



Contents lists available at ScienceDirect

Geoderma

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

Inductively mapping expert-derived soil-landscape units within dambo wetland catenae using multispectral and topographic data

Matthew K. Hansen^a, David J. Brown^b, Philip E. Dennison^{a,*}, Scott A. Graves^a, Ross S. Brickleyer^b

^a Department of Geography, University of Utah, 260 S. Central Campus Dr., Room 270, Salt Lake City, UT 84112, USA

^b Department of Crop and Soil Sciences, Washington State University, PO Box 646420, Pullman, WA 99164, USA

ARTICLE INFO

Article history:

Received 21 July 2008

Received in revised form 12 January 2009

Accepted 14 January 2009

Available online 15 February 2009

Keywords:

Wetlands

Digital soil mapping

Africa

Uganda

Remote sensing

SPOT

SRTM

ABSTRACT

Constructing a cost-effective and detailed digital soil map of Africa will require the extensive utilization of both legacy soil data and legacy soil-landscape knowledge – which in Africa is primarily available from reconnaissance-scale catena or association maps and related studies. We evaluated a hybrid approach for disaggregating reconnaissance scale soil maps: rapid and inexpensive delineation of representative soil-landscape units in the field using expert knowledge, followed by the use of inductive, empirical, and correlative modeling techniques to map these landscape units. Our 2214 km² study area, located in central Uganda, consisted predominantly of catenae that terminate in seasonal valley floor wetlands called dambos – a type of landscape that can be found throughout the African continent. For model training and validation, we identified four landscape classes in the field using published expert knowledge: well-drained *uplands* (red soils); sloping dambo wetland *margins* (gradient > 2%), frequently inundated dambo bottoms (hummocky microtopography), and flat dambo floors (default). Using binary decision trees (BDT) with multispectral and topographic remote sensing covariates, we created a 20 m resolution map of these four classes. Multispectral inputs included reflectance values, vegetation indices, and spectral mixture modeling fractions from Système Pour l'Observation de la Terre (SPOT) 4 satellite images acquired in December, 2006 and February, 2007. Topographic inputs consisted of a digital elevation model (DEM) from Shuttle Radar Topography Mission (SRTM) data, slope, and 20 relative elevation layers calculated using moving windows of various sizes. Decision rules were based upon the following input variables: the Normalized Difference Vegetation Index (NDVI), the Normalized Difference Infrared Index (NDII), the shortwave infrared (SWIR) reflectance, northing, slope, and two relative elevation layers. The overall classification accuracy of 75.5% and Kappa coefficient of 0.67 suggest that a combination of multispectral, topographic, and spatial data may be used to reliably classify landscape classes for dambo-terminated catenae. At 59% of the 2214 km² study area, the upland class was by far the most abundant, with margins at 21%, floors at 12% and bottoms at 8%. A statistical analysis of soil property data from a small catchment located within the study area showed significant class differences in soil texture, color, organic carbon (SOC), base saturation, pH, effective cation exchange capacity (ECEC), and clay mineralogy. Though detailed soil maps are rare in Africa, reconnaissance soil maps can be inexpensively disaggregated to provide a valuable starting point for digital soil mapping.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

There is growing interest in constructing a detailed (D3, 20–200 m pixel) digital soil map of the world (McBratney et al., 2003; Minasny et al., 2008) to support improved assessments of fertility, ecosystem services and potential soil degradation (Carre et al., 2007; Palm et al., 2007); scientifically based land management in both remote areas and densely populated regions (Behrens and Scholten, 2006; MacMillan et al., 2007); and process-based environmental modeling (Lin et al., 2006). The digital soil mapping paradigm (McBratney et al., 2003)

relies heavily upon modeling relationships between soil properties and easily measured, spatially exhaustive environmental covariates (e.g. surface reflectance and digital elevation values acquired from satellites). Though a well studied family of covariates and relationships has proven useful globally, digital soil mapping usually requires additional soil data for calibration in each new region to be mapped (McBratney et al., 2003).

The African continent is simultaneously the most soil data-challenged land surface in the world and the area most in need of improved soil information (Eswaran et al., 1997; Palm et al., 2007; Rossiter, 2008). However, purely empirical, inductive, environmental correlation-based, predictive soil mapping at a D3 to D4 level (20 m to 2 km pixel size) appears to require well-distributed profile characterization at a density of at least 75 km²/profile (Dobos et al., 2000;

* Corresponding author.

E-mail address: dennison@geog.utah.edu (P.E. Dennison).

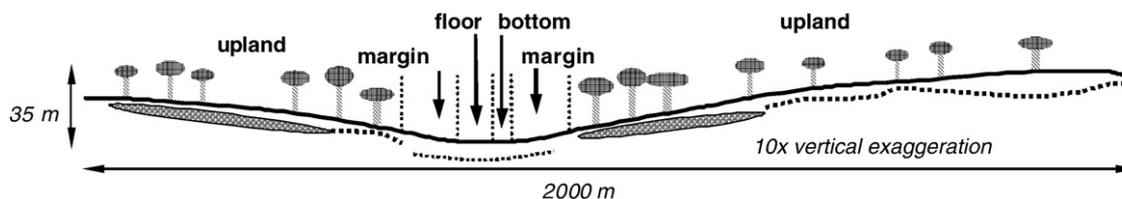


Fig. 1. A relatively narrow, incised, dambo cross-section from the middle of the study area, showing trees, ferricrete and stonelines largely confined to upland areas, constructed using elevation data from Brown et al. (2004a,b).

McBratney et al., 2003), with most published regional studies reporting significantly higher sampling densities (McKenzie and Ryan, 1999; Hengl et al., 2004; Henderson et al., 2005). Even if sampling were focused only on the 13.6 million km² of potentially arable land in Africa as opposed to the total 30.3 million km² of continental land area (Eswaran et al., 1997), a minimum of 182,000 characterized profiles would be required to inductively map the soils of Africa assuming a 75 km²/profile sampling density. Though proximal soil sensing techniques can potentially reduce characterization costs (Shepherd and Walsh, 2002), the acquisition and characterization of soil samples from >10⁵ representative soil profiles distributed throughout the African continent would likely prove cost prohibitive.

Legacy “expert” soil-landscape knowledge will by necessity play a central role in the construction of a high-resolution, D3 digital soil map of Africa. Where higher order surveys are available, “data renewal” techniques can be used to improve the spatial resolution and accuracy of older maps (Rossiter, 2008), and knowledge extraction approaches can be employed to quantify and extend implicit soil-landscape correlations to surrounding areas with similar soils and environmental conditions (Bui et al., 1999; Bui and Moran, 2001; Bui and Moran, 2003). However, for most of Africa only reconnaissance maps are available with soils mapped as catenae¹ or associations (Eswaran et al., 1997; Rossiter, 2008). Utilizing legacy soil profile characterization data for these areas will require disaggregation of landscapes into previously identified landscape units (Bui and Moran, 2001). Research publications containing soil catena landscape relationships, profile descriptions and characterization data are also available for many parts of Africa (Morison, 1948; Morison et al., 1948; Nye, 1954; Nye, 1955a,b; Ollier, 1959; Radwanski and Ollier, 1959; Watson, 1964a,b,c; Webster, 1965). Though these publications generally focus more on soil profile characterization and genesis than soil-landscape relationships, a careful reading of these papers could yield the rules required for expert-knowledge soil mapping techniques (McBratney et al., 2003).

While purely inductive, empirical, environmental correlation methods require a large number of characterized soil profiles to develop calibrations, expert-knowledge systems depend on the quality of the expert or expertise employed. Our objective in this study was to assess the viability of a hybrid approach: (1) rapidly and inexpensively delineate representative soil-landscape units in the field using expert knowledge; then (2) employ inductive, empirical, and correlative modeling techniques to map these soil-landscape units. We selected a dambo-terminated catena region in Uganda for this study as dambo catenae are well described in the literature, occupy a significant fraction of the African land surface, and contain seasonal wetlands important for both global biogeochemistry (Bartlett and Harriss, 1993) and rural livelihoods (Scoones, 1991). Both binary decision trees and random forest machine learning were employed to calibrate and validate an empirical landscape classification model. Finally, we modeled soil properties as a function of expert-

derived landscape class for a small 2nd-order watershed within the study area to provide an initial evaluation of soil-class correlations.

2. Background

2.1. Dambo-terminated catenae

Dambos—also termed mbugas, vleis, and fadamas—are seasonally saturated, grassy, channelless, gently sloping valley floors (Acres et al., 1985) that commonly occupy the lowest topographic positions in African catenae or “land systems” (Trapnell and Clothier, 1937; Trapnell, 1943; Morison et al., 1948; Radwanski and Ollier, 1959; Watson, 1964c; Webster, 1965; Ollier et al., 1969). Dambos are commonly subdivided into three basic soil-vegetation-hydrology-topography units, following the terminology of Acres et al. (1985): margins, floors and bottoms (Fig. 1). These can be considered roughly analogous to footslope, toeslope and channel terms commonly employed in geomorphology (Ruhe, 1961; Conacher and Dalrymple, 1977). Margins are generally narrow, sloping, sandy, transitional elements with sparse vegetation cover. Valley floors have barely perceptible elevation gradients, clayey grey subsoils, and an increase in vegetation density relative to margins. Black-clay dambo bottoms occupy the lowest cross-sectional position, are often inundated, and host dense grasses and sedges growing on well-defined hummocks (Mäckel, 1974; Acres et al., 1985; Mäckel, 1985).

Assuming areal densities of 5% and 2% (Bullock, 1992), respectively, for the map of “main” and “sporadic” areas of dambo occurrence in Africa published by Acres et al. (1985), a first order estimate of dambo extent would be approximately 300,000 km². Since dambos occupy at most 20% of these catenary landscapes, this would imply a total area of 1.5 million km² of dambo-terminated catenae, or approximately 5% of the African land surface and 11% of the arable land in Africa. This estimate may be conservative as published estimates of valley floor wetland extent are as high as 1.35 million km² (Frenken and Mharapara, 2002), implying that associated catenae could cover as much as 20% of the African land surface.

2.2. Wetland delineation techniques

Three prevalent techniques currently used for wetland identification on a landscape scale are field investigation, aerial photo interpretation, and multispectral image analysis (Baker et al., 2006). Automated classification using high-resolution satellite imagery offers the best wetland mapping alternative in many situations. In fact, in comparison to human interpretation of aerial photos, multispectral image classification of wetlands has demonstrated similar accuracy and greater repeatability (Harvey and Hill, 2001; Baker et al., 2006). However, relying solely on the vegetation cues provided from satellite imagery to identify seasonally saturated wetlands is difficult and potentially inaccurate (Tiner, 1993). This is because the transition from wetland to non-wetland plant species is gradual, making it challenging to distinguish wetland extents, much less distinct elements within wetlands.

Recent land cover classification studies have shown the benefits of combining multiple types of input data (Pohl and Van Genderen, 1998;

¹ The term “catena” is used in Africa much as “association” is used elsewhere with the notable difference that the catena term implies a specific set of landscape processes governing soil genesis and soil-topography relationships (Brown, 2005).

McBratney et al., 2003; Varma et al., 2003; Zhu and Tateishi, 2006). For example, Amarsaikhan and Douglas (2004) saw significant improvements in distinguishing land cover types in Mongolia by combining SPOT-XS satellite and ERS-SAR (Synthetic Aperture Radar) imagery. Similarly, Wright and Gallant (2007) increased palustrine wetland classification accuracy in Yellowstone National Park by combining multispectral Landsat Thematic Mapper (TM) data with terrain and other ancillary inputs. Scull et al. (2005a) combined soil samples with vegetation and terrain metrics derived from Landsat TM imagery and a digital elevation model (DEM) to produce predictive soil maps.

Among the data mining tools that have been used effectively for land system classification from multiple inputs is the binary decision tree (BDT). A BDT consists of a system of nodes and connections analogous to the hierarchical branch pattern of a tree (Pal and Mather, 2003). At the “root” of a BDT are the input data, which are passed through a series of decision rules and categorized into specific classes at the terminal nodes. BDT classifiers generate decision rules by passing a set of training data through a series of recursive yes/no tests. The binary splits that result in more homogeneous subsets of the training data become the nodes of the BDT (Pal and Mather, 2003). These nodes also identify which of the input variables were selected to distinguish each of the classes (Simard et al., 2000). BDTs are considered superior to many conventional statistical classification techniques because they are nonparametric and make fewer assumptions about the data (Baker et al., 2006). BDTs are also advantageous because they are simpler to train and interpret than most neural network classifiers (Pal and Mather, 2003).

Numerous remote sensing studies have demonstrated the effective use of BDTs for land cover classification. For example, Simard et al. (2002) used a BDT to combine data from multiple SAR bands. These combined data improved the overall classification accuracy of tropical

vegetation by 18% in comparison to a single band classification. Work by Scull et al. (2005b) showed that BDT analysis can be applied to predictive soil mapping. Xu et al. (2005) explored the potential for decision tree regression as a classification technique focusing on mixed pixels. More recently, Lowry et al. (2007) employed a decision tree classifier in a five-state biodiversity assessment because of its past successes in previous moderate-scale classifications. Although much of the prior BDT classification research focused on continental or even global extents, Brown de Colstoun et al. (2003) demonstrated with their 270 km² study area that BDTs can also effectively classify high-resolution imagery of smaller areas. The land cover classifications of study areas in Yellowstone National Park and Mongolia mentioned previously also made use of BDT classifiers (Amarsaikhan and Douglas, 2004; Wright and Gallant, 2007).

3. Methods

3.1. Study area

The study area, comprising of 2214 km², was selected from a relatively undisturbed region of the central Ugandan plateau underlain by Pre-Cambrian granitic gneiss which drains into the perennially swampy Mayanja, Towa, and Lugogo Rivers (Fig. 2). This area is representative of the geomorphology, climate, and vegetation regimes associated with African dambos (Mäckel, 1974; Acres et al., 1985; Mäckel, 1985; von der Heyden, 2004). The contemporary climate in this area is wet tropical with a mean annual precipitation of ~120 cm (distinctly bimodal distribution), and a mean annual temperature of 23 °C at an elevation over 1 km above sea level (Survey Department, 1967). Annual precipitation declines significantly from south to north, though precise measurements are not available. Additionally, the



Fig. 2. The location of the study area in Uganda.

study area offers a range of topographic relief with maximum slopes of 7% in the south declining to 2% in the far north.

Upland areas are covered with a tall grass–bush–bare soil mosaic classified as “dry *combretum* savanna” (Langdale-Brown et al., 1964, p.59–63). Upland woody vegetation becomes denser at the upland–margin interface, presumably due to greater water availability. In contrast, dambos are open and grassy with woody vegetation concentrated in dense thickets on large and widely spaced termite mound “islands.” Major dambo bottom species include the sedges *Cyperus alba* and *Cyperus denudatus*, wetland grasses *Commelina subulata* and *Setaria sphacelata*, and forbs *Dyschoriste magchena* (snake herb) and *Emilia javanica* (Tassel flower). Similar species are found in dambo floors, though perhaps due to grazing they form a low turf as opposed to the tall and thick tussocks found in the bottoms. Moving up into the margins, the forb *Murdannia simplex* and grasses *Paspalum scrobiculatum* and *Hyparrhenia filipendula* become more prevalent. Major dambo species (almost all perennials) are found in all three landscape classes, though in different proportions. Moving into the upland areas, the *C. subulata* and drought-tolerant *Brachiaria brizantha* grasses become more prevalent, though even here wetland species were sometimes observed. The primary land use is non-intensive grazing for which dambo grasses are burnt annually.

3.2. Expert-derived soil-landscape units

Previous survey work from this region (Radwanski, 1959) and published research on a small 2nd-order catchment within the study area (Brown, 2007; Brown et al., 2004a) shows relatively little variability in well-drained upland soil mineralogy, chemistry, texture and soil organic matter. When this area was previously surveyed, upland associates of the dominant Buruli and Buyaga catenae were distinguished primarily on the basis of stone line and ferricrete depth (Radwanski, 1959). However, subsequent research suggests that these features are artifacts from a past episode of landscape genesis and might therefore be difficult to map using information derived from the contemporary land surface (Brown et al., 2003; Brown et al., 2004b). We therefore mapped uplands as a single soil-landscape class, defined by high-chroma, reddish hue subsoils (see Table 1 for details) with the quantitative limits derived from previous soil reflectance measurements in this study area (Brown et al., 2004a; Brown, 2007). Radwanski (1959) described all upland soils as having a subsoil hue of 5YR or redder, but given that only exemplars are reported for that survey, our rules were consistent with previous survey results. In the field, uplands were distinguished by reddish subsurface soils as observed (without digging) by comparing termite mounds to Munsell® color charts, with hue and chroma satisfying either of two sets of criteria provided in Table 1. While upland vegetation generally consisted of a bush–savannah mosaic with abundant bare soil, the woody vegetation often extended down into margins making this a poor indicator for upland delineation.

Radwanski (1959) mapped dambos as a single associate in the catena, and gave these wetland areas relatively little consideration. Since that survey was completed, a consensus has emerged in the literature regarding the distinction of the *margin*, *floor* and *bottom* landscape units within dambos (Mäckel, 1974; Acres et al., 1985; Mäckel, 1985; von der Heyden, 2004). Based upon the published

literature both throughout Africa (cited above) and from Uganda specifically (Brown et al., 2006), we formally defined the *margin* class as having a slope >2% without meeting the subsoil color requirements of the *upland* class – as measured using handheld inclinometers in the field. We also considered using the density of grass cover (thinner for margins) as a distinguishing characteristic but this was harder to measure quickly and reliably in the field. The *bottom* class was defined by a characteristic hummocky microtopography that could be easily observed in the field. Again, vegetation density could have been employed as dambo bottoms generally contain thick, 1-m high grasses and sedges vs. the lawn-like vegetation of dambo floors. However, due to regular dambo burning, this was not always a reliable indicator. The default *floor* class captured relatively flat wetland areas that lacked hummocks. Though landscape classes were clearly and unambiguously defined, transitions between classes were often diffuse – making it sometimes difficult to map class boundaries with meter-scale precision.

To evaluate the explanatory power of these expert landscape units, we applied these criteria to 193 soil profiles previously sampled and characterized from a small 2nd-order catchment within the study area (Brown, 2007; Brown et al., 2004a,b). Samples from this catchment were previously crushed, sieved (<2 mm), and air-dried and scanned with an Analytical Spectral Devices (Boulder, CO, USA) Fieldspec Pro® spectroradiometer to obtain precise color measurements (Brown et al., 2004a,b). These spectral data were subsequently employed to characterize soil clay mineralogy (Brown, 2007) using a global calibration previously evaluated for this study area (Brown et al., 2006). Soil organic carbon (SOC) was determined using the dry combustion method and soil texture was measured with both the pipette and hydrometer methods (Gee and Bauder, 1986). Soil pH (in H₂O), Effective Cation Exchange Capacity (CEC) and base saturation were measured using standard laboratory techniques as reported in Brown et al. (2003). Using soil data from 0–10, 20–30, 50–60, 90–100, 140–150 and 200–210 cm, we employed generalized least squares regression within the R *nlme* package (R Development Core Team, 2008) to model soil properties by depth as a function of landscape class (bottom, floor, margin and upland encoded as a single factor variable). Spatial correlation models were fit to the residuals for each property–depth model as required and overall significance was extracted using analysis of variance (ANOVA).

3.3. Field observation design

Field measurements were taken during January and February, 2007, with the intention of observing the vegetation variability that typically occurs during one of the two annual dry seasons in the study area. To ensure sampling at a variety of scales, stream order values were used to create a stratified random sample set. Some of the randomly determined sites were inaccessible in the field; however, the sample set did include observations from all stream orders within the drainage system. Two sampling techniques were employed in the field. First, polygons were used to delineate areas with relatively homogenous vegetation cover, topographic relief, and soil properties, indicating that they belonged to single dambo class. These homogenous polygons were typically captured using a GPS, although large or irregularly shaped areas were occasionally outlined on a paper map for subsequent digitization. Second, to sample the variability within each dambo class and particularly in the transition areas between classes, cross-section transects with a width of 120 m were sampled perpendicular to the hydrologic flow. Transect length varied according to drainage size at the sample location. Transects were sampled by using a GPS to record the estimated points of transition between the dambo classes. Buffers with a width of 60 m on each side of the centerline were then applied to the linear transects to create rectangular dambo class areas. Because the transect segments were not perfectly aligned, the 60-m buffers sometimes resulted in

Table 1
Dambo class field identification criteria

Dambo class	Identification criteria
Upland	Subsoil chroma ≥ 4 and hue ≤ 7.5 YR, OR chroma ≥ 3 and hue ≤ 5 YR
Margin	Slope > 2% and not upland
Floor	Not upland, margin, or bottom
Bottom	Hummocky microtopography

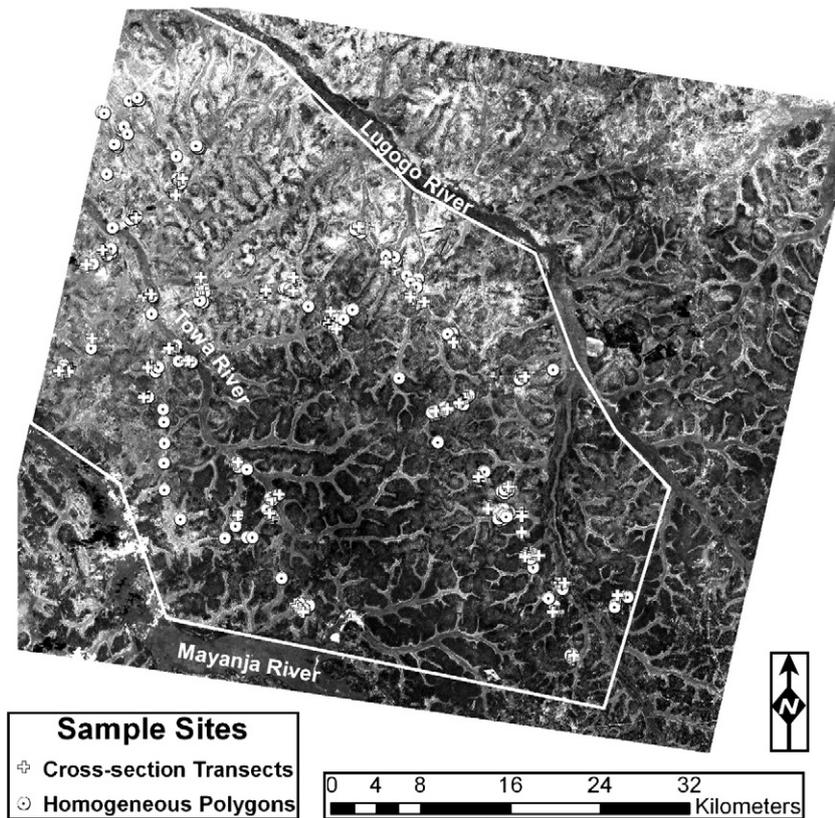


Fig. 3. Sample sites overlaid on the February 21, 2007, SPOT 4 image of study area (includes material © CNES 2007, Distribution Spot Image S.A., France SICORP, USA, all rights reserved).

overlapping dambo classes. The dambo class areas were shortened on each end by 20 m (equivalent to 1 pixel of the highest resolution remote sensing data used) to eliminate areas of overlap. Points were sampled only at the transition from one dambo class to another; therefore, the beginning and end of each transect were typically located at the upland/margin interfaces. Upland areas, which are usually recognizable in satellite imagery by their woody vegetation, were extended from these endpoints for use in the training and validation sets. A total of 190 homogeneous dambo class polygons and 46 cross-section transects were sampled in the field (Fig. 3). In addition to the field samples, several large polygons clearly identified as uplands in multispectral imagery used for this study (discussed below) were added to the sample data. As dambo bottoms provided such a small percentage of the total cross-section transect area, the multispectral images were also used to add several bottom sites during the classification training process.

3.4. Modeling

A set of 66 variables, mostly derived from remotely sensed data, served as inputs to the landscape classification (Table 2). The majority of these input variables were measures of topographic relief or vegetation cover. While most of these variables were not ultimately used for the classification, it was necessary to test each of them comparatively to ascertain which variables contributed to an accurate classification of the study area.

Two SPOT 4 multispectral satellite images of the study area, atmospherically corrected to apparent surface reflectance using ACORN (ImSpec LLC), provided the primary inputs for the classification (Fig. 3). The use of these images also established a common spatial resolution of 20 m, to which other inputs were resampled as necessary using bilinear interpolation. The SPOT scenes were captured December 10, 2006, and February 21, 2007, dates that roughly

coincided with the beginning and end of the field campaign, as well as the local dry season. While this period was chosen with the intent of monitoring dambo vegetation phenology as moisture decreased

Table 2
Classification input variables subdivided according to source

Multispectral inputs ^a	Topographic inputs	Spatial inputs
SPOT 10 Dec 06	DEM	UTM northing ^b
Green (band 1) ρ	Slope ^b	UTM easting
Red (band 2) ρ	Relative elevation 11	
Near infrared (band 3) ρ	Relative elevation 21 ^b	
Shortwave infrared (band 4) ρ	Relative elevation 31	
SPOT 21 Feb 07	Relative elevation 41	
Green ρ	Relative elevation 51	
Red ρ	Relative elevation 61	
Near infrared ρ	Relative elevation 71	
Shortwave infrared ρ^b	Relative elevation 81	
Vegetation indices	Relative elevation 91	
Dec. NDVI	Relative elevation 101	
Feb. NDVI	Relative elevation 111	
NDVI difference ^b	Relative elevation 121	
Dec. NDII	Relative elevation 131	
Feb. NDII ^b	Relative elevation 141	
NDII difference	Relative elevation 151	
Spectral mixture analysis	Relative elevation 161	
Dec. PV	Relative elevation 171	
Dec. NPV	Relative elevation 181 ^b	
Dec. shade	Relative elevation 191	
Feb. PV	Relative elevation 201	
Feb. NPV		
Feb. shade		
PV difference		
NPV difference		
Shade difference		

The symbol ρ signifies reflectance.

^a Standard deviation was also included as a variable for all multispectral inputs.

^b Denotes variable selected as a decision tree node.

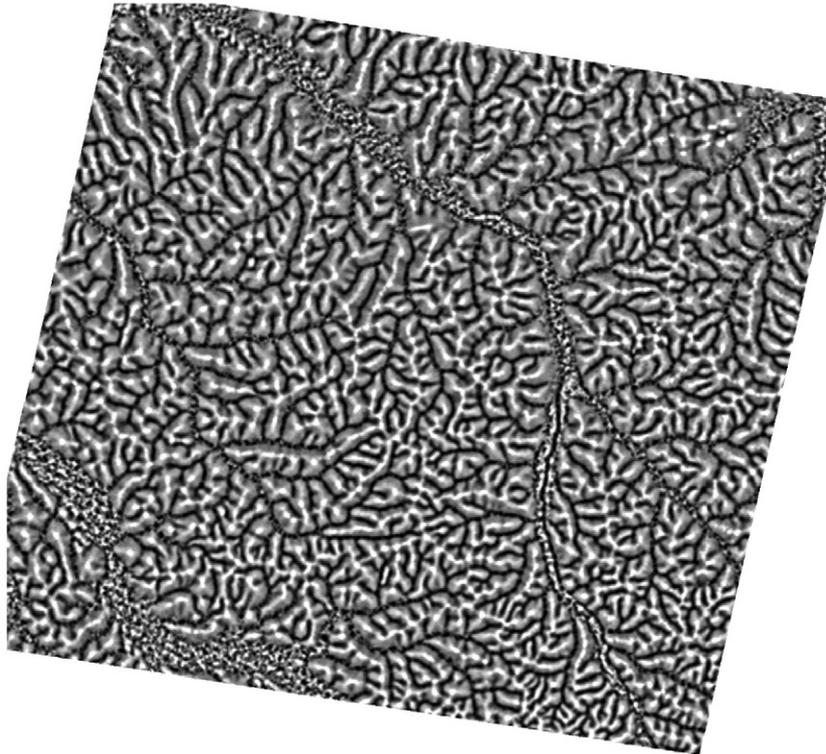


Fig. 4. Ranked relative elevation using 11 pixel by 11 pixel (990 m×990 m) window. Brighter tones indicate higher relative rank.

within the study area, the contrast between the two scenes may not be as drastic as in other years because the winter 2006–2007 dry season was particularly wet. The reflectance value of each of four image bands was included as a separate input to the classification. A green band covers a spectral range of approximately 0.50 to 0.59 μm , a red band covers approximately 0.61 to 0.68 μm , a near infrared band covers

approximately 0.79 to 0.89 μm , and a shortwave infrared band covers approximately 1.58 to 1.75 μm .

Reflectance from three of the SPOT bands also provided the source data for the calculation of two common vegetation indices: the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Infrared Index (NDII). NDVI is a measure of vegetation

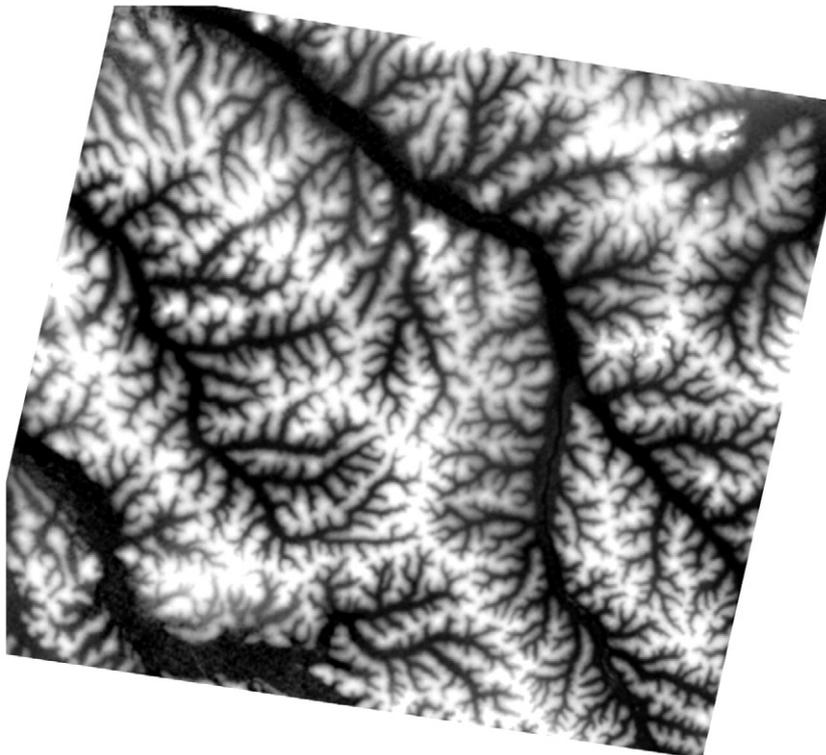


Fig. 5. Ranked relative elevation using 201×201 pixel (18,090 m×18090 m) window. Brighter tones indicate higher relative rank.

greenness that relies on a spectral change in reflectance caused by chlorophyll absorption in red wavelengths and leaf additive reflectance in near infrared wavelengths (Rouse et al., 1973). NDVI is calculated using near infrared wavelength reflectance (ρ_{NIR}) and red wavelength reflectance (ρ_{red}):

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad (1)$$

NDII is a measure of canopy water absorption in the shortwave infrared (Hardisky et al., 1983; Hunt and Rock, 1989). NDII is calculated using ρ_{NIR} and shortwave infrared reflectance (ρ_{SWIR}):

$$NDVI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}} \quad (2)$$

Spectral mixture analysis (SMA) was also used to provide classification inputs. SMA models a mixed spectral signature within a remotely sensed pixel as a combination of purer reference endmembers (Adams et al., 1993). SMA fractions describe image pixels in terms of the relative abundance of each of the endmembers. For the dambo classification, photographic land cover samples were captured in the field at 10 m intervals along a reference line bisecting plots composed of a particular class. Upon returning from the field, these photographs were used to determine the fractional coverage of the following three endmembers within each plot: photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and shade (White et al., 2000). Image endmembers used to model the SPOT scene were then selected using Constrained Reference Endmember Selection (Roberts et al., 1993). The results of the SMA, namely the fractional abundance of PV, NPV, and shade for the December and January images, provided additional inputs to the classification.

Several input variables were calculated to capture spatio-temporal patterns in index values. First, the difference between February, 2007 and December, 2006 variables was calculated for each multispectral input variable to assess the impacts of changing moisture availability during the dry season. Standard deviation was then calculated for nine-pixel moving windows as a measure of the spatial heterogeneity of each layer.

Shuttle Radar Topography Mission (SRTM) terrain data provided the elevation values for the classification; they also served as the basis for the calculation of slope and 20 ranked relative elevation calculations. The SRTM data were resampled from 90-m to 20-m resolution and smoothed using a nine-pixel neighborhood mean to reduce the effects of data anomalies and local topographic variation not related to the surrounding catchment (e.g., trees and large termite mounds).

Moving from south to north, the study area exhibited a general decrease in elevation and a flattening in topographic gradient. In other words, the change in elevation from upland to bottom was generally greater and more abrupt in the southern part of the study area than in the north. Due to these terrain attributes, and because dambos exist at various scales within the drainage system, relative elevation was more effective at distinguishing classes than overall elevation. BDTs are capable of processing large volumes of data to select the most significant input variables for a landscape classification. This allowed for the inclusion of multiple relative elevation layers calculated for different spatial scales. Relative elevation was calculated within windows of varying sizes. These windows ranged in extent from 11 by 11 to 201 by 201 pixels in width and height (Figs. 4 and 5), which translates to 990 by 990 m and to 18,090 by 18,090 m respectively. Each pixel was ranked according to its elevation relative to the other pixels within a moving window of a given size. To normalize relative elevation values, the rank of the center pixel was then divided by the total number of pixels within the window. As the size of the windows increased, so did the topographic features they accentuated. At the smallest window sizes, lower order streams and ridges were clearly defined, while the major

rivers were too broad to be portrayed as relative low points (Fig. 4). In contrast, larger windows captured major topographic features while ignoring some upper-catchment features (Fig. 5).

In addition to the general decrease in elevation, field observations and satellite images revealed drier conditions moving south to north within the study area. The main indicators of this variation were an increase in senescent vegetation and a decrease in standing water within the dambo bottoms. These observations led to the inclusion of northing and easting values derived from a Universal Transverse Mercator (UTM) projection of the study area to serve as proxies for climatic gradients that may exist within different spatial sub-regions of the study area.

Layers containing each of the 66 input variables for the extent of the SPOT scenes were then consolidated into a single stack in preparation for use in the landscape classification. Next, field data were divided into training and validation sets. The homogeneous polygon samples were stratified according to dambo class, while each cross-section transect was preserved as an individual unit composed of multiple classes. The training and validation sets were random selections consisting of data from both field sampling techniques. Pixel values from each of the input variables were then extracted for regions of interest representing the training set. These values provided the source data from which the “tree” package (Ripley, 2007) in R constructed the BDT classifier. Construction of the tree also determined which of the input variables were significant in identifying the dambo classes for the training set. When applied to the inputs

Table 3

Generalized least squares regression predictions of soil properties by landscape unit and depth, including number of samples and overall significance of the regression model

Depth (cm)	N	p-value	Bottom	Floor	Margin	Upland
<i>% clay (in fine <2 mm fraction)</i>						
5	193	<0.0001	31.3	21.3	17.5	34.1
25	130	<0.0001	30.0	25.5	22.2	41.4
55	117	<0.0001	43.7	57.1	22.1	45.8
95	52	<0.0001	44.5	56.3	25.8	44.2
145	51	0.0017	35.8	41.8	26.6	43.1
205	34	0.1221	32.4	36.3	29.4	43.6
<i>% silt (in fine <2 mm fraction)</i>						
5	193	<0.0001	29.6	24.3	10.0	7.3
25	130	<0.0001	23.8	15.8	5.8	6.8
55	117	0.0010	11.5	8.7	7.1	5.2
95	52	0.0099	10.3	5.3	4.0	5.1
145	51	0.3450	7.0	6.8	4.9	5.1
205	34	0.5413	9.2	5.2	7.2	7.9
<i>SOC g kg⁻¹ (in fine <2 mm fraction)</i>						
5	61	<0.0001	30.0	27.3	11.8	11.6
25	47	0.0190	14.1	13.6	9.1	8.4
55	50	0.0052	8.0	8.8	4.1	5.8
95	31	0.0391	4.6	2.0	2.6	4.1
145	27	0.0010	3.3	1.5	1.6	3.3
205	18	0.0698	1.5	NA	1.1	2.3
<i>Smectite proportion in clay fraction (VisNIR, ordinal units from 0–5)</i>						
5	193	<0.0001	0.3	0.8	0.2	0.0
25	193	<0.0001	0.6	0.5	0.2	0.0
55	184	<0.0001	1.3	1.3	0.3	0.0
95	181	<0.0001	1.9	1.3	0.5	0.1
145	157	<0.0001	2.0	1.6	0.6	0.0
205	130	<0.0001	2.6	1.8	1.8	0.6
<i>Munsell® color, 50–60 cm depth</i>						
Hue YR	193	<0.0001	5.7	5.8	5.5	3.4
Value	193	0.3550	4.1	4.1	4.2	4.2
Chroma	193	<0.0001	2.6	2.9	3.1	5.1
<i>Fertility measures, all depths</i>						
pH	120	0.0001	6.2	6.3	5.5	5.6
ECEC	120	<0.0001	10.8	10.2	3.3	1.7
Base sat.	120	0.0002	97	100	58	59

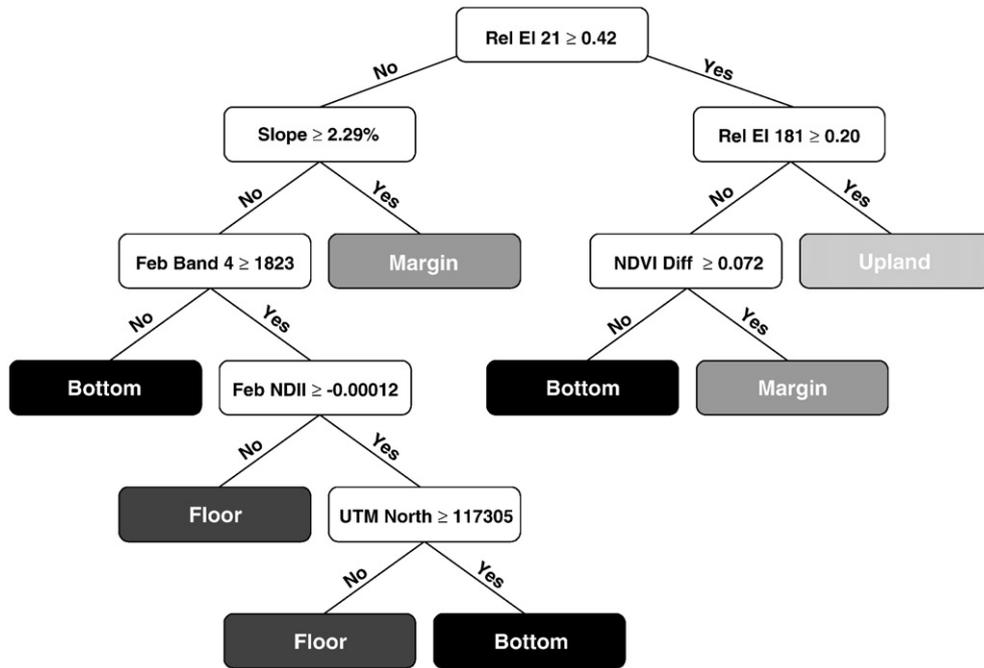


Fig. 6. Binary decision tree classifier diagram. Intermediate decision nodes shown in white, with terminal dambo class nodes shown in shades of gray.

collected for the entire study area, these decision rules produced a map of dambo classes.

After several iterations of the BDT, it was determined that a single bottom class could not adequately classify both the swampy rivers and the narrow, channel-like bottoms in the upper catchments. A secondary

bottom class was added to capture the rivers, with training sites selected visually from the SPOT images. Following the creation of the BDT, the two bottom classes were combined to simplify the results of the classification.

The results of the classification were first evaluated qualitatively, with a particular focus on the continuity of the classes and the

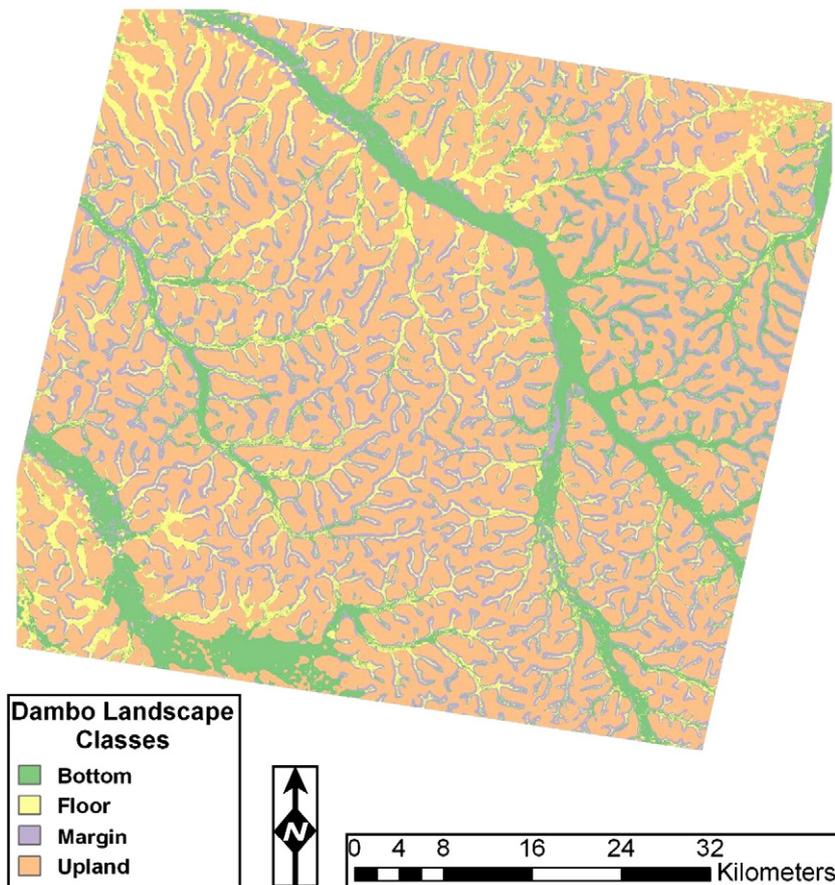


Fig. 7. Classified map of study area.

agreement of the classified image with the dambo classes observed in the field and described in the literature. Quantitative accuracy was assessed by comparing the validation set with the classified image. Two common image classification accuracy assessment techniques were employed. The first was the error matrix, including overall, producer's, and user's accuracy (Congalton, 1991). Second, because overall accuracy may not indicate the actual effectiveness of a classification technique, the more rigorous Kappa analysis was also performed for the BDT classification (Congalton, 1991; Fitzgerald and Lees, 1994). The Kappa test is based upon the null hypothesis that there is no statistically significant relationship between the ground truth data and the results of the classification. The Kappa coefficient (K) is calculated as:

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^r x_{i+} * x_{+i}} \quad (3)$$

where N is the total number of error matrix values, r is the number of rows in the error matrix, x_{ii} is the number of values within row i and column i , x_{i+} is the marginal sum of row i , and x_{+i} is the marginal sum of column i . Kappa values range between 0 and 1, with 0 indicating no agreement and 1 indicating total agreement.

To provide a comparative evaluation of BDT predictive accuracies, we used all of the same predictors with an ensemble “bagging” decision tree classification method, Random Forest (Breiman, 2001; Liaw and Wiener, 2008). We experimented with the following model combinations

(a) 100, 300, 500 and 1000 trees, (b) 1–3 pixel minimum node sizes and (c) with and without Northing and Easting in the calibration. As we found little difference in the results overall (given at least 300 trees), only results from the best model are reported in this paper: 500 trees, minimum node size of 1 and inclusion of Northing and Easting.

4. Results

4.1. Landscape class as a soil predictor

The expert-derived soil-landscape units employed in this study significantly explained soil variability within one 2nd-order catchment located in the larger study area (Table 3). Comparing the top 30 cm in particular, upland and margin soils generally contained less SOC than floor and bottom soils. Clay content was relatively high for all soils except the sandy margins and within 50 cm of the surface of floors. Silt-size particles were found primarily in the upper horizons of the floors and bottoms. Substantial amounts of kaolinite were found in all soils (results not reported) with smectites concentrated in the more poorly drained floor and bottom classes. Soil pH, base saturation and ECEC were all higher for floor and bottom relative to margin and upland soils. Similarly, we found substantially redder hues and higher chromas for upland vs. dambo soils (not surprising given that these soils were used to establish upland color rules). Overall, upland and margin soils appear to have distinctive properties with little apparent separation between soils found in floors and bottoms – despite important differences in hydrology and vegetation for the lowest two landscape units.

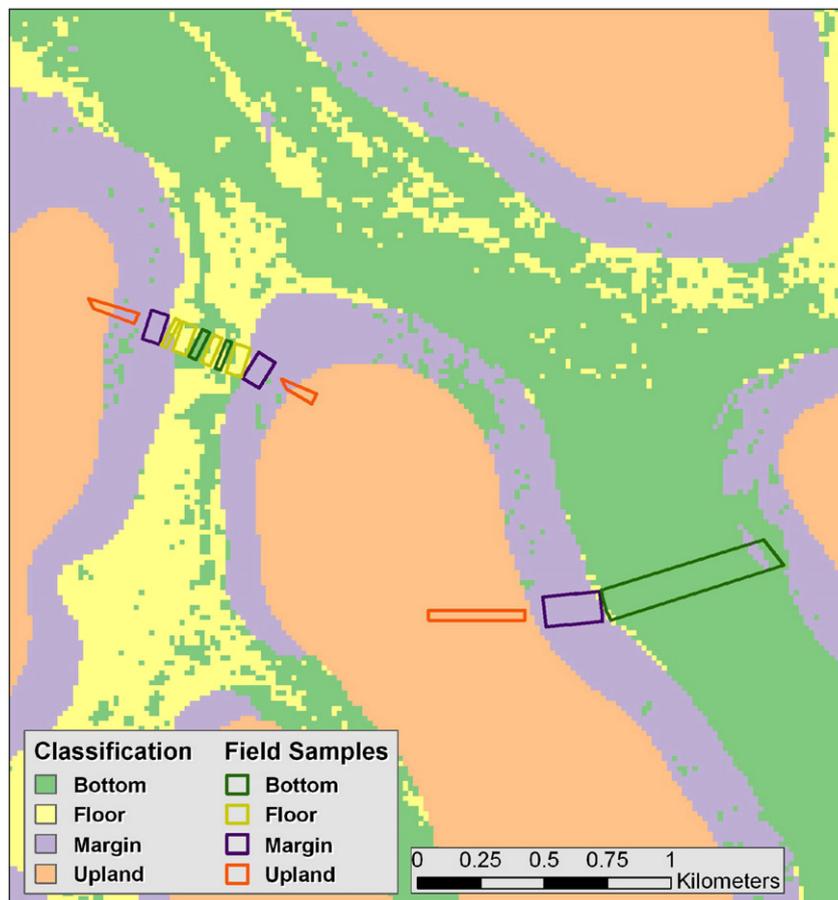


Fig. 8. Subset of the classified map overlaid with validation transects. Validation transects are outlined with colors corresponding to their class.

Table 4

The confusion matrix for the BDT classification of the study area shows the number and percentages of correctly classified pixels

		Ground truth				Total	User's accuracy (%)
		Bottom	Floor	Margin	Upland		
Classified	Bottom	520	142	88	22	772	67.4
	Floor	130	1002	232	0	1364	73.5
	Margin	56	258	1159	294	1767	65.6
	Upland	0	0	102	1396	1498	93.2
	Total	706	1402	1581	1712	5401	
Producer's accuracy		73.7%	71.5%	73.3%	81.5%		75.5

Column totals indicate the number of ground truth pixels from each class. Row totals indicate the number of pixels assigned to each class by the BDT classifier. Bold values along the diagonal indicate the number of pixels classified as the correct class.

4.2. Landscape classification with binary decision trees (BDT)

The BDT classifier constructed from the training data consisted of seven intermediate nodes and eight terminal nodes (Fig. 6). In essence, seven decision rules were used to assign the input data to membership in one of the four dambo classes. The classification relied upon a combination of topographic, spectral, and spatial inputs. Uplands were identified using relative elevation, while a combination of slope, SPOT bands, vegetation indices, and UTM northing coordinates were used to identify the other three classes.

Applying the BDT-derived decision rules to the entire extent of the input data resulted in a map of dambo classes (Fig. 7). A visual assessment of the classification map showed general agreement with the idealized dambo cross-section. The lowland classes were surrounded by the higher terrain of the upland class, with the margins as a transitional band between them. As was observed in the field, dambo bottoms were more stream-like in the upper catchments, becoming broader as they approached the rivers. Beyond the boundaries of the study area, the classification showed decreased visual agreement with the typical dambo cross-section. Floors and margins appeared to be overclassified in these areas, although no accuracy assessment data were available to test this observation. Even with the addition of a second bottom class, the swampy expanses of the Lugogo, Towa, and Mayanja rivers were occasionally misclassified as upland or margin.

A randomly selected, independent validation set was used to assess the accuracy of the classification. Examples of qualitative agreement are shown in a comparison of the classified map with the validation transects (Fig. 8). On the left, the classification successfully identified convergent branches of dambo bottom and the floor between them. On the right, the classification accurately portrayed an exceptional case in which the margin transitioned directly to bottom without the typical floor in between. A qualitative assessment of the classification also revealed items in need of improvement. In Fig. 8, small portions of the margin were erroneously identified as bottom, and on the left, the transition from margin to upland did not agree as well with the validation transect.

The quantitative accuracy assessment found an overall accuracy of 75.5%. Subdividing the validation set to construct an error matrix provided information about individual class accuracy (Table 4; Congalton, 1991). In an error matrix, the cell values along the diagonal

Table 5

Classification results by area (km²) and as a percentage of total area

Dambo class	Area (km ²)	Area %
Upland	1306	59.0%
Margin	467	21.1%
Floor	265	12.0%
Bottom	176	7.9%
Total	2214	100.0%

Table 6

The confusion matrix for the Random Forest (500 trees, node size = 1) classification of the study area shows the number and percentages of correctly classified pixels

		Ground truth				Total	User's accuracy (%)
		Bottom	Floor	Margin	Upland		
Classified	Bottom	616	362	44	13	1035	59.5
	Floor	79	758	82	0	919	82.5
	Margin	11	282	1311	147	1751	74.9
	Upland	0	0	144	1552	1696	91.5
	Total	706	1402	1581	1712	5401	
Producer's accuracy		87.3%	54.1%	82.9%	90.7%		78.5

Column totals indicate the number of ground truth pixels from each class. Row totals indicate the number of pixels assigned to each class by the Random Forest classifier. Bold values along the diagonal indicate the number of pixels classified as the correct class.

give the quantity of accurately classified pixels for each class. These values were compared with corresponding row totals to provide user's accuracy, which is the number of correctly identified pixels as a percentage of the total pixels assigned to the class. Diagonal values were compared to corresponding column totals to calculate producer's accuracy, which is class assignment as a percentage of the actual number of pixels belonging to that class on the ground. Producer's accuracy and user's accuracy are measures of omission and commission errors, respectively. The BDT classifier identified uplands more effectively than any other class, probably due to their direct connection to relative elevation. The most common classification errors involved confusion with adjacent ordinal classes, which is not surprising when assigning discrete class divisions to a continuous landscape. To illustrate, only 3% of the accuracy assessment pixels were misclassified by two or more ordinal classes. In other words, 97% of the pixels were either classified correctly or confused with an adjacent class. A Kappa value of 0.67 was calculated for the classification, showing significant agreement between the BDT classifier and the field data (Fitzgerald and Lees, 1994). This suggests that the classified map should be relatively reliable for discerning dambo classes within the study area.

Class membership as a percentage of the overall study area varied greatly (Table 5). At 59% of the entire study area, the upland class was by far the most abundant. The margins were the next most plentiful at 21%. The floor and bottom classes represented only 12% and 8% of the 2214 km² study area, respectively. These class membership percentages seem to approximate the actual relative abundance of the classes observed in the field.

4.3. Landscape classification with Random Forest

Random Forest yielded only minor improvements in overall accuracy. Results for the Random Forest run using 500 trees with northing and easting, and a minimum node size of 1 are shown in Table 6. This modeling run resulted in an overall accuracy of 78.5%. Producer's accuracy increased for bottom, margin, and upland classes, but decreased to 54.1% for the floor class. User's accuracy was higher for the floor and margin classes but lower for the bottom class. Bottom was substantially overclassified and floor was substantially underclassified relative to the simple BDT.

5. Discussion

5.1. Utility of expert landscape classes

Soil surveyors have long used readily observable ground features to inexpensively delineate landscape units in the field, often in combination with auger probing for rapid detection of soil changes (Clarke, 1936; Kellogg, 1937). We modernized this traditional soil

mapping technique through the use of handheld GPS receivers to georeference our ground observations, a task traditionally accomplished by drawing on aerial photographs in the field. (In a few instances, we selected pixels through visual examination of satellite imagery, to augment our field observations.)

Three conditions must be met for this expert approach to work: (1) landscape classes should be easily identified in the field with *consistent, unambiguous and easily implemented rules*; (2) landscape classification rules must be related directly or indirectly to the exhaustive spatial covariates available for predictive mapping (e.g. multispectral imagery and digital elevation); and (3) landscape classes must predict soil properties of interest. In this study, we initially used augers to obtain and measure subsoil color in order to delineate upland areas – a task that slowed field work significantly for the first week in the field. Recognizing that subsoil colors could be rapidly observed at the surface due to termite activity allowed us to meet criterion 1 as measuring slope gradient and observing hummocks took little time. Despite a steep initial learning curve, we were able to acquire approximately 10,000 pixels of “ground truth” in just four weeks of field work. Given our validation results, criterion 2 was also met. For criterion 3, our landscape classes captured important differences in soil texture, clay mineralogy, chemistry, color and SOC. Though dambo floor and bottom soils were quite similar, this could be due to the relatively narrow and dry dambos found in the 2nd-order catchment used for this analysis. Valley floors generally widen and are saturated for longer periods “downstream.” Future work should more systematically evaluate landscape class–soil relationships for the study area as a whole.

5.2. Landscape classification with binary decision trees (BDT)

As mentioned previously, one of the advantages BDTs offer is the identification of the significant input variables and threshold values used to create the decision rules for the image classification (Fig. 6, Table 2). The results of this classification add credence to Tiner's (1993) warning to avoid relying solely on vegetation data to classify seasonal wetlands. DEM-derived relative elevation and slope variables provided three of the seven decision tree nodes. Given the relationship between topography and hydrology, it was not surprising that these variables were significant inputs for dambo classification. Uplands were identified by their high relative elevation at fine and coarse scales (1890 by 1890 m and 16,290 by 16,290 m). On the left branches of the decision tree, margins were distinguished from the other classes by their relatively steep slopes. UTM northing was the other decision node variable not directly linked to vegetation, although it may have been an indirect measure of relative greenness due to localized climatic variation. While conducting field work in January and February, 2007, it was apparent that green vegetation and standing water decreased from south to north within the study area. This latitudinal variation coincides with a decrease in elevation and may also be linked to precipitation gradients.

The remaining decision node variables were derived from satellite imagery and all seemed to be measures of vegetation phenology, either directly or indirectly. Changes in NDVI values between December, 2006 and February, 2007 indicate senescence during the dry season. The steeply sloping margins with their sandy soils experienced greater changes in vegetation greenness than the relatively flat, saturated bottoms. Both shortwave infrared reflectance and NDII change with grass senescence. These variables allowed separation of bottoms and floors, since saturated bottom soils sustain green vegetation even during prolonged dry periods.

Several data-related factors may have limited the accuracy of the classification. The greatest limitation was the lack of a higher-resolution DEM. If LIDAR, radar interferometry, or other high-resolution elevation data were available for this area, classification accuracy may have been increased. Another limitation may have been

the size and spatial extent of the sample data. Even after four weeks of rigorous field sampling, some less accessible sections of the study area remained unvisited (Fig. 3).

The overall performance of the BDT-derived dambo wetland classifier was promising, although its application in other regions would most likely require local training and accuracy assessment data. But the landscape rules derived from this modeling exercise also contribute directly to the larger body of “expert knowledge” that can be employed to map dambo catenae elsewhere in Africa. The actual utility of the class map resulting from this project will be evaluated as it is employed in ongoing biogeochemical process measurement and modeling. At the time of this writing, the results of the classification had already been used to select representative sites for detailed soil, hydrology, vegetation and methane gas flux measurements.

6. Conclusions

Utilizing 193 soil profiles previously sampled and characterized from a 2nd-order catchment within the study area, the expert landscape classes employed in this study yielded highly significant generalized least squares regression models for most soil properties at depths of 5 to 205 cm. To highlight major soil differences, we found redder and higher chroma soils in uplands, sandy soils in margins, silty surface soils in floors and bottoms, more acidic soils in uplands and margins, significant amounts of smectites in floor and bottom subsoils, and much higher SOC in the top 30 cm of floor and bottom soils. These results were generally consistent with descriptions and data found in published surveys and peer-reviewed studies, though floor and bottom soils were not as distinct as anticipated.

Using multispectral and topographic remote sensing inputs, this study produced a relatively accurate (75% overall, 97%+/- one class, and Kappa=0.67), 20 m spatial resolution map of four dambo wetland landscape classes. At 59% of the 2214 km² study area, the upland class was by far the most abundant, with margins at 21%, floors at 12% and bottoms at 8%. For comparison, Random Forest modeling with the same predictors yielded only a slight overall improvement (78.5%), substantial problems in bottom classification, and a difficult to interpret model structure.

Several observations were made in the process of constructing this model which may be beneficial in future digital soil mapping and wetland delineation dambo research. The results of this study demonstrate that binary decision trees (BDT) can use spatial data to reliably classify soil-landscape elements within dambo-terminated catenae, including three different dambo wetland classes. Valuable classification inputs were shown to come from a variety of sources. In particular, shortwave infrared reflectance, NDVI, NDII, relative elevation, slope, and northing (possibly a proxy for precipitation) were important for this classification. We found that field sampling was critical to the success of this project, as was the capacity to calculate and compare a variety of topographic and spectral metrics.

We propose the approach described in this paper as a relatively inexpensive way to accurately disaggregate association or catena map units. In areas where formal soil surveys are not available, it might still be possible to disaggregate catenae based upon published information from similar landscapes, as we did with the three dambo classes in this study. As such, this study represents an intermediate step between: (1) expert derivation of landscape rules based upon prior knowledge; and (2) full-scale field soil sampling and analysis, with subsequent soil-landscape modeling. Ideally, field-supported landscape classification would be followed by targeted soil sampling and soil-landscape modeling. But for many parts of Africa, simply disaggregating catenae and assigning the best available soil property data to these associates would represent a valuable contribution to soil science and land management even if additional soil sampling was not possible.

Acknowledgements

This work was supported by the National Science Foundation Geography and Regional Science Program, under Grant Nos. 0620206 and 0620142. Kawanda Agricultural Research Institute, Uganda, and Dr. Moses Isabirye are thanked for providing field assistance.

References

- Acres, B.D., Rains, A.B., King, R.B., Lawton, R.M., Mitchell, A.J.B., Rackham, L.J., 1985. African dambos: their distribution, characteristics and use. In: Thomas, M.F., Goudie, A.S. (Eds.), *Dambos: Small Channelless Valleys in the Tropics*. Zeitschrift für Geomorphologie, Supplementband, vol. 52, pp. 63–86.
- Adams, J.B., Smith, M.O., Gillespie, A.R., 1993. Imaging spectroscopy: interpretations based on spectral mixture analysis. In: Pieters, C.M., Englert, P.A. (Eds.), *Remote Geochemical Analysis: Elemental and Mineralogical Composition*. Cambridge University Press, Cambridge, UK, pp. 145–166.
- Amarsaikhan, D., Douglas, T., 2004. Data fusion and multisource image classification. *International Journal of Remote Sensing* 10, 3529–3539.
- Baker, C., Lawrence, R., Montagne, C., Patten, D., 2006. Mapping wetlands and riparian areas using Landsat ETM+ imagery and decision-tree-based models. *Wetlands* 26, 465–474.
- Bartlett, K.B., Harriss, R.C., 1993. Review and assessment of methane emissions from wetlands. *Chemosphere* 26 (1–4), 261–320.
- Behrens, T., Scholten, T., 2006. Digital soil mapping in Germany – a review. *Journal of Plant Nutrition and Soil Science-Zeitschrift Für Pflanzenernahrung Und Bodenkunde* 169 (3), 434–443.
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- Brown, D.J., 2005. A historical perspective on soil-landscape modeling. In: Grunwald, S. (Ed.), *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics*. CRC Press, Boca Raton, FL, pp. 61–103.
- 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 (4), 444–453.
- Brown, D.J., Helmke, P.A., Clayton, M.K., 2003. Robust geochemical indices for redox and weathering on a granitic laterite landscape in central Uganda. *Geochimica Et Cosmochimica Acta* 67 (15), 2711–2723.
- Brown, D.J., Clayton, M.K., McSweeney, K., 2004a. Potential terrain controls on soil color, texture contrast and grain-size deposition for the original catena landscape in Uganda. *Geoderma* 122 (1), 51–72.
- Brown, D.J., McSweeney, K., Helmke, P.A., 2004b. Statistical, geochemical, and morphological analyses of stone line formation in Uganda. *Geomorphology* 62 (3–4), 217–237.
- Brown, D.J., Shepherd, K.D., Walsh, M.G., Mays, M.D., Reinsch, T.G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma* 132 (3–4), 273–290.
- Brown de Colstoun, E.C., Story, M.H., Thompson, C., Commisso, K., Smith, T.G., Irons, J.R., 2003. National Park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier. *Remote Sensing of Environment* 85, 316–327.
- Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma* 103 (1–2), 79–94.
- Bui, E.N., Moran, C.J., 2003. A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray-Darling basin of Australia. *Geoderma* 111 (1–2), 21–44.
- Bui, E.N., Loughhead, A., Corner, R., 1999. Extracting soil-landscape rules from previous soil surveys. *Australian Journal of Soil Research* 37 (3), 495–508.
- Bullock, A., 1992. Dambo hydrology in Southern Africa – review and reassessment. *J. Hydrol.* 134 (1–4), 373–396.
- Carre, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: beyond DSM. *Geoderma* 142 (1–2), 69–79.
- Clarke, G.R., 1936. *The Study of Soil in the Field*. Clarendon Press, Oxford.
- Conacher, A.J., Dalrymple, J.B., 1977. The nine-unit landsurface model: an approach to pedogeomorphic research. *Geoderma* 18, 1–154.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37, 35–46.
- Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L., Helt, T., 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma* 97 (3–4), 367–391.
- Eswaran, H., Almaraz, R., vandenBerg, E., Reich, P., 1997. An assessment of the soil resources of Africa in relation to productivity. *Geoderma* 77 (1), 1–18.
- Fitzgerald, R.W., Lees, B.G., 1994. Assessing the classification accuracy of multisource remote sensing data. *Remote Sensing of Environment* 47, 362–368.
- Wetland development and management in SADC countries. In: Frenken, K., Mharapara, I. (Eds.), *Proceedings of a Sub-regional Workshop*, Nov. 19–23, 2001. FAO SAFR, Harare, Zimbabwe.
- Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. In: Klute, A. (Ed.), *Methods of Soil Analysis: Part I. Physical and Mineralogical Methods*. Soil Science Society of America, Madison, WI, pp. 383–411.
- Hardisky, M.A., Klemas, V., Smart, R.M., 1983. The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of *Spartina alterniflora* canopies. *Photogrammetric Engineering and Remote Sensing* 49, 77–83.
- Harvey, K.R., Hill, G.J.E., 2001. Vegetation mapping of a tropical freshwater swamp in the Northern Territory, Australia: a comparison of aerial photography, Landsat TM and SPOT satellite imagery. *International Journal of Remote Sensing* 22, 2911–2925.
- Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. *Geoderma* 124 (3–4), 383–398.
- Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120 (1–2), 75–93.
- Hunt, E.R., Rock, B.N., 1989. Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment* 30, 43–54.
- Kellogg, C.E., 1937. *Soil Survey Manual*, Misc. Pub. No. 274. U.S.D.A., Washington, D.C., 136 pp.
- Langdale-Brown, I., Osmaston, H.A., Wilson, J.G., 1964. *The Vegetation of Uganda: and its Bearing on Land Use*. Government of Uganda, Entebbe, Uganda, 159 pp.
- Liaw, A., Wiener, M., 2008. The Random Forest Package for R (Version 4.5-25). <http://www.r-project.org/>.
- Lin, H., Bouma, J., Pachepsky, Y., Western, A., Thompson, J., van Genuchten, M.T., Vogel, H.-J., Lilly, A., 2006. Hydropeology: synergistic integration of pedology and hydrology. *Water Resources Research* 42 (5).
- Lowry, J., Ramsey, R.D., Thomas, K., Schrupp, D., Sajwaj, T., Kirby, J., Waller, E., Schrader, S., Falzarano, S., Langs, L., Manis, G., Wallace, C., Schulz, K., Comer, P., Pohs, K., Rieth, W., Velasquez, C., Wolk, B., Kepner, W., Boykin, K., O'Brien, L., Bradford, D., Thompson, B., Prior-Magee, J., 2007. Mapping moderate-scale land-cover over very large geographic areas within a collaborative framework: a case study of the Southwest Regional Gap Analysis Project (SWReGAP). *Remote Sensing of Environment* 108, 59–73.
- Mäckel, R., 1974. Dambos: a study in morphodynamic activity on the plateau regions of Zambia. *Catena* 1, 327–365.
- Mäckel, R., 1985. Dambos and related landforms in Africa – an example for the ecological approach to tropical geomorphology. In: Thomas, M.F., Goudie, A.S. (Eds.), *Dambos: Small Channelless Valleys in the Tropics*. Zeitschrift für Geomorphologie, Supplementband, vol. 52, pp. 1–23.
- MacMillan, R.A., Moon, D.E., Coupe, R.A., 2007. Automated predictive ecological mapping in a forest region of BC, Canada, 2001–2005. *Geoderma* 140 (4), 353–373.
- McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117 (1–2), 3–52.
- McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma* 89 (1–2), 67–94.
- Minasny, B., McBratney, A.B., Lark, R.M., 2008. Digital soil mapping technologies for countries with sparse data infrastructures. *Digital Soil Mapping with Limited Data*, pp. 15–30.
- Morison, C.G.T., 1948. The catena concept and the classification of tropical soils. First Commonwealth Conference on Tropical and Subtropical Soils. Commonwealth Bureau of Soil Science, Harpenden, England, pp. 124–128.
- Morison, C.G.T., Hoyle, A.C., Hope-Simpson, J.F., 1948. Tropical soil-vegetation catenas and mosaics. *Journal of Ecology* 36, 1–84.
- Nye, P.H., 1954. Some soil-forming processes in the humid tropics I. A field study of a catena in the west African forest. *Journal of Soil Science* 5, 7–21.
- Nye, P.H., 1955a. Some soil-forming processes in the humid tropics II. The development of the upper-slope member of the catena. *Journal of Soil Science* 6, 51–62.
- Nye, P.H., 1955b. Some soil-forming processes in the humid tropics III. Laboratory studies on the development of a typical catena over granitic gneiss. *Journal of Soil Science* 6, 63–72.
- Ollier, C.D., 1959. A two-cycle theory of tropical pedology. *Journal of Soil Science* 10, 137–148.
- Ollier, C.D., Lawrence, C.J., Beckett, P.H.T., Webster, R., 1969. Terrain classification and data storage: land systems of Uganda. M.E.X.E. Report No. 959, Military Engineering Experimental Establishment, Christchurch, Hampshire, U.K.
- Pal, M., Mather, P.M., 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment* 86, 554–565.
- Palm, C., Sanchez, P., Ahamed, S., Awiti, A., 2007. Soils: a contemporary perspective. *Annual Review of Environment and Resources* 32, 99–129.
- Pohl, C., Van Genderen, J.L., 1998. Multisensor image fusion in remote sensing: concepts, methods, and applications. *International Journal of Remote Sensing* 19, 823–854.
- Radwanski, S.A., 1959. *The Soils and Land Use of Buganda*. 4, Department of Agriculture, Uganda.
- Radwanski, S.A., Ollier, C.D., 1959. A study of an East African catena. *Journal of Soil Science* 10, 149–168.
- R Development Core Team, 2008. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 3-900051-07-0. URL <http://www.R-project.org>.
- Ripley, B.D., 2007. The Tree Package for R (Version 1.0-26). <http://www.r-project.org/>.
- Roberts, D.A., Adams, J.B., Smith, M.O., 1993. Discriminating green vegetation, non-photosynthetic vegetation and soils in AVIRIS data. *Remote Sensing of Environment* 44, 255–270.
- Rossiter, D.G., 2008. Digital soil mapping as a component of data renewal for areas with sparse soil data infrastructures. In: Hartemink, A.E., McBratney, A., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer, pp. 69–80.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.V., 1973. Monitoring vegetation systems in the Great Plains with ERTS. Third ERTS Symposium, NASA Publication No. SP-351, pp. 309–317.
- Ruhe, R.V., 1961. Elements of the soil landscape. *Transactions of the 7th International Congress of Soil Science*, IV, pp. 165–169.
- Scoones, I., 1991. Wetlands in drylands – key resources for agricultural and pastoral production in Africa. *Ambio* 20 (8), 366–371.
- Scully, P., Okin, G., Chadwick, O.A., Franklin, J., 2005a. A comparison of methods to predict soil surface texture in an alluvial basin. *Professional Geographer* 57, 423–437.
- Scully, P., Franklin, J., Chadwick, O.A., 2005b. The application of classification tree analysis to soil type prediction in a desert landscape. *Ecological Modelling* 181, 1–15.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* 66 (3), 988–998.

- Simard, M., Saatchi, S., De Grandi, G., 2000. The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. *IEEE Transactions on Geoscience and Remote Sensing* 38, 2310–2321.
- Simard, M., De Grandi, G., Saatchi, S., Mayaux, P., 2002. Mapping tropical coastal vegetation using JERS-1 and ERS-1 radar data with a decision tree classifier. *International Journal of Remote Sensing* 23, 1461–1474.
- Survey Department, 1967. Atlas of Uganda. Government Printer, Entebbe, Uganda.
- Tiner, R.W., 1993. Using plants as indicators of wetland. *Proceedings of the Academy of Natural Sciences of Philadelphia* 144, 240–253.
- Trapnell, C.G., 1943. *The Soils, Vegetation and Agriculture of Northeastern Rhodesia*. Government Printer, Lusaka.
- Trapnell, C.G., Clothier, J.N., 1937. *The Soils, Vegetation and Agricultural Systems of Northwestern Rhodesia*. Government Printer, Lusaka.
- Varma, H., Fadaie, K., Habbane, M., Stockhausen, J., 2003. Confusion in data fusion. *International Journal of Remote Sensing* 24, 627–636.
- von der Heyden, C.J., 2004. The hydrology and hydrogeology of dambos: a review. *Progress in Physical Geography* 28 (4), 544–564.
- Watson, J.P., 1964a. A soil catena on granite in southern Rhodesia, II. Analytical data. *Journal of Soil Science* 15, 251–257.
- Watson, J.P., 1964b. A soil catena on granite in southern Rhodesia, III. Clay minerals, IV. Heavy minerals, V. Soil evolution. *Journal of Soil Science* 16, 158–169.
- Watson, J.P., 1964c. A soil catena on granite in southern Rhodesia, I. Field observations. *Journal of Soil Science* 15, 238–250.
- Webster, R., 1965. A catena of soils on the Northern Rhodesia plateau. *Journal of Soil Science* 16, 31–43.
- White, M.A., Asner, G.P., Nemani, R.R., Privette, J.L., Running, S.W., 2000. Measuring fractional cover and leaf area index in arid ecosystems: digital camera, radiation transmittance, and laser altimetry methods. *Remote Sensing of Environment* 74, 45–57.
- Wright, C., Gallant, A., 2007. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. *Remote Sensing of Environment* 107, 582–605.
- Xu, M., Watanachaturaporn, P., Varshney, P.K., Arora, M.K., 2005. Decision tree regression for soft classification of remote sensing. *Remote Sensing of Environment* 97, 322–336.
- Zhu, L., Tateishi, R., 2006. Fusion of multisensor multitemporal satellite data for land cover mapping. *International Journal of Remote Sensing* 27, 903–918.