the entire NRCS global soil spectral library with up to 4184 samples.

^b Local refers to local data sets from one of three study sites (BBar, Decker/Bales, or Rock Creek).

^c Global-CaCO₃ includes only those NRCS samples taken from profiles with detectable CaCO₃ at some depth.

PLS models might produce more accurate predictions using the small site-specific data sets (i.e. small training data sets) than BRT models. This was intended to provide more accurate local predictions with which to compare to global and local+global predictions.

We tested the importance of local calibration by sample weighting for the combined local and global models at the semiarid grassland sites, by sequentially applying lower weight to the global samples relative to local samples in the BRT modeling procedure. Local samples from the two study sites were always given a full weight of 1, and global samples were given weights of 0.01, 0.25, 0.50, 0.75, and 1 for five iterations of model development. These sample weighting models were constructed in the same fashion as described for BRT models.

3.4. Model evaluation

Predicted and measured values for all three study sites were compared for validation by calculating a mean squared deviation (MSD) and root mean square deviation (RMSD). The MSD was broken into components of standard bias (SB), non-unity (NU), and lack of correlation (LC) which along with the standard error of prediction (SEP) and residual prediction deviation (RPD), were calculated with the following equations (Gauch et al., 2003):

$$SB\% = \frac{(\mu(\text{predicted}) - \mu(\text{validation}))^2}{MSD}$$

$$NU\% = \frac{(1-b)^2 * \sum_n (n(\text{predicted}) - \mu(\text{predicted}))^2}{MSD}$$

$$LC\% = \frac{(1-r^2) * \sum_n (n(\text{validation}) - \mu(\text{validation}))^2}{MSD}$$

$$SEP = \sqrt{(SSE/(N-1))}$$

$$RPD = s/SEP$$

where μ refers to the mean of either predicted values or the validation samples, b is the slope of the least squares regression line through the plot of measured values as a function of predicted values, n is the sample size of either the predicted values or validation samples, N is the number of pairs of predicted values and validation samples, r^2 is the square of the correlation, SSE is the sum of squared errors between predictions and observations, and s is the standard deviation of the validation samples. Standard Bias quantifies the proportion of the MSD related to the deviance of the least squares fit from a 1:1 relationship in the y direction (intercept). Non-unity quantifies the proportion of the MSD related to the deviance of the least squares fit from a 1:1 relationship in the slope of

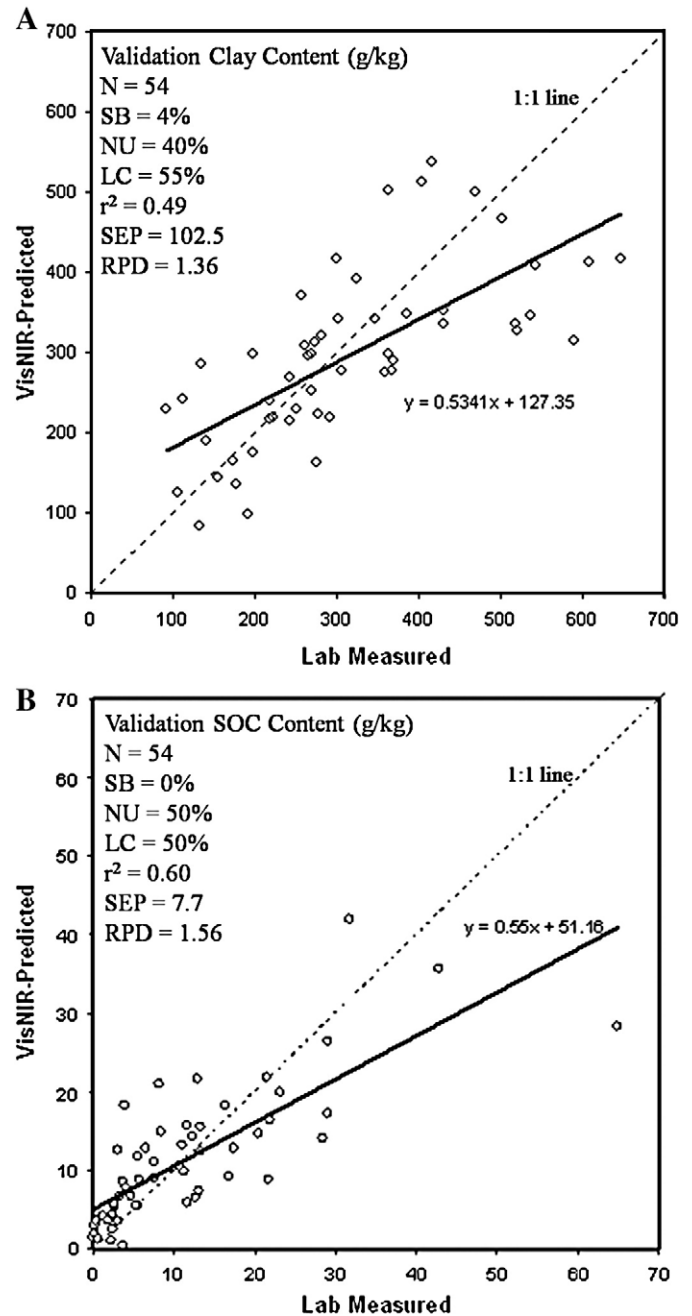


Fig. 4. Decker/Bales ranch (semiarid grassland) VisNIR-predicted versus laboratory measured for full local calibration set+global library for (A) clay content and (B) SOC. Solid line represents simple linear regression of the measured soil property as a function of the predicted soil property for specific model. Dashed line represents $y=0+x$.

the fitted line. Lack of correlation quantifies the proportion of the MSD related to the scatter of the points in relation to the regression line.

4. Results

We obtained the most accurate clay predictions for two of the three sites with local+global calibrations (Table 3, and Figs. 4–7). The differences between predictions using global and local samples versus predictions using local samples alone were substantial, though no explicit test of statistical significance was performed, and different trends were observed at the sites. The SEP decreased from 108.5 to 95.4 g kg⁻¹

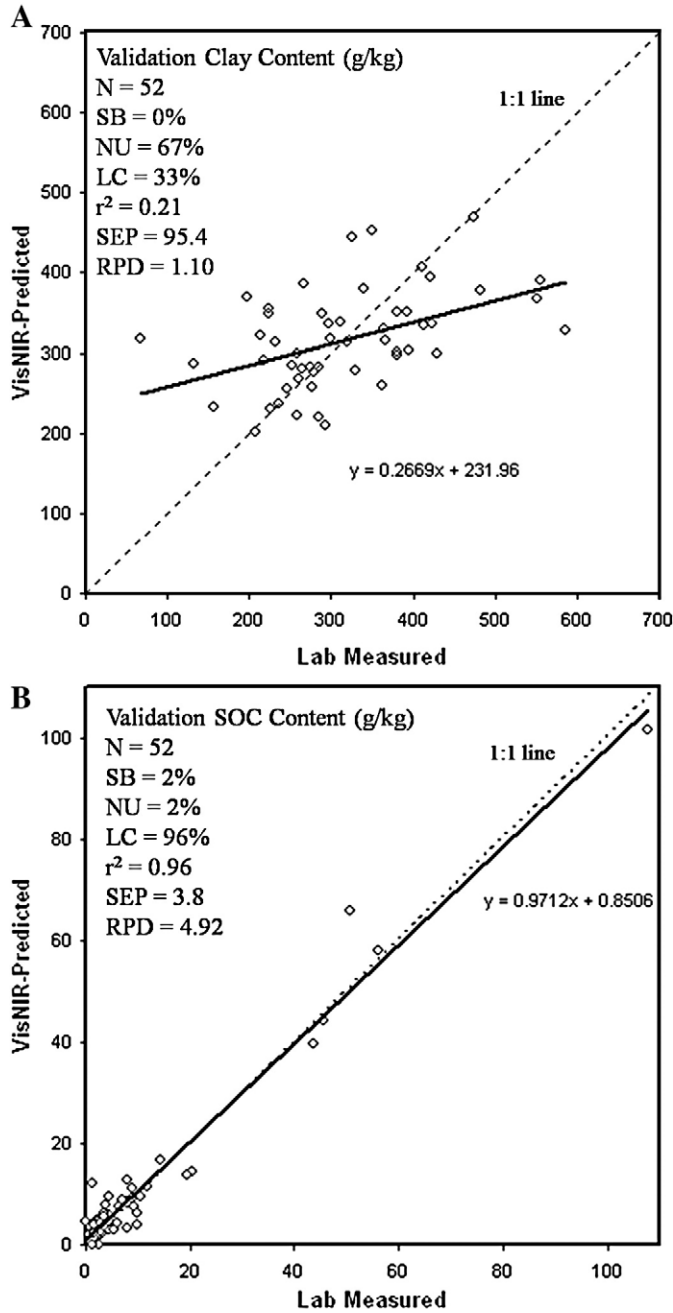


Fig. 5. BBar ranch (semiarid grassland) VisNIR-predicted versus laboratory measured for full local calibration set+global library for (A) clay content and (B) SOC. Solid line represents simple linear regression of the measured soil property as a function of the predicted soil property for specific model. Dashed line represents $y=0+x$.

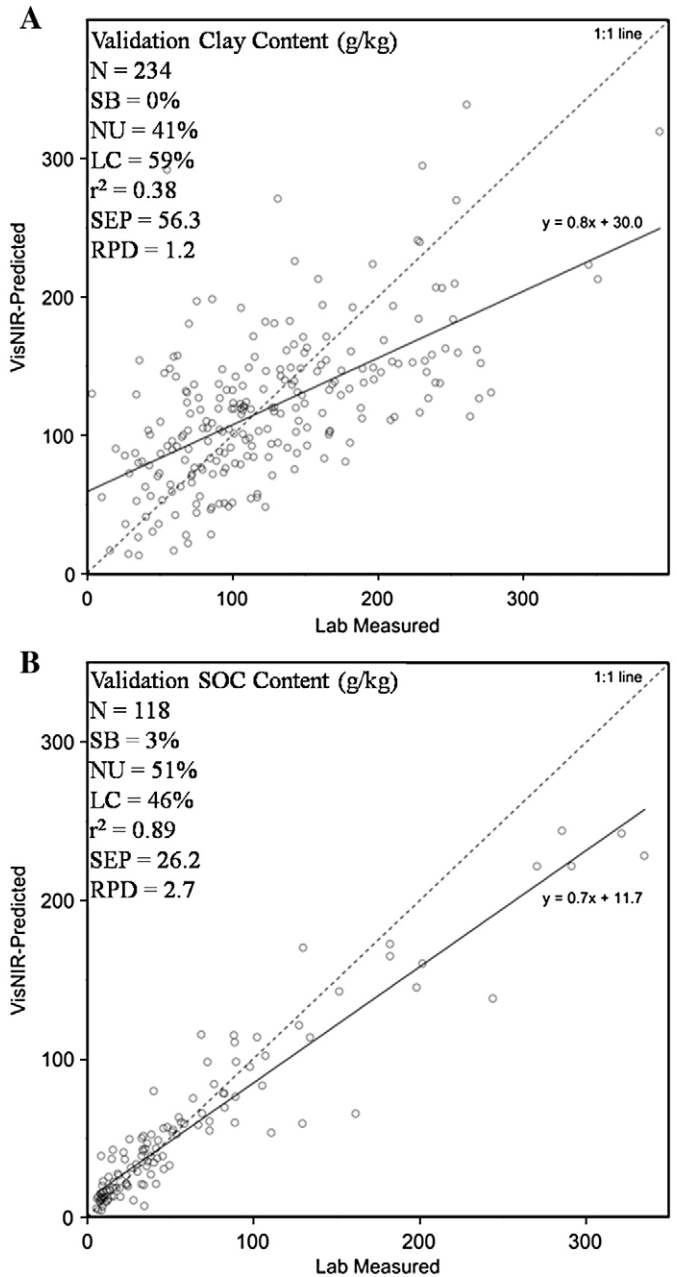


Fig. 6. Subalpine forest VisNIR-predicted versus laboratory measured for full local calibration set+global library for (A) clay content and (B) SOC. Solid line represents simple linear regression of the measured soil property as a function of the predicted soil property for specific model. Dashed line represents $y=0+x$.

at the BBar (semiarid grassland) site, and from 137.6 to 102.5 g kg⁻¹ at the Decker/Bales (semiarid grassland) site, when global samples were used in addition to local samples. The SEP increased from 49.2 g kg⁻¹ to 56.3 g kg⁻¹ when global samples were used in addition to local samples at the subalpine forest site, but this was also the site and soil property with the largest number of local calibration samples (N=210). The difference between predictions using global and local samples versus predictions using global samples alone, were notable for the three sites. The SEP improved from 123.1 to 102.5 g kg⁻¹ when local samples were added to the global data set at the Decker/Bales (semiarid grassland) site. The SEP for clay predictions improved from 65.1 g kg⁻¹ to 56.3 g kg⁻¹ when the local samples were added to the global library at the subalpine forest site. The SEP very slightly improved (by <1.0 g kg⁻¹ clay) when global and local samples were combined relative to global samples alone at the BBar (semiarid grassland) site. We estimated an SEL of 25.3 g kg⁻¹

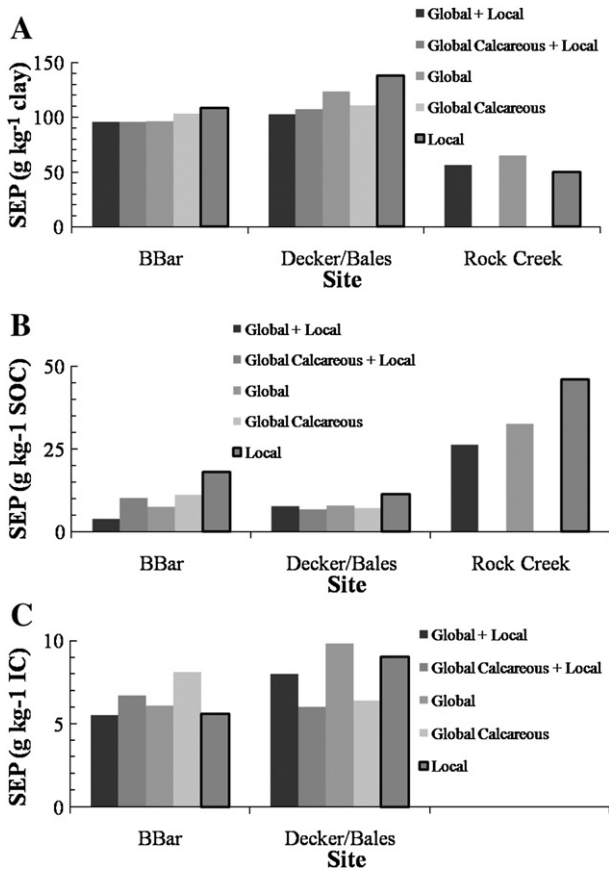


Fig. 7. Overview of (A) clay content, (B) SOC, and (C) IC prediction results for calibration data combinations tested at the semiarid grassland (BBar and Decker/Bales) and subalpine forest (Rock Creek) sites. IC was not predicted for the subalpine forest site. Note that BRT (not PLSR) results are presented for the local predictions of SOC at the BBar semiarid grassland and subalpine forest sites.

for clay characterization based on replicate analysis of subalpine forest samples.

Soil organic carbon predictions using global and local samples at all sites were very promising and produced moderate to strong r^2 between measured and predicted SOC (Table 4). The combination of global and local data predicted SOC with an SEP 3.8 g kg^{-1} at the BBar (semiarid grassland) site and 7.7 g kg^{-1} at the Decker/Bales (semiarid grassland) site, where measured SOC ranged from 0 to 108 g kg^{-1} and 0 to 65 g kg^{-1} , respectively. The local+global combination predicted SOC with an SEP of 26.2 g kg^{-1} at the subalpine forest site where measured SOC ranged from 3 to 335 g kg^{-1} . The combination of local samples with the calcareous

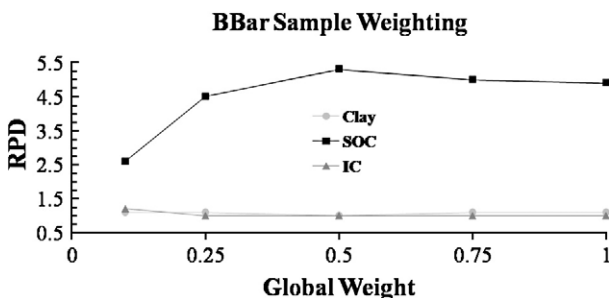


Fig. 8. Sample weighting validation results for clay, SOC, IC predictions at BBar (semiarid grassland) study site.

Table 4
Validation results for soil organic carbon content (g kg^{-1} SOC) predictions

	Global ^a & local ^b	Global-CaCO ₃ ^c & local	Global	Global-CaCO ₃	Local
<i>BBar (semiarid grassland)</i>					
N (validation)	52	52	52	52	52 ^d
SEP (g kg^{-1} SOC)	3.8	10.1	7.4	11.2	7.2
RPD	4.92	1.85	2.51	1.66	2.6
SB (%)	2	4	42	0	4
NU (%)	2	83	27	95	32
LC (%)	96	12	31	5	64
r^2	0.96	0.88	0.93	0.91	0.86
<i>Decker/Bales (semiarid grassland)</i>					
N (validation)	54	54	54	54	54
SEP (g kg^{-1} SOC)	7.7	6.7	7.8	7.1	11.3
RPD	1.56	1.81	1.55	1.70	1.07
SB (%)	0	1	0	0	4
NU (%)	50	64	55	70	93
LC (%)	50	35	45	30	3
r^2	0.60	0.75	0.60	0.72	0.31
<i>Subalpine forest (Rock Creek)</i>					
N (validation)	118	–	118	–	118 ^e
SEP (g kg^{-1} SOC)	26.2	–	32.6	–	29.4
RPD	2.7	–	2.1	–	2.4
SB (%)	3	–	2	–	0
NU (%)	51	–	42	–	25
LC (%)	46	–	56	–	75
r^2	0.89	–	0.80	–	0.82

^a Global refers to the entire NRCS global soil spectral library with up to 4184 samples.

^b Local refers to local data sets from one of three study sites (BBar, Decker/Bales, or Rock Creek).

^c Global-CaCO₃ includes only those NRCS samples taken from profiles with detectable CaCO₃ at some depth.

^d PLSR results presented in table. BRT comparison results: SEP=18.1; RPD=1.03; r^2 =0.39.

^e PLSR results presented in table. BRT comparison results: SEP=46.0; RPD=1.5; r^2 =0.73.

global samples produced the lowest SOC average prediction errors at the Decker/Bales (semiarid grassland) site (SEP= 6.7 g kg^{-1}). We estimated an SEL of 0.64 g kg^{-1} for SOC characterization based on replicate analysis of subalpine forest samples.

Inorganic carbon was modeled for the semiarid grassland sites only (Table 5 and Fig. 7). The local data set produced the second most accurate IC predictions at the BBar (semiarid grassland) site, interestingly. The SEP of 5.6 g kg^{-1} produced using just the local set

Table 5
Validation results for inorganic carbon content (g kg^{-1} IC) predictions at the semi-arid grassland sites

	Global ^a & local ^b	Global-CaCO ₃ ^c & local	Global	Global-CaCO ₃	Local
<i>BBar (semiarid grassland)</i>					
N (validation)	52	52	52	52	52
SEP (g kg^{-1} IC)	5.5	6.7	6.1	8.1	5.6
RPD	1.02	0.84	0.92	0.70	1.00
SB (%)	0	4	2	32	1
NU (%)	51	41	49	26	96
LC (%)	49	55	49	41	3
r^2	0.17	0.07	0.09	0.08	0.02
<i>Decker/Bales (semiarid grassland)</i>					
N (validation)	54	54	54	54	54
SEP (g kg^{-1} IC)	8.0	6.0	9.8	6.4	9.0
RPD	1.18	1.56	0.95	1.46	1.04
SB (%)	29	7	34	9	2
NU (%)	46	36	44	33	94
LC (%)	24	56	22	58	3
r^2	0.52	0.63	0.29	0.58	0.17

^a Global refers to the entire NRCS global soil spectral library with up to 4184 samples.

^b Local refers to local data sets from one of three study sites (BBar, Decker/Bales, or Rock Creek).

^c Global-CaCO₃ includes only those NRCS samples taken from profiles with detectable CaCO₃ at some depth.

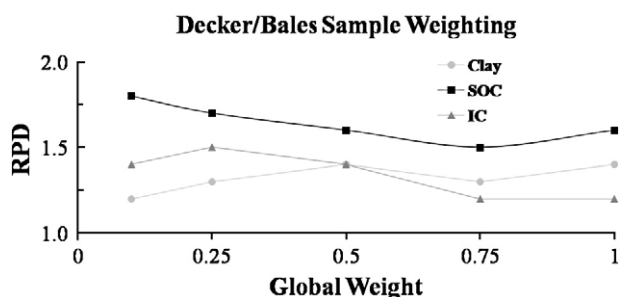


Fig. 9. Sample weighting validation results for clay, SOC, IC predictions at Decker/Bales (semiarid grassland) study site.

was comparable to the SEP of 5.5 g kg^{-1} resultant from the combined global and local data sets at this site. The SEP decreased from 9.0 to 8.0 g kg^{-1} when the local data set was augmented with the entire global set at the Decker/Bales (semiarid grassland) site. Using only the calcareous global samples combined with local samples yielded notably better results for the Decker/Bales (semiarid grassland) site with $\text{SEP}=6.0 \text{ g kg}^{-1}$. We estimated an SEL of 0.17 g kg^{-1} IC based on replicate analysis of semiarid grassland samples.

Local-only calibrations, based on smaller datasets, yielded less accurate predictions than local+global calibrations in six of the eight instances where this was tested. The two exceptions included clay content at the subalpine forest site (best results) and the semiarid grassland IC results at the BBar site (~equal to local+global calibration). In most instances, the use of PLS in place of boosted regression tree models for site-specific data sets did not substantially improve upon this general outcome (results not presented). Partial least squares produced more accurate SOC predictions for the BBar semiarid grassland and subalpine forest local calibrations and are therefore presented (Table 4) with BRT model results for comparison.

Weighting global samples less heavily (in BRT modeling only) did not improve predictions of clay content at either site (Figs. 8 and 9). Such sample weighting improved SOC predictions slightly at the Decker/Bales semiarid grassland site, where the RPD increased from 1.6 to 1.8 (Fig. 9) and the SEP decreased from 7.7 to 6.6 g kg^{-1} (results not presented) when the global samples were given a weight of 0.1 versus 1.0. Soil organic carbon predictions showed a marginal improvement in accuracy when global samples were given a weight of 0.5 versus 1.0 at the BBar semiarid grassland site (Fig. 8). Sample weighting improved predictions slightly for IC at both semiarid grassland sites, where less than full weights assigned to global data produced the most accurate IC predictions (Figs. 8 and 9).

5. Discussion

5.1. Does a larger global spectral library augmented with local samples provide better predictions than global samples alone?

The combination of a larger global data set and a smaller local data set produced more accurate predictions than global samples alone in all instances for this study (Fig. 7). This observed improvement in prediction gained through the addition of a proportionately small number of local samples is possibly due to the fact that the global library used in this study did not contain any samples from the local study site locations. The added local calibration samples were therefore considerably more similar to the validation samples than were most of the global library samples. Different results might be expected from a study based on other local sites, as well as a study in which the global library included samples from the local site of prediction.

The clay prediction accuracies at the subalpine forest site were consistent with or better than two previous studies that employed regional or global calibration data sets (Shepherd and Walsh, 2002; Brown

et al., 2006). Soil organic carbon predictions for the semiarid grassland sites were consistent with previous regional and global studies (McCarty et al., 2002; Shepherd and Walsh, 2002; Brown et al., 2006). It appeared that prediction accuracy for both properties might be influenced not only by local calibration sample size, but the complexity of local soil processes (i.e. sampled range of soil properties) as well, assuming our local data sets accurately reflect the population of soil properties at their respective study sites. The combination of local and global data produced slightly more accurate predictions at both sites for which IC was modeled. Inorganic carbon predictions were less accurate than the best local/regional calibrations reported in two previous studies, but are comparable to some of the geographically independent validations of those same predictions (McCarty et al., 2002; Brown et al., 2005). Slightly different methods were used to determine IC for the local datasets [modified pressure calcimeter (Sherrod, 2002)] compared to the global dataset [HCl treatment and manometer (Soil Survey Staff, 1996)] used in this study. Brown et al. (2006) found IC determined by the two methods to be correlated ($r^2=0.97, n=198$) with an RMSD of 4.3 g kg^{-1} IC, and attributed the observed differences to intra-lab variability and not variability between the two similar methods. Nonetheless, a lack of precision between IC estimates by the two characterization methods might translate to a decreased precision in combined local+global prediction.

Augmenting the “calcareous” global library with local samples produced more accurate predictions relative to the “calcareous” global library alone. We expected the calcareous global samples to more accurately predict soil properties for the semiarid grassland sites (relative to the entire global library) as these sites were generally calcareous. However, the entire global data set (both “noncalcareous” and “calcareous” profiles) augmented with local samples produced more accurate predictions of SOC and IC at the BBar ranch than either the global calcareous samples alone or the global calcareous/local combination. The global calcareous samples augmented with local samples did produce some of the most accurate predictions of all three properties at the Decker/Bales ranch. The Decker/Bales ranch is located in the heart of the semiarid, non-glaciated, Northern Great Plains. The calcareous global data set was possibly a more appropriate spectral library for this site compared to the BBar ranch which is located at the westernmost edge of the non-glaciated plains, where climate and soil development are directly affected by the proximity to the Rocky Mountain front.

5.2. Does a larger global spectral library augmented with local samples provide better predictions than local samples alone?

The results for all three predicted properties suggest that a larger global library augmented with local samples often provides better predictions than local samples alone. Instances were observed where local samples alone produced some of the best predictions, however. The difference in both cases was often substantial. Clay, for example, was predicted with approximately 13 g kg^{-1} and 35 g kg^{-1} better accuracy at the semiarid grassland sites and, conversely, 7 g kg^{-1} worse accuracy at the subalpine forest site, by the local+global combination compared to local samples alone. The subalpine forest had a much larger local calibration sample size (~4 times larger than semiarid grassland datasets) and smaller sampled range and mean of clay content. The large sample size also allowed for one half of the samples to be reserved for independent validation at the subalpine forest, whereas cross-validation was employed at the semiarid sites. The global spectral library used in this study was constructed from United States Department of Agriculture sample archives and is inherently biased towards soils from farm and range landscapes of the world (Brown et al., 2006). The subalpine forest site is a mountain environment that might not be well represented by samples included in this global library.

In contrast to clay predictions, SOC predictions at the subalpine forest site appeared to be more accurate for local+global calibration compared to local calibration. PLSR produced predictions by local calibration at the

subalpine forest site that, while more accurate than BRT local predictions, still had higher SEP, and lower RPD and r^2 values than local+global SOC predictions at this site. It is important to note that the subalpine forest dataset was characterized for SOC using the dry combustion method, whereas local+global analysis at the subalpine forest site employed a subset of the global dataset characterized for SOC with the Walkley–Black method. In construction of the global dataset, Brown et al. (2006) found SOC from dry combustion and SOC from Walkley–Black to be correlated ($r^2=0.95$, $n=1175$) with an RMSD of 6 g kg^{-1} SOC. The observed difference between local and local+global SOC prediction accuracies at the subalpine forest site might not be significant in light of the potential lack of precision between SOC estimates using the two different methods.

Soil inorganic carbon prediction for the BBar ranch, the more humid and less calcareous (i.e. smaller IC mean and range) of the semiarid sites, was an example where predictions by local samples alone were similar in accuracy compared to local+global predictions. The low r^2 statistic for the predictions by BBar local samples suggests that predictive accuracy was quite low. Different results might be expected for the semiarid grassland sites, could more local samples be feasibly collected.

5.3. Do weighted samples from a global spectral library augmented with un-weighted local samples provide better predictions than un-weighted samples from both sources?

Clay content predictions by the models that weighted local and global samples equally were not substantially different than the predictions by the weighted models that produced the most accurate predictions for clay content at both semiarid grassland sites. Weighting did produce IC and SOC prediction accuracy that, in instances, improved on non-weighting. The choice of the optimum weight for prediction of each of the properties appeared to be site-specific, and possibly specific to the estimated property as well. The previously untested approach, at best, slightly improved SOC and IC prediction accuracy in this study.

6. Conclusions

This study found that a larger global spectral library augmented with a small set of local samples was often capable of providing improved VisNIR diffuse reflectance spectroscopy predictions of clay, SOC and IC, compared to calibrations derived from global or local samples alone. Furthermore, the use of a potentially more site-specific subset of only the calcareous global samples in conjunction with local samples improved prediction accuracy of some variables. Preferentially weighting local samples in Boosted Regression Tree (BRT) local+global calibrations yielded only slight improvements in SOC and IC predictions for some modeling scenarios.

This study was based on samples collected for three local sites in a temperate soil environment, and inferences about the relative performance of global, local, and combined local+global predictions in such an environment are limited to the sites employed. The future production of more local soil spectral libraries from a variety of soils and landscapes would allow for more robust comparisons of the performance of global, local, and combined local+global predictions. Future research on local+global calibrations should also address the optimum local sample size as well as approaches for selecting samples similar to local soils from larger global or regional soil spectral libraries. Though a great deal of work is required to fine tune this general approach, we anticipate that future success in VisNIR DRS soil characterization will rely heavily on optimizing local+global calibrations.

Acknowledgements

Financial support for this project was provided in part by Big Sky Regional Carbon Sequestration Partnership (DOE Award No. DE-FC26-05NT42587) and USDA-CSREES-NRI Award No. 2003-35107-13774.

References

- Bernard, M.L. (in preparation). Fuzzy landscape modeling for predictive soil mapping of the Beartooth Mountains, Montana. M.S. Thesis, Montana State University, Bozeman, MT, USA.
- Brown, D.J., 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. *Geoderma* 140, 444–453.
- Brown, D.J., Bricklemeyer, R.S., Miller, P.R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VisNIR soil C prediction in Montana. *Geoderma* 129, 251–267.
- Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, D.M., Reinsch, T.G., 2006. Global soil characterization with VisNIR diffuse reflectance library. *Geoderma* 132, 273–290.
- Chang, C., Laird, D.L., Mausbach, M.J., Hurburgh, C.R., 2001. Nearinfrared spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal* 65, 480–490.
- Clark, R.N., 1999. Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz, N. (Ed.), *Remote Sensing for the Earth Sciences: Manual of Remote Sensing*. John Wiley and Sons, New York, pp. 3–52.
- Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australia. *Australian Journal of Experimental Agriculture* 42, 607–614.
- Forest Service Staff, 1986. Custer National Forest: Final Environmental Impact Statement. U.S. Dept. of Agriculture, Forest Service, Billings, MT, USA.
- Gauch, H.G., Hwang, J.T., Fivk, G.W., 2003. Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal* 95, 1442–1446.
- Gee, G.W., Bauder, J.W., 1986. Particle size analysis. In: Klute, A. (Ed.), *Methods of Soil Analysis: Part 1. Physical and Mineralogical Methods*. Soil Science Society of America, Madison, WI, pp. 383–411.
- Henderson, T.L., Baumgardner, M.F., Franzmeier, D.P., Stott, D.E., Coster, D.C., 1992. High dimensional reflectance analysis of soil organic matter. *Soil Science Society of America Journal* 56, 865–872.
- Hunt, G.R., 1989. Spectroscopic properties of rocks and minerals. In: Carmichael, R.S. (Ed.), *Practical Handbook of Physical Properties of Rocks and Minerals*. CRC Press, Boca Raton, FL, pp. 295–385.
- Islam, K., Singh, B., McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research* 41 (6), 1101–1114.
- McCarty, G.W., Reeves III, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal* 66, 640–646.
- Montagne, C., Munn, L.C., Nielsen, G.A., Rogers, J.W., Hunter, H.E., 1982. Soils of Montana. Bulletin 744, Montana Agricultural Experiment Station, Montana State University, Bozeman, MT, USA.
- Sankey, J.B., 2005. Modeling soil water for precision range management. M.S. Thesis, Montana State University, Bozeman, MT, USA.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* 66, 988–998.
- Sherrod, L.A., 2002. Inorganic carbon analysis by modified pressure calcimeter method. *Soil Science Society of America Journal* 66, 299–305.
- Soil Survey Staff, 1971. Soil Survey for the Powder River Area Montana. United States Department of Agriculture, Soil Conservation Service, Washington, D.C.
- Soil Survey Staff, 1996. Soil Survey Laboratory Methods – Soil Survey Investigations Report Number 42. United States Department of Agriculture, Natural Resources Conservation Service, National Soil Survey Center, Washington D.C.
- Soil Survey Staff, 2004. Soil Survey (SSURGO) database for Sweet Grass County Area, Montana. United States Department of Agriculture, Natural Resources Conservation Service, Fort Worth, TX, USA. <http://soildatamart.nrcs.usda.gov>.
- Sorensen, L.K., Dalsgaard, S., 2005. Determination of clay and other soil properties by near infrared spectroscopy. *Soil Science Society of America Journal* 69, 159–167.
- VanGosen, B.S., et al., 2000. Generalized geologic map of the Absaroka–Beartooth study area, south-central Montana. Miscellaneous Field Studies Map, Report Number 2338, U.S. Geological Survey, Denver, CO, USA.
- Veseth, R., Montagne, C., 1980. Geologic parent materials of Montana Soils. Bulletin 721, Montana Agricultural Experiment Station, Montana State University, Bozeman, MT, USA.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131, 59–75.