

DETECTION OF SOIL PROPERTIES WITH AIRBORNE HYPERSPECTRAL MEASUREMENTS OF BARE FIELDS

W. R. DeTar, J. H. Chesson, J. V. Penner, J. C. Ojala

ABSTRACT. Remote sensing with aircraft-based sensors can provide the fine resolution required for site-specific farming. The within-field spatial distribution of some soil properties was found by using multiple linear regression to select the best combinations of wave bands, taken from among a full set of 60 narrow bands in the wavelength range of 429 to 1010 nm. The resulting regression equations made it possible to calculate the value of the soil property at every pixel, with a spatial resolution of 1.2 m. Both surface and subsurface samples of soil were taken from the center of each of 321 equal-sized grids on 128 ha of recently seeded and nearly bare soil. The soil samples were tested in a laboratory for 15 different properties. The percent sand in surface samples was found to be detectable with a reasonable degree of accuracy with $R^2 = 0.806$ for a four-parameter model; the best combination of wavelengths was 627, 647, 724, and 840 nm. For silt, clay, chlorides, electrical conductivity, and phosphorous, the results were somewhat less satisfactory with a range of $0.66 < R^2 < 0.76$. The poorest fit was for carbon with $R^2 = 0.27$. Organic matter and saturation percentage had $R^2 < 0.49$. For the remaining properties, i.e., pH, Ca, Mg, Na, K, and bicarbonates, the correlation was intermediate and statistically significant, but with a great deal of scatter around the regression lines. An example of an image map was produced showing the percent sand at every pixel location in one field. New spectral indices were developed; one index ($I = R_{763} - 0.85 \cdot R_{753} - 0.24 \cdot R_{657} - 0.40 \cdot R_{443}$) was found to work well with five of the soil properties (EC, Ca, Mg, Na, and Cl), indicating some commonality in the manner in which they affected the reflectance from the soil surface, possibly due to a salinity effect. Multiple linear regressions were also run on every possible combination of four broader bands in the blue, green, red, and near-infrared regions of the spectrum, resulting in R^2 values lower than with the various combinations of narrow bands. The main findings were that (1) some soil properties can be accurately detected using airborne remote sensing over nearly bare fields, and (2) it is possible to produce a fine-resolution, farm-size, soils map showing the in-field distribution of these properties.

Keywords. Band selection, Bare fields, Hyperspectral imagery, Multiple linear regression, Precision agriculture, Remote sensing, Soils map, Soil properties, Spectral indices.

Site-specific crop management requires soil property data with fine spatial resolution. Remote sensing in the form of hyperspectral imagery (HSI) on an aircraft platform can provide the necessary resolution, and in addition it supplies a complete spectrum of data for every pixel location. This combination creates a potential for accurate detection of soil properties. Recent studies have shown the usefulness of spectral indices developed from HSI data to characterize biophysical properties in agriculture (Haboudane et al., 2002; Goel et al., 2003). Haboudane et al. (2002) combined a radiative simulation model with HSI data to generate predictive equations that worked quite well for chlorophyll. Goel et al. (2003) used multiple linear regression of HSI reflectance data to find the best combination of narrow wave bands for detecting nine different properties of a corn crop. They found that some modified forms of the

normalized difference vegetation index (NDVI) performed better than five-term linear regression models.

Thomasson et al. (2001) used spectrophotometer readings of soil samples to measure nine soil properties, including pH, Ca, Mg, Zn, Na, P, K, sand, and clay. The best responses were with Ca and Mg for soil from one field where the R^2 values were 0.72 and 0.73, respectively. Palacios-Orueta and Ustin (1998), using remote-sensing reflectance data, found that Fe, organic matter, and sand were the main factors affecting the spectral response curves. Low reflectance was noted in soils with low sand content. Barnes and Baker (1999) found that aerial and satellite images could be used to map soil texture in individual fields with a reasonable degree of accuracy. Barnes et al. (1996) suggested that spectral reflectance data could be used to improve the interpolation of gridded soil samples. Chen et al. (2000) used two different approaches for remote sensing of a large bare field to detect soil organic carbon. In one method, using red, green, and blue bands, a regression equation with $R^2 = 0.93$ was used to calculate values at every pixel. The resulting distribution agreed well with a classification approach.

Hong et al. (2002) attempted to measure the within-field variation of seven soil properties (Mg, K, P, pH, cation exchange capacity, organic matter, and apparent electrical conductivity) using airborne hyperspectral images on 35 ha of bare soil. They applied both stepwise multiple regression and principal component regression (PCR) on the full HSI data

Submitted for review in February 2007 as manuscript number SW 6864; approved for publication by the Soil & Water Division of ASABE in March 2008.

The authors are **William R. DeTar**, ASABE Member Engineer, Research Agricultural Engineer, **Joseph H. Chesson**, Agricultural Engineer, **John V. Penner**, Engineering Technician, and **John C. Ojala**, Research Agronomist, USDA-ARS Western Integrated Cropping Systems Research Unit, Shafter, California. **Corresponding author:** William R. DeTar, USDA-ARS, 17053 N. Shafter Ave., Shafter, CA 93263; phone: 661-746-8011; fax: 661-746-1619; e-mail: bill.detar@ars.usda.gov.

set and also used multiple regression on four broad band ranges of HSI data corresponding to Landsat TM bands 1-4. Although stepwise multiple regression on the full HSI data set gave good results, the inclusion of many variables increased the chance of overfitting. The results with the PCR and the Landsat-like bands were generally almost as good and showed potential for estimating the distribution of soil properties with less chance of overfitting.

Selige et al. (2003) used airborne HSI over two bare fields. Using single-band regression on each of 128 narrow bands, a very good non-linear correlation was found for clay content using the wave band in the range of 2427-2436 nm, with $R^2 = 0.78$. The best band for organic matter was found at 344-357 nm, with $R^2 = 0.68$. Selige et al. (2006) used airborne HSI over 12 bare fields with a total area of 700 ha at a spatial resolution of 6 m. There were 128 bands in the range of 420-2480 nm. After first removing some bands using partial least-square regression (PLSR), they used multiple linear regression of the remaining bands to find the best bands and predictive equations for four soil properties: nitrogen, organic carbon, sand, and clay. In this case, the best four bands for clay were in the range of 902-1165 nm with $R^2 = 0.65$. The best four bands for organic matter were in the range of 800-1322 nm with $R^2 = 0.86$, and for sand they were all greater than 2200 nm with $R^2 = 0.87$. Bajwa and Tian (2005) used PLSR analysis to relate several fertility-related soil properties to airborne HSI data from four bare fields. The spectral range was 471-828 nm. The R^2 for Ca and Mg were 0.82 and 0.72, respectively.

The current study shows the degree to which 15 soil properties can be detected with airborne HSI data taken over large, essentially bare fields, using a multiple linear regression procedure for band selection where every possible combination of 60 narrow bands (429 to 1010 nm) is regressed in two-, three-, and four-parameter equations. The ultimate objective is to accurately show the in-field distribution of soil properties.

MATERIALS AND METHODS

STUDY AREA AND SOIL SAMPLING

The study site included two fields of a large commercial farm on the western side of the San Joaquin Valley of California, in Kings County, near the town of Stratford. The center of one field, called 4-1, was located at 36.222° N and 119.918° W; the field had an area of 72 ha with an average elevation of 71 m. The center of the other field, called 6-4, was located at 36.214° N and 119.954° W; the field had an area of 63 ha with an elevation of 76 m. Both fields were nearly square in shape. Acala cotton was planted in mid-April, and the plants had reached about 5% canopy cover by mid-May. The soil in both fields was classified as Lethent silty clay loam, slight to strong alkali (fine, smectitic, thermic Typic Natrargids). The parent material was sedimentary rock alluvium. It was a deep, poorly drained soil on nearly level land. For the area in general, perched groundwater over deep clay layers transported salts from the higher regions to the lower fans, basin, and basin rims, imposing the need of drainage management (USDA, 1978).

Soil samples were collected in April 2002 from 173 locations in field 4-1 and from 148 locations in field 6-4. The location of each sample site was set by global positioning sys-

tem (GPS) coordinates as the center of a 0.4 ha polygon grid area. Two samples were taken at each of the 321 sites, one scraped from the soil surface and the other a combination of 8 subsamples taken with a 19 mm diameter soil probe to a depth of 300 mm at random points within 3 m of the center of the sample site. Each sample weighed about 1700 g. Sample preparation consisted of air-drying in a laboratory and then grinding until the entire sample passed a 2 mm sieve. There was no tillage of the soil between the time of the sampling and time of the flight, and the soil surface was dry at both times. It is assumed that other differences were minor.

LABORATORY ANALYSIS OF SOIL SAMPLES

Half of each sample was sent for analysis to the Agriculture and Natural Resources Analytical soils laboratory at the University of California in Davis, California. Starting with a vacuum extraction of a saturated paste, the first property measured was saturation percentage (SP). Soil pH was determined directly on the saturation paste using an electrode probe; from the extract, estimates were made of electrical conductivity (EC), Ca, Mg, Na, Cl, and HCO_3 . Carbon (C) was determined by flash combustion coupled with a gas chromatograph separation system. Extractable phosphorus (P) was determined by the Olsen method, and exchangeable potassium (K) was determined by extraction with sulfuric acid. Organic matter (OM) was measured by sample ignition. Particle size (sand, silt, and clay) was measured by the hydrometer method.

HYPERSPECTRAL DATA ACQUISITION

The Airborne Visible and Near-Infrared (AVNIR) hyperspectral sensor (OKSI, Inc., Torrance, Cal.) was flown over the site on 22 May 2002. The AVNIR images, acquired at 2200 m above ground level, produced 1.2 m spatial resolution with 60 bands, covering the spectral range from 429 to 1010 nm with a 10 nm bandwidth. The wavelength for the center of each band is shown in table 1. As with Landsat bands, it was found to be more convenient to refer to the bands by number, rather than by wavelength, and this convention is used throughout the rest of this article. Atmospheric calibration was performed using black and gray, 8×8 m reference panels located near the fields. Images were georeferenced and registered using small white panels located at the corners of the fields. Spectral radiometer readings of the calibration panels were taken using a model LI-1800 from LICOR, Inc. (Lincoln, Neb.).

DATA ANALYSIS

Image processing and initial statistical analysis were conducted using ENVI 4.1 software (Research Systems, Inc., Boulder, Colo.). The points representing the location of the soil sample sites were superimposed on the HSI images of each field. A circular area, made large enough to include 150 pixels (about a 7.6 m radius), was selected at each sample site, and the average reflectance for each of the 60 bands was exported as an ASCII file for further processing. Microsoft Excel was then used to convert these files to a format useable by CoPlot v3.0 (CoHort Software, Monterey, Cal.) and by ArcView GIS v3.3 (ESRI, Redlands, Cal.). The master processing file in Excel contained 321 rows of data, one for each sampling site, and 60 columns, one for each band. Soil property data were added one at a time in column 61, and this file

Table 1. Wavelengths and corresponding band numbers used in the hyperspectral analysis.

Wavelength (nm)	Band No.						
1005.11	B1	859.86	B16	714.61	B31	569.36	B46
995.43	B2	850.18	B17	704.93	B32	559.68	B47
985.75	B3	840.49	B18	695.24	B33	549.99	B48
976.06	B4	830.81	B19	685.56	B34	540.31	B49
966.38	B5	821.13	B20	675.88	B35	530.63	B50
956.70	B6	811.44	B21	666.19	B36	520.94	B51
947.01	B7	801.76	B22	656.51	B37	511.26	B52
937.33	B8	792.08	B23	646.83	B38	501.58	B53
927.65	B9	782.39	B24	637.14	B39	491.89	B54
917.96	B10	772.71	B25	627.46	B40	482.21	B55
908.28	B11	763.03	B26	617.78	B41	472.53	B56
898.60	B12	753.34	B27	608.09	B42	462.84	B57
888.91	B13	743.66	B28	598.41	B43	453.16	B58
879.23	B14	733.98	B29	588.73	B44	443.48	B59
869.55	B15	724.29	B30	579.04	B45	433.79	B60

became the primary source for multiple linear regression analysis, which was performed using the CoPlot software.

MULTIPLE LINEAR REGRESSION

The soil property was considered the dependent variable, and the various band reflectances were the independent variables. There is an automatic procedure available in CoPlot, in which, after a choice of the number of bands to include in each multiple regression is made, the program looks at every possible combination, returning only the R^2 values. The number of regressions required for pairs of bands out of 60 available is 1770. To find the best three-band combinations required 34,220 regressions. Four-band combinations required 487,635 regressions. The program automatically ranks and stores the results of the 100 best combinations (models).

IMAGE MAP

The entire HSI image for each field was converted to ASCII format and loaded into Microsoft Access; there were nearly 500,000 rows and 62 columns (60 columns for reflectance plus two more for longitude and latitude) of data. Using the bands and equations developed with the multiple linear regression, queries were formed to calculate the predicted value of a soil property at each pixel. These values along with

their coordinates were then loaded into ArcView for color selection. The final step for the image map was to convert it to a TIFF format.

RESULTS AND DISCUSSION

BAND SELECTION AND SPECTRAL INDICES

The coefficient of determination (R^2) obtained between the various soil properties and the multi-parameter spectral reflectance models are shown in table 2 for the samples taken from the soil surface. Table 3 shows the same type of data using the subsurface samples. It is clear that some of the properties correlate quite well, especially sand, which had an $R^2 > 0.80$ for both surface and subsurface samples using the four-parameter model. In comparing tables 2 and 3, one notes that for sand, the best models contain almost the same bands for both the surface and subsurface samples. For the other properties, however, the correlation is much better with the surface samples. In addition, the surface and subsurface tests do not share many of the same bands. For surface samples of silt, clay, chlorides, EC, and P, the results were a little less satisfactory than with sand, with $0.66 < R^2 < 0.76$ using four-parameter models. The poorest fit was for carbon, with $R^2 = 0.27$. OM and SP had $R^2 < 0.49$. For the remaining properties, i.e., pH, Ca, Mg, Na, K, and bicarbonates, the correlation was

Table 2. Bands for the highest R^2 values for surface properties of soil.

Soil Property	One Parameter		Two Parameters		Three Parameters		Four Parameters	
	Band	R^2	Bands	R^2	Bands	R^2	Bands	R^2
SP (%)	B10	0.3978	B4, B30	0.4109	B30, B50, B53	0.4518	B4, B33, B36, B59	0.4859
pH	B3	0.5474	B2, B56	0.5710	B23, B30, B33	0.5900	B3, B7, B13, B26	0.6164
EC (dS m ⁻¹)	B52	0.4633	B57, B58	0.5441	B26, B27, B54	0.6359	B26, B27, B38, B59	0.6693
Ca (meq L ⁻¹)	B52	0.4325	B57, B58	0.5167	B26, B27, B57	0.5918	B26, B27, B38, B60	0.6188
Mg (meq L ⁻¹)	B55	0.3761	B51, B52	0.4508	B26, B27, B57	0.5433	B26, B27, B38, B60	0.5820
Na (meq L ⁻¹)	B52	0.3428	B51, B52	0.4478	B26, B27, B54	0.5850	B26, B27, B38, B59	0.6224
Cl (meq L ⁻¹)	B52	0.6026	B57, B59	0.6613	B26, B27, B54	0.7249	B26, B27, B37, B59	0.7376
HCO ₃ (meq L ⁻¹)	B21	0.3189	B39, B47	0.4019	B27, B30, B47	0.5183	B27, B30, B48, B53	0.5668
C (%)	B9	0.0245	B44, B50	0.1912	B26, B47, B50	0.2527	B32, B44, B47, B50	0.2692
P (mg kg ⁻¹)	B22	0.4452	B39, B46	0.5138	B38, B48, B55	0.6540	B27, B30, B50, B53	0.6975
K (mg kg ⁻¹)	B21	0.3729	B57, B59	0.4820	B27, B29, B47	0.5897	B14, B27, B30, B47	0.6391
OM (%)	B27	0.2617	B47, B50	0.3517	B26, B27, B47	0.4384	B26, B27, B47, B50	0.4857
Sand (%)	B13	0.7623	B2, B38	0.7865	B2, B4, B7	0.7965	B18, B30, B38, B40	0.8063
Silt (%)	B18	0.6878	B2, B7	0.7187	B27, B30, B45	0.7326	B2, B6, B39, B45	0.7518
Clay (%)	B4	0.5699	B4, B35	0.6016	B4, B35, B44	0.6516	B13, B29, B36, B44	0.6708

Table 3. Bands for the highest R² values for subsurface properties of soil.

Soil Property	One Parameter		Two Parameters		Three Parameters		Four Parameters	
	Band	R ²	Bands	R ²	Bands	R ²	Bands	R ²
SP (%)	B10	0.3873	B10, B38	0.4073	B9, B34, B36	0.4413	B4, B9, B32, B38	0.4703
pH	B59	0.1660	B4, B20	0.2320	B12, B31, B32	0.2557	B12, B31, B32, B47	0.2784
EC (dS m ⁻¹)	B57	0.0100	B50, B52	0.1873	B10, B14, B55	0.2937	B8, B14, B27, B59	0.3375
Ca (meq L ⁻¹)	B37	0.0333	B33, B36	0.1105	B21, B32, B37	0.2240	B21, B31, B32, B37	0.2640
Mg (meq L ⁻¹)	B17	0.0311	B50, B52	0.1675	B21, B32, B37	0.2741	B21, B31, B32, B37	0.3089
Na (meq L ⁻¹)	B26	0.0373	B50, B52	0.2048	B11, B14, B55	0.3445	B15, B26, B27, B60	0.4281
Cl (meq L ⁻¹)	B57	0.2515	B51, B52	0.3043	B26 B27, B58	0.3834	B9, B27, B31, B43	0.4359
HCO ₃ (meq L ⁻¹)	B57	0.0999	B33, B44	0.1606	B33, B44, B51	0.1961	B9, B31, B47, B52	0.2029
C (%)	B38	0.0075	B44, B50	0.1909	B44, B50, B55	0.2737	B33, B44, B50, B55	0.3078
P (mg kg ⁻¹)	B47	0.3158	B9, B21	0.4142	B38, B50, B55	0.4921	B26, B30, B48, B58	0.5482
K (mg kg ⁻¹)	B21	0.3490	B50, B59	0.4363	B38, B47, B59	0.5077	B6, B21, B26, B47	0.5591
OM (%)	B34	0.2361	B14, B31	0.3217	B26, B47, B50	0.3995	B23, B27, B47, B50	0.4360
Sand (%)	B13	0.7682	B2, B38	0.7938	B2, B4, B7	0.8012	B18, B30, B38, B39	0.8148
Silt (%)	B18	0.6782	B2, B7	0.7012	B27, B30, B46	0.7168	B1, B35, B45, B60	0.7308
Clay (%)	B33	0.6162	B4, B35	0.6492	B4, B35, B44	0.6927	B3, B35, B44, B55	0.7111

intermediate and statistically significant, but with a great deal of scatter around the regression lines. Increasing the number of parameters in the model from two to three improved the correlation greatly for some soil properties (P and Na) and hardly at all for others (pH, sand, and silt). Going from three parameters to four parameters usually did not improve the R² very much.

For each of the top-ranked band combinations shown in tables 2 and 3, there are 99 more of lower rank stored by the CoPlot program. In some cases, the lower-ranked models have nearly the same R² as the top-ranked model. For example, with the four-parameter model for surface sand, the R² declines from 0.8063 for the top model to 0.7989 for the model ranked number 100. Thus, the top model is definitely not a unique answer. As suggested by Lillesand and Kiefer (1999), a unique solution would only be possible under ideal conditions.

There are certain bands that appear frequently in the top 100 models for surface sand. For example, B38 appears in 43 of the 100 top four-parameter models, and B29 appears 39 times. Forty-two of the top 50 two-parameter models for surface sand include the combination of either B1 or B2 with any band in the range of B32 to B52. It is evident in tables 2 and 3 that the wavelength found for the best single-parameter model often does not appear at all in any of the best multi-parameter models. In the best three-parameter models for surface soil, the same three bands (B26, B27, and B54) appear for three different soil properties (EC, Na, and Cl). The combination of these bands also appears in the top two rankings of five surface soil properties (EC, Ca, Mg, Na, and Cl), and the combination of bands B26, B27, and B57 do likewise. In fact, for these same five soil properties, bands B26 and B27 work well in combination with any band between B52 and B58. For the four-parameter surface soils models, the combination of bands B26, B27, B37, and B59 are in the top three rankings for the same five surface soil properties (EC, Ca, Mg, Na, and Cl). It is clear that all five of these properties affected the character of the soil surface in a similar manner, and consequently produced some similarity in the spectral response patterns. All could be indicators of salinity. Selige et al. (2006) found that the best bands for their four properties (organic carbon, nitrogen, sand, and clay) did not coincide with any of the bands above, reinforcing the idea that separate models are needed for each field, as mentioned by Thomas-

son et al. (2001). Bajwa and Tian (2005), using PLSR on the first derivative, found that bands near our B27, B36, B37, and B40 were important in many of the models for the eight properties they tested, but they did not say which bands were best for each property.

Table 4 shows how the best combinations of bands can vary in the top 20 models for surface sand. Here one can see that the R² value declines faster in the lower rankings of the single-parameter models than in the multi-parameter models. The best four-parameter regression equation for each surface soil property is in shown in table 5. The factor in parentheses is the spectral index. Table 6 shows the similarity in the four-parameter spectral indices for EC, Ca, Mg, Na, and Cl. The average spectral index ($I = R26 - 0.855 * R27 - 0.241 * R37 - 0.400 * R59$) was determined by simply averaging the coefficients.

Figure 1 is a scatter diagram showing how well the percent sand in the soil surface is correlated to its four-parameter spectral index, from model 13 in table 5. As seen in figure 2, the scatter is also low around the regression line for percent silt in the surface soil versus the spectral index from model 14. The third best correlation is shown in figure 3, where the scatter for chlorides, model 7, is lower for Cl < 60 meq L⁻¹. Figure 4 shows a fairly good correlation for EC, from model 3. It is clear that there is less scatter in figures 1 and 2 than in figures 3 and 4, especially at the higher values of the soil property. Examples of spectral response patterns are shown in figures 5 and 6. The higher potassium levels tended to be associated with a somewhat higher reflectance over almost the entire spectrum. The slight undulation in the difference curve at about 750 nm might explain the appearance of band B27 in the three- and four-parameter models for potassium. For sand, there was a large difference in reflectance over the entire spectrum, with the higher sand levels causing higher reflectance, as noted by Palacios-Orueta and Ustin (1998). The greatest difference was in the range of 950-1000 nm, perhaps explaining why the best three-parameter models included bands B2, B4, and B7.

Four broader bands (blue, green, red, and NIR) were used in a multiple linear regression analysis of the 15 surface soil properties, and the results are shown in table 7. The exact wavelengths used for each band are given in the table caption. Sand was found again to be the easiest property to detect. Using red and NIR produced an R² of 0.77. Adding one or two

Table 4. Bands for the 20 highest R² values for percent sand in surface soil.

Rank	One Parameter		Two Parameters		Three Parameters		Four Parameters	
	Band	R ²	Bands	R ²	Bands	R ²	Bands	R ²
1	B13	0.7623	B2, B38	0.7865	B2, B4, B7	0.7965	B18, B30, B38, B40	0.8063
2	B3	0.7610	B2, B37	0.7842	B2, B3, B7	0.7933	B14, B27, B29, B38	0.8056
3	B31	0.7516	B2, B36	0.7840	B2, B7, B36	0.7932	B18, B30, B38, B39	0.8053
4	B14	0.7493	B2, B35	0.7827	B2, B7, B42	0.7931	B13, B27, B29, B38	0.8050
5	B12	0.7386	B2, B42	0.7826	B2, B7, B38	0.7928	B14, B26, B38, B60	0.8045
6	B2	0.7380	B2, B34	0.7823	B2, B7, B37	0.7927	B14, B29, B38, B60	0.8043
7	B1	0.7318	B1, B42	0.7823	B2, B7, B54	0.7927	B14, B30, B38, B60	0.8041
8	B18	0.7315	B1, B45	0.7817	B2, B7, B45	0.7926	B18, B30, B38, B41	0.8041
9	B10	0.7303	B2, B32	0.7816	B2, B7, B58	0.7926	B13, B23, B29, B38	0.8039
10	B22	0.7270	B1, B36	0.7815	B2, B7, B40	0.7925	B2, B25, B29, B38	0.8037
11	B4	0.7265	B2, B33	0.7815	B2, B7, B53	0.7923	B10, B25, B29, B38	0.8036
12	B32	0.7240	B1, B34	0.7813	B2, B7, B43	0.7923	B12, B25, B29, B38	0.8031
13	B15	0.7237	B2, B39	0.7813	B2, B7, B39	0.7922	B10, B24, B29, B38	0.8025
14	B11	0.7162	B2, B45	0.7812	B2, B7, B44	0.7921	B1, B25, B29, B38	0.8022
15	B25	0.7126	B1, B40	0.7811	B2, B7, B57	0.7921	B13, B30, B38, B60	0.8022
16	B16	0.7110	B1, B43	0.7810	B2, B7, B35	0.7920	B13, B25, B29, B38	0.8020
17	B26	0.7098	B1, B35	0.7809	B2, B7, B41	0.7920	B18, B20, B30, B38	0.8020
18	B19	0.7095	B1, B38	0.7808	B2, B7, B52	0.7918	B13, B25, B26, B38	0.8019
19	B24	0.7073	B2, B40	0.7808	B2, B7, B55	0.7917	B11, B25, B29, B38	0.8019
20	B17	0.7054	B2, B41	0.7808	B2, B4, B6	0.7917	B13, B27, B29, B37	0.8018

Table 5. Regression equations relating surface soil properties to best four-parameter model. The factor in parentheses is the spectral index (S = predicted value of soil property, and R = reflectance for indicated band).

Model No.	Soil Property	Regression Equation	R ²	RMSE
1	SP (%)	S = 78.29 - 425.0*(R4 - 4.136*R33 + 4.000*R36 - 0.601*R59)	0.4859	3.07
2	pH	S = 8.353 - 8.130*(R3 - 1.025*R7 + 2.137*R13 - 1.643*R26)	0.6164	0.076
3	EC (dS m ⁻¹)	S = -24.28 - 829.5*(R26 - 0.841*R27 - 0.253*R38 - 0.403*R59)	0.6696	1.96
4	Ca (meq L ⁻¹)	S = -126.5 - 3663*(R26 - 0.887*R27 - 0.321*R38 - 0.218*R60)	0.6188	9.51
5	Mg (meq L ⁻¹)	S = -48.18 - 1681*(R26 - 0.870*R27 - 0.304*R38 - 0.219*R60)	0.5820	4.32
6	Na (meq L ⁻¹)	S = -83.97 - 4190*(R26 - 0.822*R27 - 0.236*R38 - 0.364*R59)	0.6224	9.11
7	Cl (meq L ⁻¹)	S = -221.1 - 4744*(R26 - 0.889*R27 - 0.267*R37 - 0.445*R59)	0.7376	12.24
8	HCO ₃ (meq L ⁻¹)	S = 5.630 - 87.47*(R27 - 1.697*R30 + 2.587*R48 - 1.938*R53)	0.5668	0.345
9	C (%)	S = 1.377 + 15.77*(R32 - 2.156*R44 - 2.741*R47 + 3.986*R50)	0.2692	0.081
10	P (mg kg ⁻¹)	S = 96.24 - 1107*(R27 - 1.681*R30 + 5.017*R50 - 4.333*R53)	0.6975	4.32
11	K (mg kg ⁻¹)	S = 904.6 + 6390*(R14 - 2.711*R27 + 2.839*R30 - 1.899*R47)	0.6391	45.03
12	OM (%)	S = 2.083 + 18.17*(R26 - 0.600*R27 - 3.261*R47 + 2.521*R50)	0.4857	0.083
13	Sand (%)	S = -92.6 + 1530*(R18 - 1.152*R30 + 1.537*R38 - 0.940*R40)	0.8063	4.83
14	Silt (%)	S = 82.59 - 573.0*(R2 - 0.645*R6 - 1.570*R39 + 1.860*R45)	0.7518	3.34
15	Clay (%)	S = 90.25 - 698.7*(R13 - 0.796*R29 + 1.905*R36 - 1.730*R44)	0.6708	3.23

Table 6. Spectral indices that are similar for five surface soil properties (R = reflectance for indicated band).

Model No.	Rank	Soil Property	Spectral Index
16	2	EC (dS m ⁻¹)	R26 - 0.847*R27 - 0.253*R37 - 0.392*R59
17	2	Ca (meq L ⁻¹)	R26 - 0.866*R27 - 0.233*R37 - 0.403*R59
18	3	Mg (meq L ⁻¹)	R26 - 0.842*R27 - 0.220*R37 - 0.405*R59
19	2	Na (meq L ⁻¹)	R26 - 0.829*R27 - 0.232*R37 - 0.355*R59
7	1	Cl (meq L ⁻¹)	R26 - 0.889*R27 - 0.267*R37 - 0.445*R59
Average			R26 - 0.855*R27 - 0.241*R37 - 0.400*R59

bands did little to increase the R² value. By comparison, using narrow bands produced an R² of 0.81. Silt was the next easiest to detect with the broad bands, just as it was with the narrow bands, but again the R² was considerably lower with the broad bands. By comparing table 7 to table 2, one can see that for every soil property, the best four-parameter correlation was consistently better with the narrow bands than with these broad bands.

Surface sand was chosen to be used in a test and validation procedure. One out of every ten rows of data in the master processing file was randomly selected; they were then removed and placed in a separate validation file. The remaining

data were put in a test file. MLR for a three-parameter model was run on the test file, producing a best fit with bands B2, B4, and B7. The results are shown in figure 7, with R² = 0.7942 and RMSE = 4.94%. The regression equation for the test data shown in figure 7 is also shown in figure 8 along with the data from the validation file. The validation data fit the test equation quite nicely, with RMSE = 5.07%.

The in-field distribution of sand in the surface soil for field 6-4 is shown in figure 9. The equation in model 13 (table 5) was used in an Access query to calculate the level of sand at every pixel. With this image loaded into ArcView, the level of sand at any point and the average for any area can easily

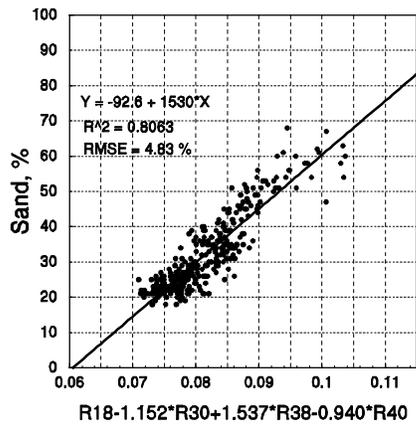


Figure 1. Scatter diagram and regression equation for percent surface sand as a function of the spectral index for model 13; data for fields 6-4 and 4-1 from flight of 22 May 2002.

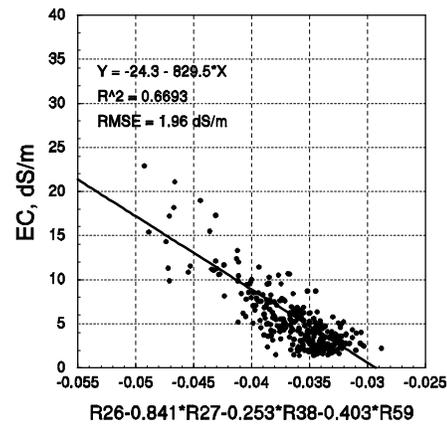


Figure 4. Scatter diagram and regression equation for surface EC as a function of the spectral index for model 3; data for fields 6-4 and 4-1 from flight of 22 May 2002.

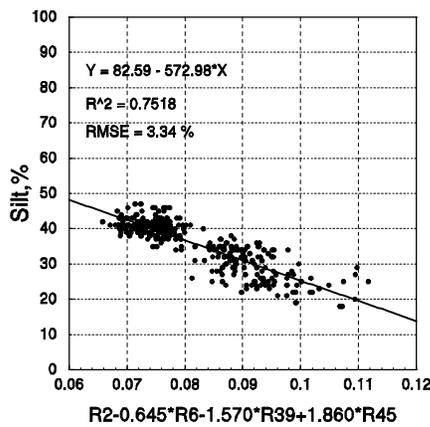


Figure 2. Scatter diagram and regression equation for percent surface silt as a function of the spectral index for model 14; data for fields 6-4 and 4-1 from flight of 22 May 2002.

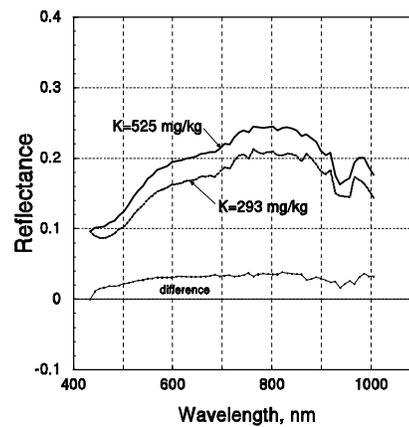


Figure 5. Spectral response patterns for high and low potassium levels.

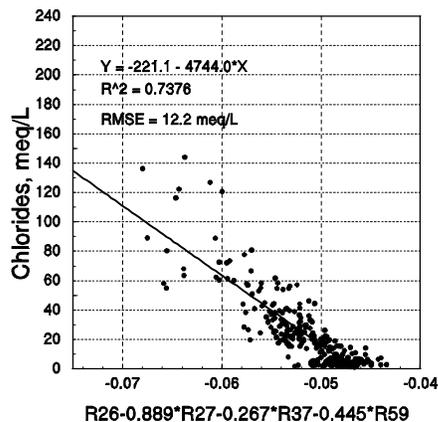


Figure 3. Scatter diagram and regression equation for surface chlorides as a function of the spectral index for model 7; data for fields 6-4 and 4-1 from flight of 22 May 2002.

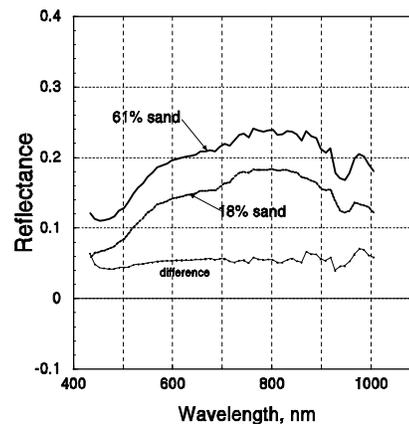


Figure 6. Spectral response patterns for high and low sand levels.

be found. The scatter around the regression line in figure 1 indicates the degree of accuracy for these values. In most cases, the predicted value shown in the image in figure 9 will have an error of less than 4.8% sand (the RMSE), and the error will seldom exceed 9.6% sand (twice the RMSE). This type of image becomes one of the basic maps needed in many precision agriculture applications.

The use of narrow bands with a spectral resolution of 10 nm may have been a key element in this study. Thomasson et al. (2001) suggested that perhaps stronger correlations could be obtained with a spectral resolution less than 50 nm. Although some specific broad bands might work well, the ones selected for blue, green, red, and NIR in this study did not correlate as well as narrow bands to the soil properties. In comparing our results to some similar work, i.e., for those studies using remote sensing flights over bare soil as in Selige

Table 7. Bands with the highest R² values for surface properties of soil, using the broad bands blue, green, red, and NIR. Band width for blue: 468-516 nm; for green: 535-565 nm; for red: 652-690 nm; and for NIR: 826-875 nm.

Soil Property	One Parameter		Two Parameters		Three Parameters		Four Parameters	
	Band	R ²	Bands	R ²	Bands	R ²	Bands	R ²
SP (%)	NIR	0.3549	Red, NIR	0.3708	Green, red, NIR	0.3726	Blue, green, red, NIR	0.3845
pH	green	0.5242	Blue, NIR	0.557	Blue, green, NIR	0.5609	Blue, green, red, NIR	0.5619
EC (dS m ⁻¹)	green	0.4544	Blue, NIR	0.4611	Blue, red, NIR	0.4631	Blue, green, red, NIR	0.4662
Ca (meq L ⁻¹)	green	0.4265	Blue, NIR	0.4361	Blue, red, NIR	0.4371	Blue, green, red, NIR	0.4406
Mg (meq L ⁻¹)	blue	0.3688	Blue, NIR	0.376	Blue, green, NIR	0.378	Blue, green, red, NIR	0.3842
Na (meq L ⁻¹)	green	0.333	Blue, NIR	0.3355	Blue, red, NIR	0.3366	Blue, green, red, NIR	0.3369
Cl (meq L ⁻¹)	green	0.5984	Blue, NIR	0.6065	Blue, red, NIR	0.6075	Blue, green, red, NIR	0.6081
HCO ₃ (meq L ⁻¹)	NIR	0.2956	Green, red	0.3404	Blue, green, red	0.4781	Blue, green, red, NIR	0.4789
C (%)	red	0.0162	Blue, red	0.1088	Blue, green, red	0.1089	Blue, green, red, NIR	0.1089
P (mg kg ⁻¹)	NIR	0.4378	Green, red	0.4657	Blue, green, red	0.6334	Blue, green, red, NIR	0.6402
K (mg kg ⁻¹)	NIR	0.3212	Blue, green	0.3714	Blue, green, red	0.4531	Blue, green, red, NIR	0.4539
OM (%)	NIR	0.2228	Blue, green	0.3054	Blue, green, red	0.3111	Blue, green, red, NIR	0.3189
Sand (%)	NIR	0.7401	Red, NIR	0.7724	Blue, red, NIR	0.7728	Blue, green, red, NIR	0.7748
Silt (%)	NIR	0.6787	Green, NIR	0.6836	Blue, green, NIR	0.6939	Blue, green, red, NIR	0.711
Clay (%)	red	0.5559	Red, NIR	0.5782	Green, red, NIR	0.5872	Blue, green, red, NIR	0.6659

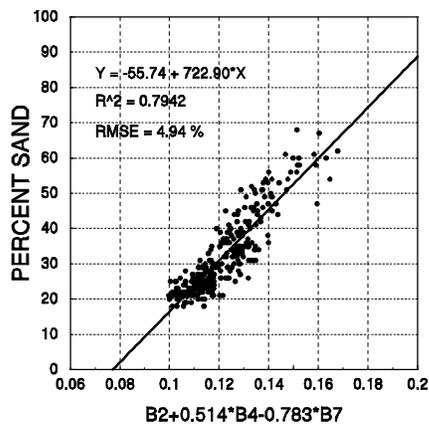


Figure 7. Scatter diagram and best-fitting three-parameter regression equation for test data, using surface sand; data for fields 6-4 and 4-1 from flight of 22 May 2002.

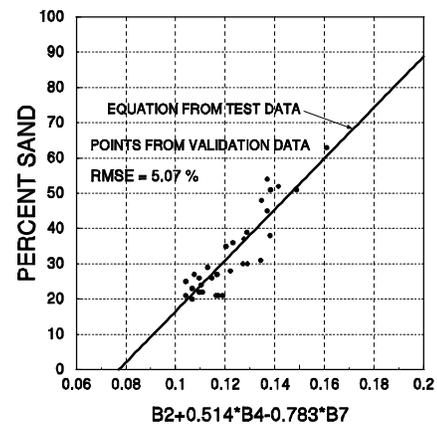


Figure 8. Scatter diagram for validation data, with regression line from test data; data for surface sand in fields 6-4 and 4-1 from flight of 22 May 2002.

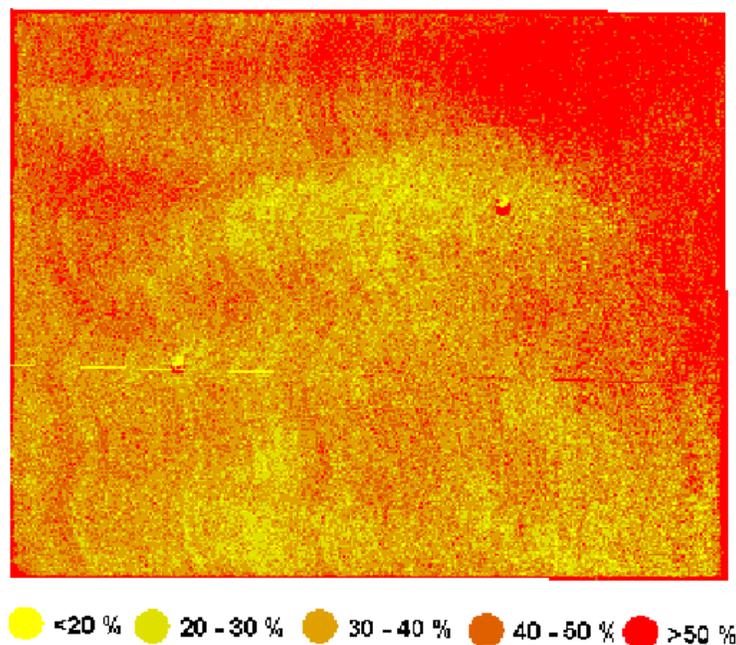


Figure 9. High-definition image of field 6-4 on 22 May 2002, showing the distribution of surface sand with a spatial resolution of 1.2 m. Equation for each pixel is from model 13.

et al. (2006) and Hong et al. (2002), we could find none that regressed every possible combination of all the HSI bands available using MLR. Likewise, none grid-sampled the soil at two different depths, and none looked at more than seven properties. Goel et al. (2003), looking at several biophysical properties of corn, used multiple linear regression for narrow band selection and presented the resulting regression equations, as we have done here. They were concerned that collinearity or codependence of the many band reflectances could be a problem, as suggested by Longley (1967) and Beaton et al. (1976). However, with the goal of prediction rather than explanation, collinearity is less of a problem (Yu, 2000). CoPlot has a diagnostic tool that automatically checks for collinearity using a procedure from Maindonald (1984). Band selection using multiple linear regression of narrow bands was also used successfully by DeTar et al. (2006). Multiple linear regression is a simple and straightforward method for obtaining optimum bands, and the resulting spectral indices worked well for several soil properties. The exact model for this and other fields depends greatly on the type of tillage operation applied just prior to the flight. According to Jensen (2000), some soil properties may be measured remotely under ideal conditions, and it is possible for remotely sensed data to be of value, and in some cases essential, to accurate soil mapping.

CONCLUSIONS

The primary result of this study is that airborne remote sensing with a hyperspectral camera over nearly bare soil can measure the within-field distribution of several soil properties with a spatial resolution of 1.2 m. Using multiple linear regression, every possible combination of 60 narrow bands was regressed to find the very best one-, two-, three-, and four-parameter models for 15 different soil properties. The percent sand in the surface samples was estimated with a reasonable degree of accuracy with $R^2 = 0.806$ for a four-parameter model; the best combination of wavelengths was 627, 647, 724, and 840 nm. For the silt, clay, EC, chlorides, and phosphorus, the results were somewhat less satisfactory with a range of $0.66 < R^2 < 0.76$. The least detectable of the surface soil properties was carbon with $R^2 = 0.27$. Organic matter and saturation percentage had $R^2 < 0.49$. The other six properties (pH, Ca, Mg, Na, HCO_3 , and K) were all in the intermediate range. For samples taken with cores from the 0 to 300 mm depth, the textural properties were also estimated accurately with $R^2 > 0.71$; however, the non-textural properties of these subsurface soils had a much poorer correlation to reflectance data than did the surface soils. Regression equations are presented for the best models for each soil property so that values can be calculated at every pixel on the image map. New spectral indices were developed; one index ($I = R_{763} - 0.85 * R_{753} - 0.24 * R_{657} - 0.40 * R_{443}$) was found to work well with five of the soil properties (EC, Ca, Mg, Na, and Cl), indicating some commonality in the manner in which they affected the reflectance from the soil surface, possibly due to a salinity effect. Using combinations of selected narrow bands produced higher R^2 values than did combinations of the broad bands of blue, green, red, and NIR. The information presented here should be useful in selecting filters for multi-spectral cameras and for selecting wavebands for HSI cameras when attempting to map soil properties with remote sensing.

ACKNOWLEDGEMENTS

Special thanks go to the reviewers who contributed many good ideas to the manuscript.

REFERENCES

- Bajwa, S. G., and L. F. Tian. 2005. Soil fertility characterization in agricultural fields using hyperspectral remote sensing. *Trans. ASABE* 48(6): 2399-2406.
- Barnes, E. M., and M. G. Baker. 1999. Multispectral data for soil mapping: Possibilities and limitations. ASAE Paper No. 991138. St. Joseph, Mich.: ASAE.
- Barnes, E. M., M. S. Moran, P. J. Pinter, Jr., and T. R. Clarke. 1996. Multispectral remote sensing and site-specific agriculture. Examples of current technology and future possibilities. In *Proc. 3rd Intl. Conf. Precision Agric.*, 845-854. Madison, Wisc.: ASA-CSA-SSSA.
- Beaton, A. E., D. B. Rubin, and J. L. Barone. 1976. The acceptability of regression solutions: Another look at computational accuracy. *J. American Stat. Assoc.* 71(353): 158-168.
- Chen, F., D. E. Kissel, L. T. West, and W. Adkins. 2000. Field-scale mapping of soil organic carbon using remotely sensed imagery. *SSSA J.* 64(2): 746-753.
- DeTar, W. R., J. V. Penner, and H. A. Funk. 2006. Airborne remote sensing to detect plant water stress in full canopy cotton. *Trans. ASABE* 49(3): 655-665.
- Goel, P. K., S. O. Prasher, J. A. Landry, R. M. Patel, A. A. Viau, and J. R. Miller. 2003. Estimation of crop biophysical parameters through airborne and field hyperspectral remote sensing. *Trans. ASAE* 46(4): 1235-1246.
- Haboudane, D., J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze. 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* 81(3): 416-426.
- Hong, S. Y., K. A. Sudduth, N. R. Kitchen, S. T. Drummond, H. L. Palm, and W. J. Wiebold. 2002. Estimating within-field variations in soil properties from airborne hyperspectral images. In *Proc. Pecora 15/Land Satellite Information IV Conf*, CD-ROM. Bethesda, Md.: American Society for Photogrammetry and Remote Sensing.
- Jensen, J. R. 2000. *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River, N.J.: Prentice Hall.
- Lillesand, T. M., and R. W. Kiefer. 1999. *Remote Sensing and Image Interpretation*. 4th ed. New York, N.Y.: John Wiley and Sons.
- Longley, J. W. 1967. An appraisal of least squares procedures for the electronic computer from the point of view of the user. *J. American Stat. Assoc.* 62(319): 819-841.
- Maindonald, J. H. 1984. *Statistical Computation*. New York, N.Y.: John Wiley and Sons.
- Palacios-Orueta, A., and S. L. Ustin. 1998. Remote sensing of soil properties in the Santa Monica Mountains: I. Spectral analysis. *Remote Sens. Environ.* 65(2): 170-183.
- Selige, T., L. Natscher, and U. Schmidhalter. 2003. Spatial detection of topsoil properties using hyperspectral sensing. In *Precision Agriculture*, 633-638. J. Stafford and A. Werner, eds. Wageningen, The Netherlands: Wageningen Academic.
- Selige, T., J. Bohner, and U. Schmidhalter. 2006. High-resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* 136(1-2): 235-244.
- Thomasson, J. A., R. Sui, M. S. Cox, and A. Al-Rajehy. 2001. Soil reflectance for determining soil properties in precision agriculture. *Trans. ASAE* 44(6): 1445-1453.
- USDA. 1978. Soil survey of Kings County, California. Washington, D.C.: USDA Soil Conservation Service.
- Yu, C. H. 2000. An overview of remedial tools for collinearity in SAS. In *Proc. 2000 Western Users of SAS Software Conference*, 196-201. Scottsdale, Ariz.: Western Users of SAS Software, Inc.