Catena 125 (2015) 169–182

Contents lists available at ScienceDirect

Catena

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

Rescue and renewal of legacy soil resource inventories: A case study of the Limpopo National Park, Mozambique

A.H. Cambule [a,b,*], D.G. Rossiter [b,d], J.J. Stoorvogel [c], E.M.A. Smaling [b]

[a] Universidade Eduardo Mondlane, Faculdade de Agronomia e Engenharia Florestal, C.P. 257, Maputo, Mozambique
[b] University of Twente, Faculty of Geoinformation Science and Earth Observation (ITC), PO Box 217, 7500 AE Enschede, The Netherlands
[c] Wageningen University, Land Dynamics Group, PO Box 47, 6700 AA Wageningen, The Netherlands
[d] Department of Crop and Soil Sciences, Cornell University, Ithaca, 14853 NY, USA

ARTICLE INFO

Article history:
Received 25 June 2013
Received in revised form 24 September 2014
Accepted 16 October 2014
Available online 9 November 2014

Keywords:
Adequacy criteria
Soil organic carbon stocks
Digital soil mapping
Data mining
Data archaeology

ABSTRACT

Many areas of developing countries are covered by legacy soil surveys, which, however are hardly used, as they are not available in digital form, used outdated standards, and have unknown quality. There have been very few attempts to rescue and renew these surveys, nor are there established criteria for the evaluation of their quality. We therefore decided to test the applicability of the Cornell Adequacy Criteria (CAC) to assess the quality of several renewed soil surveys in or near the Limpopo National Park, Mozambique (centroid: 23° 18′ 55.57″ S, 31° 55′ 16.24″ E), using the concepts of digital soil mapping. The quality was assessed for mapping and monitoring soil organic carbon (SOC), in terms of geodetic control, positional accuracy, map scale, and texture and adequacy of map legend. Metadata was attached to the renewed maps. SOC stocks were estimated qualitatively based on the description of the map units and quantitatively by the measure-and-multiply approach from legacy laboratory measurements. The positional accuracy of georegistration was 13 to 45% of the square root of a Minimum Legible Area (MLA). Point and area-class layers could be created with high positional accuracy. However the index of maximum reduction was high, indicating that the original publication scale could be reduced. Map unit definitions and overall information content of the surveys were adequate. Integration of remotely sensed optical imagery and digital elevation models could be used to derive accurate contours, against which the positional accuracy of contour-based map borders was assessed. Less than 30% of their lengths were within a distance equal to the square root of MLA. These sources could not be used to evaluate internal map borders, due to the subdued topography and major land-use changes since the original survey. Qualitative estimates of SOC are between low and medium, consistent with other studies in this area. The CAC proved to be a useful framework for determining the fitness for use of legacy surveys.

1. Introduction

The demand for soil information to support land use planning (e.g., agricultural production, infrastructure development, re-settlement, and designation of conservation areas) in developing countries is increasing, yet resources are limited for new soil surveys. However, legacy soil surveys cover large areas and are usually the only source of soil geographic data. Despite the valuable information gathered at considerable cost, this legacy data is hardly used. Poor availability, poor documentation, outdated currency (as soil properties may have changed), and the survey concepts and standards that were used make the legacy data often not adequate for decision making. Consequently, these legacy surveys are often ignored and may get lost (Rossiter, 2008). We hypothesize that these legacy soil surveys can still play a useful role as, e.g., a baseline for monitoring of changes in soil organic carbon (SOC) stocks, land degradation or rehabilitation, or explaining land use changes since the original survey date.

In recent years soil survey procedures have been revolutionized through the use of geo-information technology, geostatistics, and publically available data, including remotely-sensed imagery, and digital elevation models (DEM). The emerging paradigm is called digital soil mapping (DSM) (Hartemink et al., 2008; Lagacherie et al., 2007; McBratney et al., 2003). DSM relies on field observations for model building and validation. Legacy surveys can provide much of this information; reducing the amount of new fieldwork required and allowing for a historical perspective. For example, Baxter and Crawford (2008) used legacy records of soil pH in a DSM exercise. Other examples are from Bui and Moran (2001) and Mayr et al. (2010). In the former the authors detailed three methods of disaggregation of reconnaissance map produced by the Murray–Darling basin (southeast Australia) project, which used scattered detailed legacy soil survey data as its basis. In the latter, the authors compared two

---

* Corresponding author at: Universidade Eduardo Mondlane, Faculdade de Agronomia e Engenharia Florestal, C.P. 257 Maputo, Mozambique. Tel.: +258 843285400.
E-mail address: armindo.cambule@uem.mz (A.H. Cambule).

http://dx.doi.org/10.1016/j.catena.2014.10.019
0341-8162/© 2014 Elsevier B.V. All rights reserved.

approaches in using legacy data as the basis of digital soil mapping: (1) using measured soil property data; and (2) using values of soil properties linked to soil classes.

The re-use of legacy data requires its renewal to meet current demands. Such renewal is not common, judging by the low number of publications on the topic. Rossiter (2008) proposed a procedure to rescue and renew legacy data through "data archaeology" (locating legacy surveys and their supporting metadata), "data rescue" (keeping them from being lost), and "data renewal" (bringing them up-to-date and compatible with other databases). The renewal phase includes: (1) geodetic control; (2) area-class delineation and sample point data as geodetically correct Geographical Information System (GIS) coverages with linked attribute databases; (3) the use of auxiliary data (e.g., multispectral images, DEM, and derived terrain parameters) as background and supplemental data; (4) generation of metadata that includes a description of semantics, laboratory methods, and classification systems; and (5) integration of the legacy data into an easily accessible geospatial data infrastructure. The European Archive of Digital Soil Maps of Africa (Selvaradjou et al., 2005), also available from the online ISRIC-World Soil Information database (http://library.isric.org), constitutes a good example of the data archaeology and on-going rescue stages. These maps are only digital scans, not "digital soil maps" as the term is used in DSM. The renewal stage has hardly been addressed except for the impressive though still incomplete progress made on constructing the Soil and Terrain (SoTer) databases (Batjes, 2004; Dijkshoorn, 2003). One other example is from Dent and Ahmed (1995) and Ahmed and Dent (1997), who used statistical techniques to test and re-interpret archival data from soil survey of the tidal floodplain of the Gambia River. The subjectively-defined and mapped soil series from the original survey did not match the soil taxonomic units derived by cluster analysis of validation observations, showing the added value of geostatistics to renew legacy soil survey data. Despite these few efforts, there are no quality criteria to guide legacy map and data renewal to meet current and future demands for soil information.

Mozambique is typical for sub-Saharan African countries with respect to its soil survey history: pre-independence by the colonial power and post-independence by international and national projects. A systematic survey with a consistent standard is lacking and, given the available resources, not likely in the near future. Yet the country depends on its soil resources for agriculture, infrastructure, and environmental services. There is plenty of data for Mozambique which still need to be rescued and renewed. On 23-January-2013, the ISRIC-World Soil Information database (http://library.isric.org/), listed 329 maps and reports covering some parts of the country. A good example is the Massingir area (Massingir town center 23° 55′ 17.04″ S, 32° 09′ 39.6″ E), located at the southern end of the Limpopo National Park (LNP). This park was established in 2001 replacing the "Coutada 16" hunting zone (Ministerio do Turismo, 2003). At that time about 20 000 people lived within the park boundaries and had to be re-settled either outside LNP or within a designated multi-use zone (Ministerio do Turismo, 2003). To assess the soil suitability of the Massingir area, a new soil survey was carried out covering about 6000 ha with only six new soil profiles and seven representative legacy profiles (Rural Consult Lda, 2008). No reference to any renewal processing was reported. As more land elsewhere around the LNP is likely to be targeted by similar land development, it is important to determine the feasibility of bringing the existing legacy soil survey up to acceptable standards so to support further land use planning as well as future soil surveys.

As part of a project on competing claims in natural resources in the trans-frontier national park areas of Mozambique, South Africa, and Zimbabwe (Giller et al., 2008), we were tasked with assessing soil resources in the LNP. SOC is a good indicator of livelihoods and ecosystem function (FAO, 2001). Our study therefore focused on mapping SOC stocks (Cambule et al., 2012, 2013, 2014). In the process of reviewing existing information on the park area, we were surprised by the number and variety of available legacy surveys. To structure the use of this legacy data, the methodology proposed by Rossiter (2008) for data rescue and renewal was applied and tested. Our specific objectives were: (1) to undertake "data archaeology" to locate and catalog relevant surveys and select and rescue the most promising ones into archival digital format, (2) to renew selected surveys following the DSM approach (Rossiter, 2008), (3) to make quantitative and qualitative inferences of SOC stocks as a test of data quality for mapping and monitoring, and (4) to assess renewal quality of renewed surveys, document all steps and evaluate the applicability of the Cornell Adequacy Criteria (CAC) for soil surveys (Forbes et al., 1982) in combination with recent computer and technological developments.

2. Material and methods

2.1. The study area

The 276 km^2 study area is located along the Elefantes River (Fig. 1) in the southern part of the 10 400 km^2 LNP in southwest Mozambique near the Massingir Dam (spillway 23° 52′ 27.12″ S, 32° 08′ 43.80″ E). We considered four legacy soil survey areas: Mavodze (north edge of Massingir Dam's reservoir), Massingir, Chibotane, and Chinhangane downstream of Massingir Dam. The area is targeted for resettlement. The warm arid climate in LNP is characterized by dry winters, a mean annual temperature exceeding 18 °C, and an annual rainfall of about 500 mm (Ministerio do Turismo, 2003; Stalmans et al., 2004). Dominant geological features are (1) sedimentary rocks (limestone, sandstone) where the sand mantle has been exposed closer to waterways, (2) colluvium-filled lowlands along the undulating slopes, and (3) alluvial deposits along main drainage lines. Soils derived from sedimentary rocks are deep, well-structured, and coarse-textured; while those soils derived from alluvium are clayey (Stalmans et al., 2004). Almost all soils have low (<2%) SOC concentrations. The area is dominated by "mopane" vegetation with *Colophospermum mopane* typical of the Sudano-Zambezian region, along with *Terminalia* and *Combretum* woodlands, on higher topographic positions. *Acacia* and *Ficus* spp. are found on lower positions (Stalmans et al., 2004).

2.2. Methodology

We began with "data archaeology" (objective 1) to investigate the history of soil surveys in the area and to locate them. These surveys were described and grouped by authors, production year, location, objectives, extent, scale, and number of profiles. We selected the most promising soil surveys for (1) rescue and renewal exercise and (2) serving as SOC stocks baseline for soil quality monitoring in nearby resettlement areas. Poorly hand-sketched surveys were not used. The selected surveys were "rescued", i.e., converted into archival digital format, by scanning the paper maps and reports. We then followed the renewal steps (objective 2) proposed by Rossiter (2008) and, at each step we assessed the quality of legacy data (objective 3) using relevant adequacy criteria (Forbes et al., 1982; Goodchild and Hunter, 1997). Renewed maps with unsatisfactory results did not meet current standards and therefore required supplemental field survey before use. We did not carry out the final step of Rossiter's (2008) recommended procedure, i.e., integration of the renewed maps into an easily-accessible Spatial Data Infrastructure (SDI), see for example Hendriks et al. (2012), since there was no target SDI. Such target SDI is crucial for effective sharing of primary data, derived information and products and is currently being developed (Batjes et al., 2013). Lastly we inferred the SOC stocks (objective 4) qualitatively and qualitatively. Sections 2.2.1–2.2.4 detail the methodology followed for objectives 2 and 3 (carried out in parallel), while Section 2.2.5 details the methodology for objective 4. Section 3 follows the same structure.

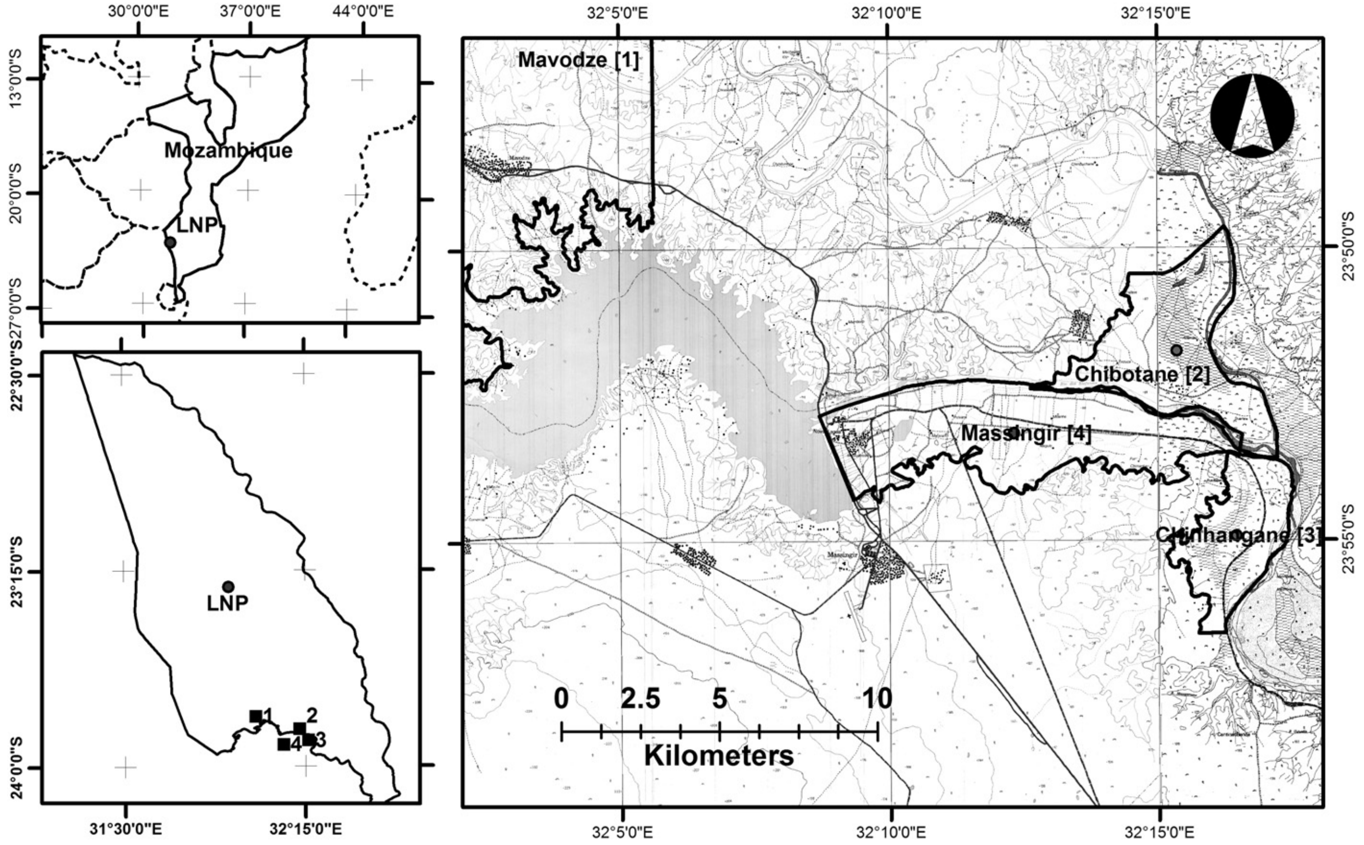

Fig. 1. Location of study area within and around Limpopo National Park (LNP, Mozambique, centroid: 23° 18′ 55.57″ S, 31° 55′ 16.24″ E), showing the four selected legacy soil survey area of Mavodze (centroid: 23° 49' 20.28′S, 31° 58′ 25.88″ E), Chibotane (centroid: 23° 57′ 44.68″ S, 31° 15′ 18.59″ E), Massingir (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E), and Chinhangane (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E).

2.2.1. Geodetic control

The first step in legacy data renewal is to improve the geodetic control, a major deficiency of many legacy survey maps. Local coordinate systems are often used instead of a proper coordinate reference system (CRS) that includes coordinate system, projection, and datum (Iliffe and Lott, 2008). Some detective work was required to identify the base map over which the legacy soil surveys were printed. The clues were cultural features, road intersections, and contour lines shown on the soil map and base maps. The base maps were then used to identify the CRS. Control points visible on the soil maps were identified on georeferenced topographic maps and remote-sensed imagery with known CRS and were used to georeference the scan. The quality of this step was assessed by the absolute Root Mean Square Error (RMSE) of the control point transformation and by the RMSE normalized by the map scale.

2.2.2. Creation of GIS layers

The second step in legacy data renewal is the conversion of the scan to GIS layers. The boundaries of the soil units and the locations of observation points were manually digitized and the attribute table was populated with the attributes from the original map and report. Lines were digitized through the middle of lines (soil unit boundaries) and points (soil profiles location) on the magnified scanned maps to reproduce the geometry of the original maps. Topology was generated from the linework. However, as is typical for renewal exercises, we could not locate the original master maps on stable mylar. Therefore, we had to work with paper prints in various states of preservation and folding. The errors in digitization are probably small compared to the errors in the scan and georeferencing of the source material. The renewed polygon soil maps were subjected to quality assessment following the adequacy criteria: (a) scale and texture, and (b) legend (Forbes et al., 1982). We did not assess the adequacy of the point coverage.

a. Scale and texture. The scale and texture of the soil map were evaluated to assess the legibility and its capacity to represent the smallest area of interest using the definitions of Forbes et al. (1982). These include the Minimum Legible Area (MLA), which indicates the smallest land area that can legibly be represented on the map at its published scale, here using the Cornell criterion of a Minimum Legible Delineation (MLD) of 0.4 cm^2. In the process of map renewal delineated areas smaller than the MLA need to be aggregated into larger units. Maps were also assessed by the Index of Maximum Reduction (IMR), which is the factor by which the map scale can be reduced before the Average Size Delineation (ASD) equals the MLD, i.e., before half of the map would become illegible. The IMR is computed as the square root of the ratio of the ASD to the MLD and reveals whether the chosen map scale matches the actual delineation sizes. An IMR of 2.0 is considered optimal (Forbes et al., 1982); at this ratio the ASD is 1.6 cm^2, i.e., four times the MLD. A large IMR means paper was wasted, i.e., the map is not as detailed as its scale indicates. A small IMR means the map is illegible and the intensity of mapping can support a larger scale. This factor can be used to adjust the scale of renewed maps, provided the sampling intensity would support a larger scale. Note that with digital displays a polygon map can be displayed at any scale, but there is always a design scale, which should match the areas to be represented.

b. Legend. The map legend identifies the map units, generally referring to a full description in the associated survey report, and may also provide a brief description and various interpretations. Identification is based on the symbols printed inside the map unit polygons. The descriptive legend provides information in narrative or tabular form about each map unit. Map unit names and definitions in descriptive and interpretive legends determine the amount and usefulness of information. Map legends may be evaluated either in terms of specific use of the soil inventory or by a more general criterion, such as a

Table 1
Legacy soil data inventory for the Limpopo National Park, Mozambique (LNP, 23° 18′ 55.57″ S, 31° 55′ 16.24″ E) and surroundings.

Item	Legacy data	Location	Objectives	Size (ha)	Nr soil profiles	Scale
1	Casimiro and Veloso (1972)	Banga-Marreguele Confluence Elefant/Shingwedzi (centroid: 23° 50′ 33.77″ S, 32° 15′ 53.49″ E)	Extension of an earlier surveyed area along right margin of Elephant River	750	50	1:20 000
2	Casimiro and Veloso (1969)	Chibotane/Machaule Confluence Elefant/ Shingwedzi (centroid: 23° 50′ 33.77″ S, 32° 15′ 53.49″ E)	Planning for resettlement of communities then to be affected by the filling of Massingir Dam (then to be built).	4400	520	1:20 000
3	COBA Consultores (1981)*	Massingir, downstream Massingir Dam, along the right margin of Elephant River. (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E)	Land suitability evaluation (for irrigation)	1157.7	22	1:10 000
4	COBA Consultores (1982)*	Chinhangane, right margin of Elefantes River, next to the COBA Consultores (1981) (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E)	To increase agricultural production for communities resettled 5–6 kms around the Massingir Dam, then affected by the filling of the reservoir	1150	14	1:10 000
5	COBA Consultores (1983a)*	Chibotane-Machaule-MadinganeConfluence Elefant/Shingwedzi (at 23° 50′ 33.77″ S, 32° 15′ 53.49″ E and nearby)	Land suitability evaluation (for irrigation) to select areas to benefit from water flowing from Massingir Dam to the benefit of the local communities	2158.2	25	1:10 000
6	COBA Consultores (1983b)*	Mavodze, Massingir-velho, Cubo, Paulo Samuel Kankhomba; northern side of the Massingir reservoir and the remainder at the southern part of the same reservoir (centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E)	Land suitability evaluation to base future land developments towards increasing agricultural production for communities resettled 5–6 kms around the Massingir Dam, then affected by the filling of the reservoir	33 000	25	1:50 000
7	Rural Consult Lda (2008)	Between Chinhangane and Banga villages, along the right margin of Elephant river at about the large meander (around 23° 54′ 55.24″ S, 32° 16′ 23.92″ E)	To study the pedology and assess grazing potential	6000	6	1:50 000

* Selected legacy survey for present study.

soil classification system. We evaluated the map units' information (description) in terms of the soil classification used in the legacy survey. Information was considered adequate if the map unit description includes the diagnostic information (horizons, properties) or the classification. The overall information quality of the soil survey was expressed by the proportion of land units or survey area evaluated as "adequate" relative to the total number of units or total area.

2.2.3. Integration of auxiliary information

The third step of legacy data renewal is the integration of auxiliary information, notably remote sensing products. These include original images (visible, near infrared, radar, and thermal) and derived products such as land cover, vegetation intensity, and terrain parameters. Soil geography is in part related to vegetation and topography. Therefore, borders of some soil units may correspond in part to limits observed on the remote sensing products. For such borders, an overlay of the digitized and georeferenced soil map and remote sensing products can show their displacement. We used a normalized difference vegetation index (NDVI) and an unsupervised land cover classification, both derived from a Landsat Thematic Mapper (TM) image, along and with a DEM derived from the Shutter Radar Topographic Mission (SRTM) imagery to check the displacement of soil units. In the case of large differences the borders of the mapping units were edited.

Table 2
Major characteristics of legacy soil surveys.

Characteristic	Description
Currency	Although most of the surveys were reported in the 80s, few date back to the late 40s–60s
Type	These "soil maps" are diverse and they go from a "sketch with simple legend" to somewhat complete map with legend, soil profile description and laboratory data
Scale	Most maps were on scales of 1:10 000 and 1:20 000, few at 1:50 000
Format	These maps are printed (hard) copies, drawn over local grids with no reference to any geodetic control and in many cases with different procedures/standards
Use	Most of these are shelved and seldom used

Multispectral satellite imagery (Landsat TM, 30 m resolution) from the end of the wet season (February 2010, row/path: 168/076) was obtained from the United States Geological Survey (USGS) website (www.usgs.gov), preprocessed as a Level 1 Terrain corrected (L1T) product; this provides radiometric and geometric accuracy, corrected for topography. Contrast enhancement was performed to increase the distinction between the features on the classified image to facilitate its visual interpretability. Unsupervised land cover classification was performed on the enhanced image, specifying the same number of classes as the maximum number of map units in the most detailed survey (Chibotane).

The DEM seemed particularly useful since most map units in the selected surveys represented physiographic units. A three arc-second (approximately 90 m) resolution DEM from the Shuttle Radar Topographic Mission was downloaded from the Jet Propulsion Laboratories (JPL) website (www.jpl.nasa.gov), preprocessed to research grade, and was used to derive contours to check boundaries declared in the soil surveys to correspond to contours. A simple positional accuracy measure (Goodchild and Hunter, 1997) was then used to evaluate the boundary displacements on the legacy map: the proportion of the total length of the digitized contour that is within a specified distance of the high accuracy representation (DEM derived contour).

2.2.4. Metadata

The fourth step is the development of appropriate metadata that includes identification information, spatial reference, attributes, information on data quality, and a description of the methods used during the original survey and the renewal exercise. In addition, the metadata should also explain the semantics e.g., soil type and soil properties. We used the Federal Geographic Data Committee (FGDC) metadata standards through the FGDC metadata editor in ArcGIS to create metadata for each soil survey.

2.2.5. Inference of SOC stocks

Finally, we evaluated the legend in terms of what information it gives explicitly or implicitly (e.g., via the soil classification or topsoil properties) about SOC concentration and stocks. The required information was extracted from either map unit descriptions or point

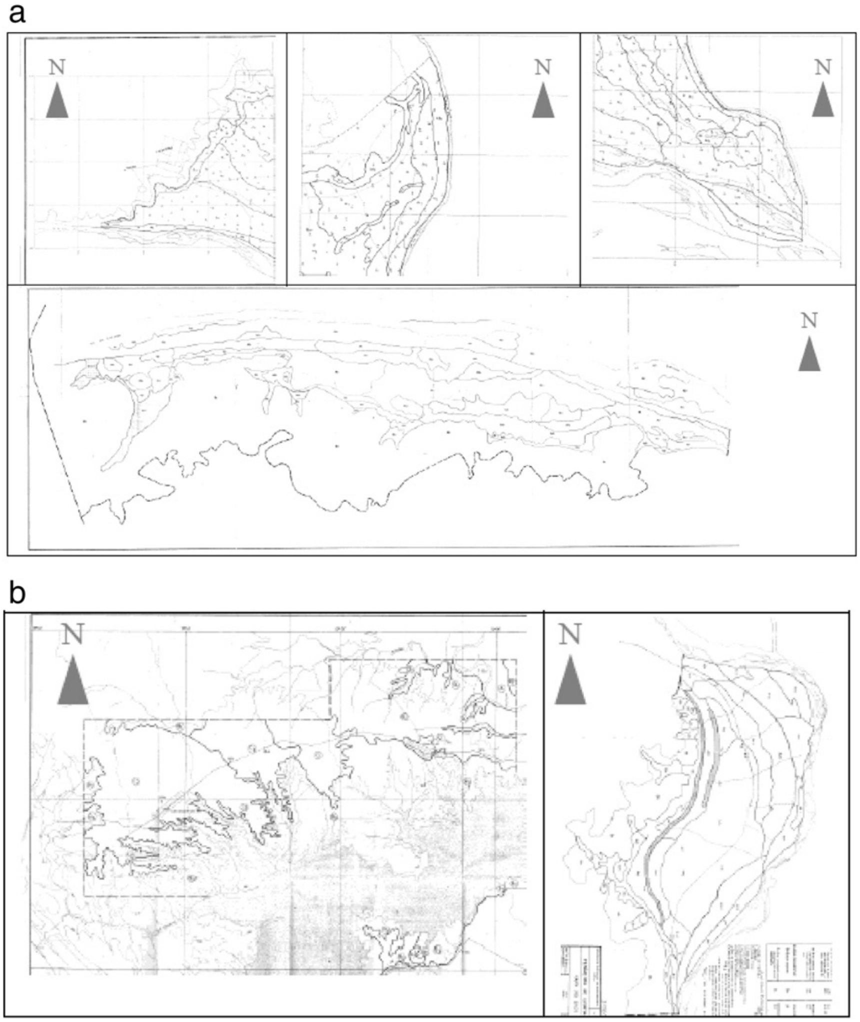

Fig. 2. a: Rescued (scanned) legacy soil maps of Chibotana, Mozambique (top; three map sheets (left; centroid: 23° 51′ 57.05″ S, 32° 14′ 12.46″ E, middle; centroid: 23° 50′ 33.77″ S, 32° 15′ 53.49″ E, right; centroid: 23° 52′ 43.07″ S, 32° 16′ 26.92″ E)) and Massingir, Mozambique (bottom, (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E). b: Rescued (scanned) legacy soil map of Mavodze, Mozambique (left; centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E)) and Chinhangane, Mozambique (right; centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E).

observations. While the former yielded a qualitative result, the latter yielded quantitative estimates, following the measure-and-multiply approach (Thompson and Kolka, 2005), making use of data populated in the attribute tables of both area-class and point data: SOC concentrations at points considered to represent a map unit, combined with soil bulk density (BD) and A-horizon thickness reported for map units, multiplied by the map unit area. Previous work (Cambule et al., 2014) has shown that SOC stocks are primarily in the A-horizon. Further, the concentrations are inversely proportional to A-horizon thickness, so that stocks are primarily controlled by climate. We applied both approaches to allow comparison of SOC stock estimates inferred independently.

3. Results and discussion

3.1. Data archeology and rescue of legacy surveys

3.1.1. Data archeology

The first soil map covering Mozambique was part of a reconnaissance report on the soils and vegetation of Africa (Shantz and Marbut, 1923). This 1:25 000 000 map was frankly described as a guess, based on regional climate and a few very scattered soil observations, of the possible soils. Upland Mozambique was simply described as “Natal red loams”. The Great Soviet World Atlas (Gorkin et al., 1937) included a

Table 3
Quality of improved geodetic control as assessed by the Root Mean Square Error (RMSE) of georeferencing. Added are the number of ground control points (GCP), map scale, maximum location accuracy and Minimum Legible Area (MLA).

Legacy survey	RMSE (m) (first-order polynomial)	Nr GCP	Map scale	Maximum location accuracy at scale (m)	MLA (ha)	Side length of MLA (m)	RMSE proportion of side length MLA
Mavodze	56.92	20	1:50 000	12.5	10	316.23	0.18
Chibotane 1	10.77	9	1:10 000	2.5	0.4	63.25	0.17
Chibotane 2	26.32	8	1:10 000	2.5	0.4	63.25	0.42
Chibotane 3	8.14	4	1:10 000	2.5	0.4	63.25	0.13
Massingir	28.64	19	1:10 000	2.5	0.4	63.25	0.45
Chinhangane	24.16	12	1:10 000	2.5	0.4	63.25	0.38

GPS coordinates: Mavodze (centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E), Chibotane 1 (centroid: 23° 51′ 57.05″ S, 32° 14′ 12.46″ E), Chibotane 2 (centroid: 23° 50′ 33.77″ S, 32° 15′ 53.49″ E), Chibotane 3 (centroid: 23° 52′ 43.07″ S, 32° 16′ 26.92″ E) and Chinhangane (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E).

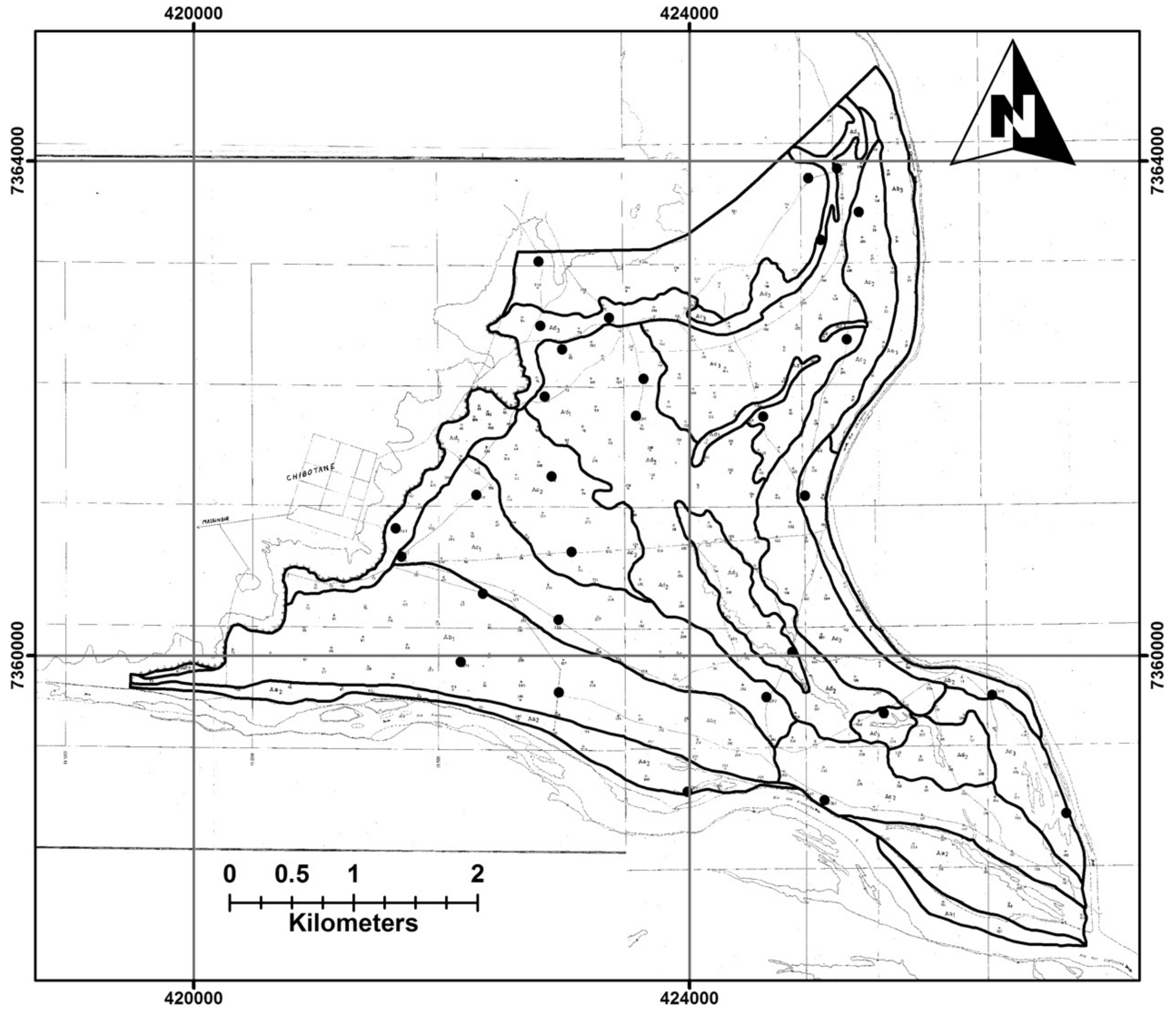

Fig. 3. Improved geodetic control of combined three legacy soil map sheets of Chibotane, Mozambique (centroid: 23° 57′ 44.68″ S, 31° 15′ 18.59″ E) soil map (Fig. 2a, top) and, overlays of digitized GIS area-class (soil units) and point (soil profiles location).

soil map at 1:50 000 000, but based on more information. The creation of this map was described by Schokalsky (1944). The Limpopo valley was simply described as "alluvial soils"; the upland portions of the LNP as "weakly leached soils of the dry evergreen forests and bushes", a reasonable inference. Gouveia and Azevedo (1955a) derived the first nationwide soil maps of Mozambique to be derived from these sources by magnifying to a scale of 1:6 000 000 while maintaining the same map units and level of detail. Map units were delineated based on climate, elevation, and geology. Soil units were broadly characterized by soil morphology and revealed little differentiation within LNP. Nevertheless these maps had an important support role for later, more detailed, soil surveys. The same authors published three more soil maps: (1) the preliminary soil map of Mozambique at 1:4 000 000 cited by Gouveia and Azevedo (1954); (2) the provisional soil map of southern Mozambique at 1:2 000 000 (Gouveia and Azevedo, 1955a); and (3) the provisional national soil map at 1:2 000 000 (Gouveia and Azevedo, 1955b). These maps were based on the amplified Marbut and Schokalsky maps and further improved by integrating soil survey data obtained since 1947 by the technical unit for cotton suitability reconnaissance (Centro de Investigação Científica Algodoeira de Moçambique). However, the new surveys only covered the cotton-growing area. So the authors followed the Marbut and Schokalsky approach to draw the maps based on climate, elevation, and geology in areas outside the cotton areas, including the future LNP.

Roeper (1984) reports that in 1950 Ripado et al. carried out one of the first soil surveys along the Elefantes and Limpopo Rivers in an area

Table 4
Assessment of map scale and map texture through the average size delineation (ASD) and index of maximum reduction (IMR) of selected Limpopo National Park (LNP, Mozambique) legacy soil maps.

Soil map	GPS coordinates (centroid)		Map unit area (ha)			Map texture	
	Latitude	Longitude	Max	Min	Mean	ASD (cm^2)	IMR (—)
Mavodze	23° 49′ 20.28″S	31° 58′ 25.88″ E	8650.2	26.1	1438.7	287.7	26.8
Massingir	23° 53′ 08.34″ S	32° 12′ 19.31″ E	1384.7	0.8	51.5	51.5	11.3
Chibotane	23° 57′ 44.68″ S	31° 15′ 18.59″ E	310.9	2.1	84.8	84.8	14.6
Chinhangane	23° 54′ 55.24″ S	32° 16′ 23.92″ E	266.1	0.6	30.5	30.5	8.7

Table 5a
Samples of soil map legend tables of Mavodze, Mozambique, at 1:50 000 scale (centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E).

Cartographic unit	Physiography	Dominant soils
A	Alluvial plains of Singuidzi River	Eutric Fluvisols (Je) Calcaric Fluvisols (Jc)
B	Sloping and undulating sedimentary areas from the tertiary	Haplic Xerosols (Xh)
B1	Colluvial lowlands and adjacent gentle slopes	Luvic Xerosols (Xl1)
B2	Undulating to gently undulating relief	Calcic Xerosols (Xk)

of about 100–200 km^2 that included the description of 15 soil profiles. In the same survey report, the "Brigada de estudo de solos" is mentioned to have surveyed the soils of Massingir District in 1964 over an area of about 250 000 ha. Their results supported the survey by Casimiro and Veloso (1969) that carried out a 1:20 000 soil survey along the left margin of the Elefantes River upstream of the confluence with the Singuedzi River covering an area of about 4400 ha. The survey includes the description of 520 soil profiles and identified 28 map units, unfortunately without georeference for either. The same authors are cited by Roeper (1984) to have surveyed both margins of the Elefantes River in 1972 covering a total of about 26 000 ha at 1:10 000 in three different reports: (1) Magajamele-Maguça, (2) Maguça-aldeia da barragen and (3) Marrenguele-Banga. We were only able to recover the report of the last survey (Casimiro and Veloso, 1972).

Roeper (1984) also reports that in 1971 Gouveia and Marques published a soil map of Mozambique at 1:5 000 000 in preparation for the FAO-UNESCO soil map of the world (FAO, 1974). The first 1:4 000 000 soil map of Mozambique was published by Gouveia and Marques (1973).

Soil surveys were then discontinued due to the increasing armed conflict before Mozambique's independence in 1975. In the first decade after independence, southern Mozambique faced repeated periods of flooding and drought. The resulting shortage in food supply resulted in widespread hunger and malnutrition. These problems stimulated the government of Mozambique to improve the agricultural infrastructure, mainly water reservoirs and irrigation systems. To this end Roeper (1984) reports that in 1978 Priporski surveyed at 1:20 000 in an area of about 2700 ha around the Massingir Dam and its irrigation schemes. These surveys were carried out as part of the relocation program for people that would be affected by the filling of the Massingir Dam's reservoir. People were relocated in seven communal settlements (Mavoze, Massingir, Chibotane, Machaule, Chinhangane, Cubo, and Paulo Samuel Kankhomba) of which four are located within today's LNP (COBA Consultores, 1981, 1982, 1983a,b). The irrigation systems were not properly managed by the beneficiary communities, with the result that the systems quickly deteriorated and were abandoned. Roeper (1984) also reports that in 1981 Sinadinov carried out a study of soil salinity problems in a new irrigation scheme for citrus orchards along the left margin of the Elefantes River. Land development projects were abandoned thereafter due to the insecurity caused by the civil war (1977–1992).

After the civil war, a 1:1 000 000 national soil map (INIA, 1995) was compiled on the basis of existing soil survey studies and supported by satellite image interpretations to extrapolate in areas for which limited soil information was available. In addition, Dijkshoorn (2003) compiled a soil and terrain database at 1:2 000 000 and legacy soil maps from ISRIC were included as rescued digital scans in the European Digital Archive of Soil Maps (EuDASM) project conducted in collaboration with ISRIC (Selvaradjou et al., 2005).

Stalmans et al. (2004) were commissioned to survey the resources of the newly-established LNP. They did not directly survey the soil resource, did not delineate soil mapping units, and did not report any point observations of soil properties. Instead the authors relied on the 1:1 000 000 national soil map and used ecological units, which they mapped at 1:250 000 (Cenacarta, 1997).

Despite the current peaceful and stable conditions in Mozambique, the only systematic soil surveys is an ongoing reconnaissance (1:250 000 nominal map scale) survey (no published results as yet), within the agro-ecological zoning project, outsourced by the Ministry of Agriculture to a private company (Rural Consult, Ltd.). The first phase of the Africa Soil Information Service (AfSIS) project (http://www.africasoils.net/) has generated soil property maps at a 1 km resolution (ISRIC, 2013) based on legacy data rescue and renewal operation for over 16 500 soil profiles (http://www.africasoils.net/data/legacyprofile), however, the LNP is not specifically targeted by these projects.

3.1.2. Selection and rescue

Data archaeology uncovered six detailed legacy soil surveys in the LNP and its vicinity as well as some reconnaissance maps (Table 1). The major characteristics of these legacy surveys are summarized in Table 2. Two surveys (Chibotane and Mavodze) within the LNP were selected to illustrate the renewal process. An additional two surveys (Massingir and Chinhangane) located downstream the Massingir Dam and along the right margin of Elefantes River outside LNP were selected to represent the baseline for soil quality monitoring in the resettlement area. The four selected soil maps are described and illustrated in Table 2 and Fig. 2. They were rescued by scanning at 300 dots per inch (dpi) resolution for subsequent renewal steps.

3.2. Renewal and quality assessment of legacy survey

3.2.1. Geodetic control

The four survey maps provide different information on their georeferencing. The Mavodze map uses a grid in geographic coordinate system (GCS) data but gives no details of the coordinate reference system. The Chibotane map shows a local kilometer grid. The Massingir and Chinhangane maps show no georeferencing information but they do show contour lines forming part of their borders. The contour information allowed us to identify the base map used for all four soil maps; the

Table 5b
Samples of soil map legend tables of Chinhangane, Mozambique, at 1:10 000 scale (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E).

Physiography	Cartographic unit	Dominant soils	Areal extent	
			ha	%
A (alluvial plain)		('Modal') Eutric Fluvisols	37.4	3.2
Aa — flooding area….	Aa1[*]			
	Aa2	Je1	52.5	4.5
	Aa3	Je2 $g_{2,3}$ Je3$g_{3,4}$; Je4$g_{3,4}$[**]	57.0	4.9
B (sedimentary zone)		Colluvial (Jb)	43.1	3.7
Ba — colluvial lowlands	Ba	Jb		

[*] Aa1 — the integer numbers 1, 2, 3 indicate textural class.
[**] Je — major soil grouping (FAO, 1988), "g" — coarse texture at depth specified by the subscript "3,4" (3 = 60–90 cm and 4 = 90–120 cm).

Table 5c
Samples of soil map legend tables of Chibotane, Mozambique, at 1:10 000 scale (centroid: 23° 57′ 44.68″ S, 31° 15′ 18.59″ E).

Cartographic Unit	Pedological Unit		Areal extent	
	Dominant	Sub-dominant	ha	%
Aa — flooding área	Eutric Fluvisols (Je)	Eutric Fluvisols (Je)		
Aa1[*]	Je1	Je1A	26.2	1.2
Aa2	Je3$g_{3,4}$[**]; Je3	Je3g; Je2g	153.3	7.1
Ab — marginal área				
Ab1	Je1A; Je1	Je2g	324.7	15.0
Ab2	Je2, Je2g_2; Je2g_3	Je3; Je3g_2; Je3g_3	49.5	2.3
Ab3	Je3	Je4	67.2	3.1

[*] Aa1 — the integer numbers 1, 2, 3 indicate textural class.
[**] Je — major soil grouping (FAO, 1988), "g" — coarse texture at depth specified by the subscript "3,4" (3 = 60–90 cm and 4 = 90–120 cm).

Table 5d
Samples of soil map legend tables of Massingir, Mozambique, at 1:10 000 scale (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E).

Cartographic Unit	Geology	Topography	Drainage	Soils*		Soil characteristics**						
				Dominant	Sub-dominant	(1)	(2)		(3)	(4)	(5)	(6)
							(2.1)	(2.2)				
Aa1	Aa		Less frequent flooding …	1	1A; 2v$_2$; 2 g$_{2,3}$**							
Aa2				2 g$_{2,3}$*								
…												
Aa6				5; 5 g$_{3,4}$	1A; 2v$_{2,3}$; 2 g$_{2,4}$ 3 g$_{2,3}$; 4 g; 1A							
Ba	Ba			Jb	–							
Bb	Bb			Xv$_1$; Xv$_2$	Qa; Rm							

* Soils in 2 g$_{2,3}$: 2 — surficial texture, g — sub-surficial (coarse) texture (lowercase “h, g, p, v or c”) and the depth range (subscript “1, 2, 3, 4 or 5”).
** Soil characteristics as follows: (1) Rooting depth; (2) Texture class (2.1 Surface 2.2 Subsurface); (3) CU water; (4) Permeability; (5) Salinity; (6) Sodicity.

1:50 000 topographic map of the national map series. This map uses the UTM projection and coordinates in UTM zone 36S projected on the Clarke 1866 ellipsoid. First, a large number of ground control points (GCPs) including cultural features and road intersections that were visible on both the soil and topographic maps were identified. With these GCPs the maps were transformed using a first-order polynomial transformation. The three map sheets covering the Chibotana area were mosaicked into a single map. Table 3 shows the quality of improved geodetic control for all soil maps. The RMSE represents well the average error of the geometric accuracy (Hughes et al., 2006). A large number of GCPs were needed to transform the maps and achieve the best achievable RMSE. For all four surveys the relative georeferencing RMSE was a substantial proportion of the side length of MLA (square geometry) (Table 3) ranging from 13% to 45%. RMSE is also 4 to 11 times the maximum location accuracy for corresponding scale. As such the geodetic control is poor, most likely due to the poor quality of the paper maps available to us.

3.2.2. GIS coverages

Fig. 3 shows the georeferenced and geodetically corrected soil map of Chibotane, the GIS coverage of soil units and soil profile locations, after digitizing. Manually digitizing the soil units is a tedious and time-consuming task. One way to minimize the work would be to use automated feature extraction from the scanned linework, but these algorithms have difficulty with finding soil boundaries on a map with other line features such as roads and contours. Attribute tables of point data were populated with profile number, associated map unit ID, soil type, A-horizon depth, SOC concentration and BD (available only in Chinhangane). Attribute tables of soil units were directly populated with the soil unit code, representative soil profile ID, and area. For the other variables average values for the points within each delineated area were computed and attributed to the polygon. For polygons without representative profiles we used the average of the mapping unit to which the polygon belongs. It would have been possible to use this approach also for polygons with representative profiles; however, since we were uncertain whether the allocation of polygons to map units was correct, we preferred to use the first approach in those cases. Since the surveys of Chibotane, Massingir, and Chinhangane share the same legend, representative profile data was shared over these three survey areas. There were a few mapping units without representative profiles, in which cases we used the profiles from the most similar map unit (determined by map unit description) from the same or nearby survey area.

3.2.3. Quality assessment

The Cornell guidelines (Forbes et al., 1982) were used to assess the quality of the legacy map data or SOC mapping and monitoring. These guidelines evaluate the scale and texture of the map as well as its legend.

3.2.3.1. Map scale and texture. The sizes of the delineated areas in the Massingir soil survey are mostly between 5 and 15 ha. Table 4 shows the MLA and IMR (see definitions in Section 2.2.2.a) for the soil maps. All delineations of all surveys are larger than the MLA, which indicates that the legacy soil maps meet the standard in this regard. However, all IMR are well above the optimal value (2.0), meaning that legacy maps are legible but scale could be substantially reduced before losing legibility, implying that renewed maps should be printed at the reduced scale. The scale could be reduced by half the IMR (resulting in an optimal IMR, i.e., 2), which would result in a scale of about 1:1 175 000 (Mavodze), 1:55 000 (Massingir), 1: 75 000 (Chibotane) and 1:45 000

Table 6
Assessment of map unit definition and overall information quality of soil survey for the selected legacy soil maps from Limpopo National Park (LNP, Mozambique) (centroid: 23° 18′ 55.57″ S, 31° 55′ 16.24″ E). In between brackets the total number of soil profiles in each survey area.

Variable	Approach	Diagnostic criterion	Units	Mavodze (10) (centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E)		Chibotana (30) (centroid: 23° 57′ 44.68″ S, 31° 15′ 18.59″ E)		Massingir (20) (centroid: 23° 53′ 08.34″ S, 32° 12′ 19.31″ E)		Chinhagane (14) (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E)	
				nr	ha	nr	ha	nr	ha	nr	ha
Map unit definition	General map unit information	Unambiguously placeable in a taxonomic class	Cartographic (mapping units)	7	21,581	12	2036	28	2833	18	1190
			Total delineated units	15	21,581	24	2036	55	2833	39	1190
			A (sampled)	15*	19,567	15	1816	13	2268	7	534
			A (not sampled, with representative profile)	8	2014	6	154	23	334	23	491
			NA (not sampled, without representative profile)	0	0	3	154	19**	231	9	166
Overall information quality	Proportion of “adequate” land size	80% or more “adequately defined”	% adequately defined	100	100	88	97	66	92	77	86
			Evaluation	A	A	A	A	NA	A	NA	A

A: adequately define, NA: not adequately defined.
* No lab data in one soil profile.
** One of them could not have a replacement.

Table 7
Improved (Restructured and unified) legend for Mavodze (centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E), Massingir (centroid: 23° 53′ 08.34″ S, 32° 12′ 19.31″ E), Chibotane (centroid: 23° 57′ 44.68″ S, 31° 15′ 18.59″ E), and Chinhangane (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E) soil surveys (all within or around LNP, Mozambique).

Land type and position	Physiography & flooding frequency	Top-soil texture	Cartographic unit	Soil association		Description	Extent	
				Dominant	Sub-domin		ha	%
A	Aa	1	Aa1	Je1	Je1A, …			
		…	…					
		6	Aa6					
	Ab	1 … 6	…					
	….	…	…					
	Ae	1 & 2	…					
B	Ba	–	Ba					

(Chinhangane). Clearly the intensity of survey information supports the map scales as reported in the survey. Some of the large IMR may be due to large areas of homogeneous soils at the chosen categorical level; in this case the categorical level could have been reduced to show, soil observation density permitting, finer distinctions (e.g., by establishing soil series or phases) or the map could be reduced. Such a reduction in scale would also affect the maximum positional accuracy of boundaries, which is equivalent to half of the conventional minimum delineation width (3 mm); e.g., the re-scaled Chibotane map implies a maximum boundary precision of 112.5 m, whereas the original scale implies 15 m. Given that the RMSE of georeferencing is on the same order (8–26 m, Table 3), the boundary precision implied by the original scale seems reasonable in this respect.

3.2.3.2. Map legend. Map units are labeled and categorized in map legends (Tables 5a–5d). All four legends in this study are based on physiographic elements at higher levels and pedological aspects at lower level. The general soil description in each map unit is included in the report, in most cases along with representative profiles. In the Mavodze soil survey, map unit definition criteria form a two-level hierarchy: (1) physiographic units and (2) association of FAO (1974) soil classes. The two aspects are linked as shown in Table 5a.

The Chinhangane, Chibotane, and Massingir soil maps, all at 1:10 000 scale, share the same legend, with minor variations (Tables 5b–5d). The most detailed legend is that of Massingir, which describes map units in terms of geomorphology, topographic characteristics, drainage conditions, soils (dominants and sub-dominants) and soil characteristics (thickness, surficial and sub-surficial texture, available water capacity, soil permeability, soil salinity and sodicity), see Table 5d.

Table 6 shows the evaluation of map unit definitions. Since map units are mostly associations of pedological units, we considered a map unit to be “adequately defined” only if all members had the same positive evaluation result. The inclusion of soil classification in map unit descriptions should have made all units adequately defined, however, we were not confident of names assigned to map units which did not contain a representative profile. As such we question the adequacy criterion “unambiguous placement in a taxonomic class” as suggested by Forbes et al. (1982). Therefore, those units were considered “not adequately defined”. The overall information quality of the surveys is “adequate” (>80%) for all surveys when the “proportion/percentage of area-size of adequately defined” map unit description is considered rather than the “proportion/percentage of number of delineated units adequately defined”.

Table 8
Simple positional measure for the 110 m above sea level (masl) of Massingir, Mozambique (centroid: 23° 53′ 08.34″ S, 32° 12′ 19.31″ E) and for 90 masl of Chinhangane, Mozambique (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E) soil border segments, respectively, for target of 99th percentile.

Iteration (i)	Massingir, test length: 17,578 m			Chinhangane, test length: 11,408 m		
	Buffer (m)	Length within (m)	Proportion (−)	Buffer (m)	Length within (m)	Proportion (−)
0	0	0	0	0	0	0
1	2.5	177	0.010	2.5	196	0.017
2	246	13,351	0.760	144	6915	0.606
3	321	14,902	0.848	236	9792	0.858
4	441	15,970	0.909	285	10,525	0.923
5	603	16,842	0.958	335	11,130	0.976
6	707	17,093	0.972	349	11,408	1.000
7	834	17,424	0.991	343		
8	826					

The Mavodze legend (Table 5a) has two conceptual flows: first, the cartographic unit entry reflects two hierarchical levels, which could well be separated in two different hierarchical levels (separate columns), to show the dichotomic subdivision of “unspecified” into physiographic units within the entry, since it is clear that, in general the “unspecified” units “A” are associated to Fluvisols, “B1-3 (and C3)” to Xerosols, and “C1-2” to Arenosols. Second, there is no indication of the proportions of pedological units forming each association, creating confusion for the map reader about which soil is more likely to be found when the same pedological unit appears in a different physiographic unit where it is part of a different association. The proportion of the different members within soil association, and thus its homogeneity, can be modeled and estimated but cannot be reliably assessed without additional fieldwork.

When a legend is built as a hierarchy, lower-level categorical units are easily interpreted within higher categories. A single harmonized legend structure for the three surveys, which integrates those from detailed surveys (Chibotane, Massingir, and Chinhangane) into the less detailed one (Mavodze), can be produced. At the higher level, the harmonized legend (Table 7) results from the separation of the “unspecified” from physiography and its upgrading to a different and highest hierarchical level of Mavodze's legend and naming it as “Land type and position”. The harmonized legend is then completed by integration of the Massingir legend “physiography and flooding frequency” into second level. This level results from amalgamation of the terms “geomorphology”, “physiography”, and “flooding frequency” used across the detailed survey legends. At the third level, “Topsoil texture” was adopted as defined in the Massingir legend. The three hierarchical levels just described were then used to define “cartographic units” further characterized in terms of “soil associations” (“dominant” and “sub-dominant” soils), their description, spatial coverage, and proportion. The improvements shown in Table 7 do not change the assessment of both map unit definition and overall legend information quality.

The FAO (1974) legend used in this survey is outdated. In 1990 the FAO legend was revised (FAO, 1988), at which time the major soil grouping of Xerosols was eliminated, since climate was no longer considered a soil classification criterion; most Luvic Xerosols were reclassified as Orthic Luvisols but the Haplic Xerosols could be Haplic Regosols or Haplic Cambisols depending on the presence of a Cambic horizon. Both FAO legends have been superseded by the World Reference Base for Soil Resources (IUSS Working Group WRB, 2006) which has similarities to the 1988 legend but introduces new concepts and some revised names (for example ‘Orthic’ is replaced with ‘Haplic’), so that for a proper renewal all legacy soil units should be reclassified based on the profile descriptions, supplemented if necessary with inferences from the 1974 names. However, since at the time we undertook this study a new version

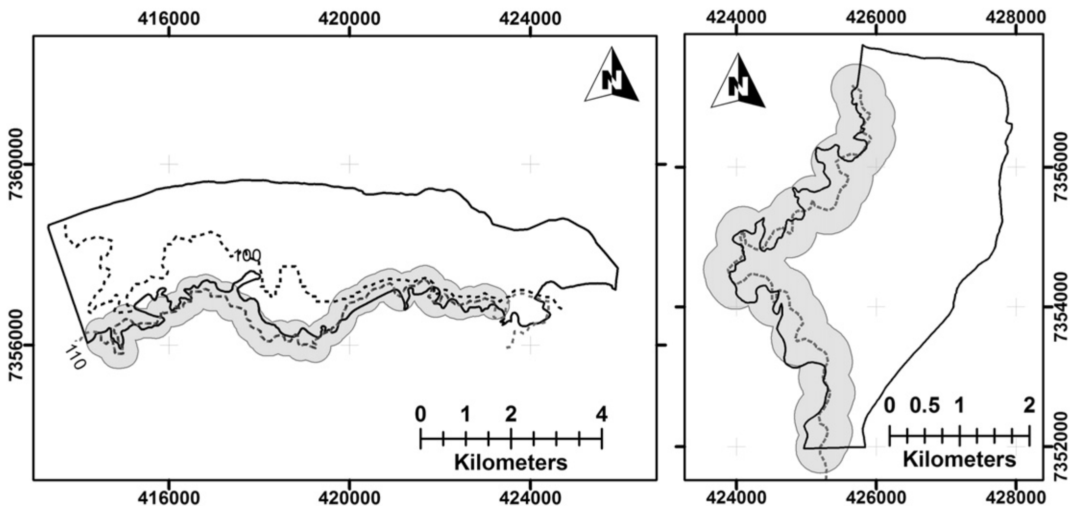

Fig. 4. The 100 and 110 mean sea level (masl) contours extracted from shuttle radar topographic mission (SRTM) digital elevation model (DEM) overlaid onto Massingir, Mozambique (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E) soil map (left) and the 90 masl contour overlaid onto Chinhangane, Mozambique (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E) soil map (right), whose southern (Massingir) and western (Chinhangane) borders were considered to be 100 and 90 masl contours, respectively. Grey buffer of 320 m around the 110 m (Massingir) and buffer of 349 m around the 90 m (Chinhangane) DEM derived contours.

of the WRB had been proposed for adoption in 2014 (and was subsequently adopted at the World Soil Congress in June 2014), we intend to review that version and then reclassify.

3.2.4. Integration of RS and DEM layers

Contours were used to define the extent of the Massingir (100 mean height above sea level (masl) contour corresponds to the southern limit) and Chinhangane (90 masl contour corresponds to the western limit) soil surveys. These contour lines were compared to those derived from the SRTM DEM (Table 8). Fig. 4 clearly shows that there is a large mismatch between the drawn and SRTM DEM derived 100 masl contours in Massingir. The drawn 100 masl is much closer to the 110 masl SRTM DEM derived contour than the 100 masl one, casting doubt on the accuracy of all other lines drawn on the map. The tested length spanned between two crossing points of the contour and soil map borders and, in the absence of local map accuracy standards, the maximum location accuracy (2.5 m at 1:10 000 map scale) was used to start the iteration as suggested by Goodchild and Hunter (1997). Linear interpolation in results (Table 8), show that in both tests, less than 30% of border length falls within a buffer length of the square root of MLA, thus the positional accuracy is poor. Only at a buffer length of 349 m do all the tested borders in the Chinhangane map fall in, while only about 90% of tested borders in the Massingir map fall in for a similar buffer length (321 m). Therefore, replacing the mapped soil borders by their high accuracy representation from contours would represent a substantial improvement for both surveys. A complication is that several internal borders are connected to the Chinhangane boundary contour. This 90 masl border thus must be revised along with the intersecting soil unit borders (see following paragraph). This step would be followed by the assessment of positional uncertainty of all polygon borders, taking the land cover classes' borders as high accuracy representations against which could be compared, using the approach proposed by Kiiveri (1997). This would be especially applicable to the Mavodze survey, where the land cover seems to match the physiography, which is not surprising given the fact that this area is covered by natural vegetation, in contrast to the primarily agricultural Massingir survey area.

Fig. 5 shows the classified (30 classes) and contrast enhanced (histogram equalization) subset of Landsat TM imagery of the study area onto which the semi-transparent SRTM DEM plus the soil map of Mavodze (top) and Massingir (bottom) are overlaid. While in Mavodze the soil unit boundaries to some degree match those of different land cover classes within the well-vegetated undulating landscape, very few land cover classes match soil units in the Massingir fluvial plain, where land cover has changed substantially since the original survey. In both cases the SRTM DEM did not add substantial relief boundary correspondence, casting doubt on whether updated physiographic/soil borders could be used as high accuracy representation against which positional accuracy of legacy delineated units' borders could be assessed. The derived NDVI classes did not match better than the unsupervised land cover classes. In both cases we did not attempt assessing the positional accuracy for reasons just explained, especially in low-relief areas (here, Massingir) where the SRTM DEM does not show enough detail to adjust boundaries based on subtle relief differences. The use of SOTER physiographic units (Dijkshoorn, 2003) would be limited for the same reasons. To obtain consistently high-accuracy representation of physiographic borders, we advise to implement (semi-)automated procedures similar to terrain analysis by Gallant and Wilson (1996), supervised landform classification by Hengl and Rossiter (2003), and automatic segmentation of landforms by MacMillan et al. (2000). However these methods should be adjusted to be sensitive to subtle relief differences. The resulting physiographic units' borders could be used to assess the positional accuracy of legacy maps.

3.2.5. Metadata

We populated most of the ArcCatalog 9.2 metadata template (Content Standard for Digital Geospatial Metadata of the US Federal Geographic Data Committee) using the supplied metadata editor, emphasizing those referring to identification information, spatial reference, entity attribute, and data quality. Table 9 shows some of the metadata, stored internally as an xml document, included in the GIS layer of Massingir renewed legacy soil map. This information should allow other users to access the data and evaluate its usefulness for their intended purposes.

3.3. Inferences about SOC stocks

We evaluated the usefulness of the legacy maps to infer SOC stocks both qualitatively and quantitatively. Although quantitative assessment is clearly preferable, with poor-quality legacy data it may only be possible to reach a qualitative assessment, which can still be useful for identifying priority areas.

3.3.1. Qualitative SOC stocks inference

The legacy map units are based on physiography and make no direct reference to SOC or many other soil properties. However, the physiographic units are described by their pedological composition (dominant and sub-dominant, relative proportions unspecified), based on

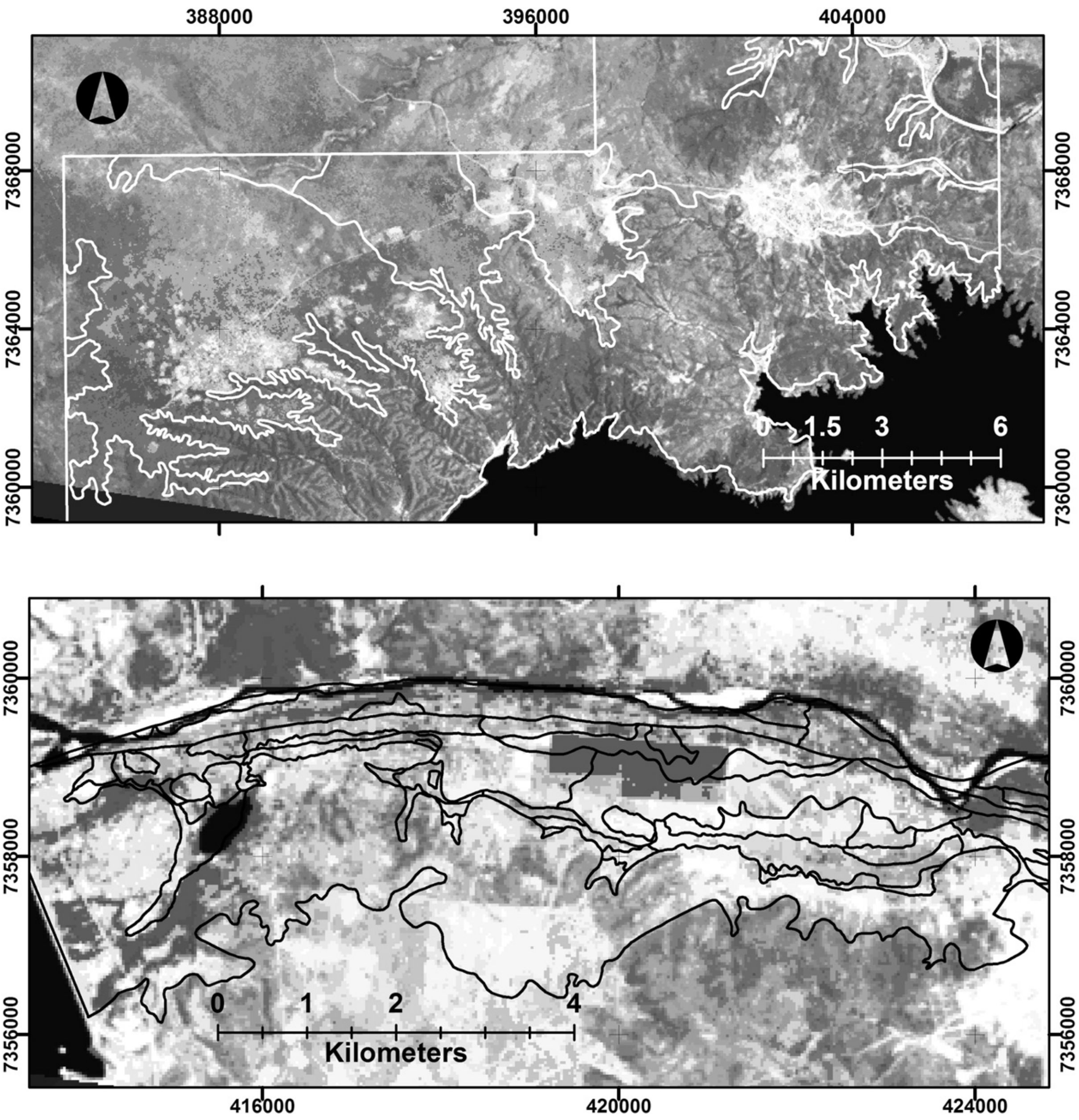

Fig. 5. The Mavodze, Mozambique (top; centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E) and Massingir, Mozambique (bottom; centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E) soil maps overlaid onto Landsat TM classified (unsupervised) image.

Table 9
Some of the metadata information included in the GIS layer of Massingir (Mozambique) soil map (centroid: 23° 53′ 08.34″ S, 32° 12′ 19.31″ E).

Item	Detail	Description
ID information	General description	The soil map was created for use in soil organic carbon (SOC) stocks inference based on the map unit information (attributes), especially the soil classification information. Specific objective was to ….
	Access constraints	This is part of the paper "Rescue and Renewal of Legacy Soil Resource Inventories: a case study of the Limpopo National Park (LNP), Mozambique" (centroid: 23° 18′ 55.57″ S, 31° 55′ 16.24″ E), copyrights by Elsevier B.V.
	Keywords	Legacy soil map Massingir (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E); Gaza, Mozambique; SOC stocks Massingir
Spatial reference	General	GCS_Tete; Tete_UTM_Zone_36S; Clarke 1866
Entity attribute	General	Attributes: ESRI FID (OID); Shape (geometry); Id (Nr); UntFisiog (string); SOC (Nr); A-horizon depth (Nr); F_Area (float); OBS (string)
Data quality	Positional accuracy	100% of digitized polylines within (paper) map line width
	Process steps	This GIS layer was created by (1) scanning at 300 dpi the legacy soil (paper) map of Massingir (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E) by COBA Consultores (1981), then (2) georeferencing using a 11:50,000 topographic map from the national topog. Series, (3) digitizing through the middle of the soil units' borders at a very high magnification, (4) populate the attribute tables with ….

Table 10
Inference of soil organic carbon (SOC) stocks based on physiographic map unit characteristics.

Unit	Physiographic characteristic				Inferences	
	Location	Texture	A-horiz. depth	Water dynamic (water level, flooding frequency)	SOC concentration	SOC stocks
Aa	Flooding area, between natural levees	Coarse	Deep	High water table, frequent flooding	Low	Low
Ab	Natural levee	Medium	Deep	Deep water table, no flooding	Medium	Medium
Ac	Outer gently sloping levee slopes (occurrence of oxbow lakes)	Medium	Deep	Deep water table, no flooding	Low	Low
Ad	Backswamps	FINE	Deep	High level, frequent flooding	Medium	Medium
Ae	Outer edge flood plain, rich in water ways	Medium	Deep	High level, moderate flooding	Medium	Medium
Ba	Colluvium filled lowlands (sediments from Tertiary and Quaternary)	Medium	Deep	Deep, rare flooding	Medium	Medium
Bb	Erosional/undulating upper terraces (Tertiary and Quaternary), under dense mopane	Coarse	Shallow	Deep water table, no flooding	Medium	Low
T	Upper terraces under dense mopane	Coarse	Shallow	Deep water table, very rare flooding	Medium	Low

representative profiles data. This combination allowed us to make qualitative inferences about SOC stocks, based on a combination of physiographic unit's expected characteristics as revealed in the taxonomic name and representative profiles: soil depth, soil texture, water table seasonal depths, and flooding frequency. Other soil properties, such as salinity and pH also affect SOC stocks, but these were not included in map unit information other than the Massingir map so that it could not be included in the inference. Table 10 presents a qualitative assessment of SOC stocks for physiographic units of Chibotane, Chinhangane, and Massingir soil areas. These vary between low and medium, the main limiting condition being coarse soil texture, thin A-horizons, high water tables, and high flooding frequency. The natural levees, the backswamps, the transition to terraces and colluvium-filled lowlands are the sites expected to have higher SOC stocks. These areas are those where condition for vegetation growth is better as a result of substantial rooting depth (levees), low rate of SOC mineralization (backswamps) and available soil moisture (transition from floodplains to terraces and colluvium filled lowlands, with water seeping from higher positions).

3.3.2. Quantitative SOC stocks inference

Section 3.2.2 presented the point data layers and described the data with which attribute tables were populated, amongst others, SOC concentration, A-horizon thickness, and BD. Data from 64 soil profiles were available. Fig. 6 shows the histogram of the SOC concentrations, ranging mainly between 4 and 14 mg kg^{-1} dry soil, well within the range of mean SOC (A-horizon) per survey area (Table 11). This range is also within the range of SOC spatial distribution predicted for the entire LNP (Cambule et al., 2013). These low values are expected, due to the large extents of coarse-textured soils and the hot semi-arid climate.

Table 11 also shows the computed SOC stocks for each the four survey areas, following the measure and multiply approach (Thompson and Kolka, 2005). The range of area-normalized mean SOC stocks is 2–4 kg m^{-2}. These quantitative values are comparable to the “low to medium” qualitative assessment, in the context of the regional climate. These values are a little higher than the 1.6 kg m^{-2} found by Cambule et al. (2014) and, given the fact that SOC concentration is comparable to those across the LNP, the higher mean SOC stocks may be explained by the thicker A-horizons typical of floodplain soils. Cambule et al. (2014) reported mean A-horizon thickness of 10–15 cm for the whole LNP, while Table 11 of the present manuscript shows much thicker A-horizon thicknesses of 24–31 cm in the floodplain region. The computed stocks are also higher than those found by Vågen et al. (2005) for southern African soils, but far lower than those found by Ryan et al. (2011) and also at the lower end of those obtained by Williams et al. (2008) in Eastern Miombo woodlands in Mozambique. These higher stocks are likely due to the high litter leaf from the leguminous trees (*Brachystegia spiciformis*) which is typical of Miombo woodlands.

The total SOC stock of the four survey areas is 596.2 Gg, which represents about 4.0% of the total LNP stocks (16 744 Gg) estimated by Cambule et al. (2014), which is proportionally about the same since the

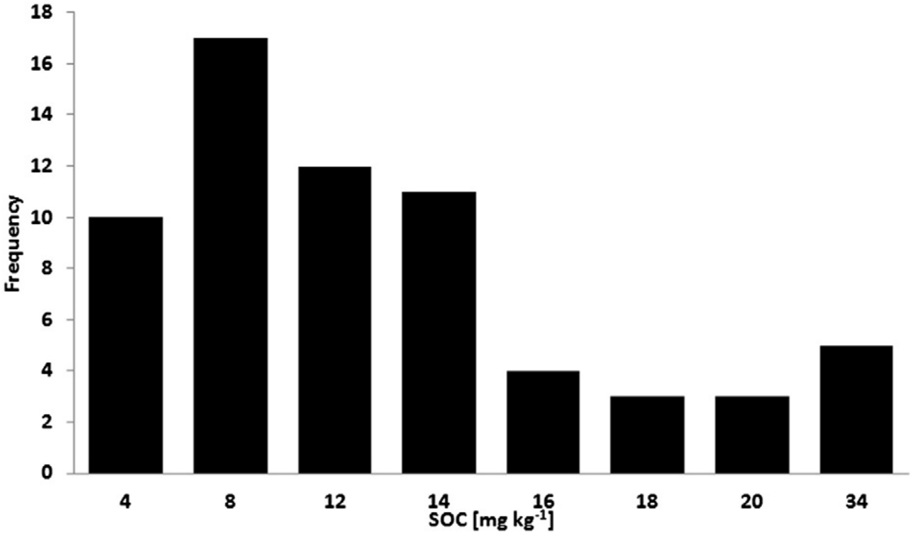

Fig. 6. Summary of retrieved Soil Organic Carbon (SOC) concentration data from legacy soil profiles data of Mavodze, Mozambique (centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E), Massingir, Mozambique (centroid: 23° 53′ 08.34″ S 32° 12′ 19.31″ E), Chibotane, Mozambique (centroid: 23° 57′ 44.68″ S, 31° 15′ 18.59″ E), and Chinhangane, Mozambique (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E).

Table 11
Inference of soil organic carbon (SOC) stocks in A-horizon of legacy soil survey areas of Mavodze, Chibotane, Massingir and Chinhangane (all within or around LNP, Mozambique) based on point data (available bulk density, BD).

Variable	Units	Legacy soil survey area			
		Mavodze (centroid: 23° 49′ 20.28″S, 31° 58′ 25.88″ E)	Massingir (centroid: 23° 53′ 08.34″ S, 32° 12′ 19.31″ E)	Chibotane (centroid: 23° 57′ 44.68″ S, 31° 15′ 18.59″ E)	Chinhangane (centroid: 23° 54′ 55.24″ S, 32° 16′ 23.92″ E)
BD mean	ton kg^{-1}	1.4	1.4	1.4	1.4
A_depth mean	m	0.24	0.25	0.31	0.28
SOC mean	mg kg^{-1}	4.6	11.2	10.1	7.5
Total area	km^2	215.8	28.3	20.4	11.9
SOC Stocks	ton	413,260.2	70,432.9	78,796.3	33,764.4
SOC stocks mean	Kg m^{-2}	1.9	2.5	3.9	2.8

study area is 276.4 km^2, about 3% of LNP area size. The small area-size and especially the unusual (considering the whole park) physiography covered by legacy data, limits its usefulness for extrapolation to the whole LNP; however the exercise is useful for the potentially agricultural areas originally selected for detailed surveys.

The SOC stocks reflect the soil conditions in the early 1980s and might not accurately represent the current situation due to changes in factors controlling the dynamic of SOC stocks, e.g. land use. Vågen et al. (2005) reported a decrease rate in SOC stocks of about 0.82 ton C ha^{-1} $year^{-1}$ in southern Africa after conversion from savanna to agriculture and of about 14.26 ton C ha^{-1} $year^{-1}$ from woodlands to agriculture in the east Sudanian savanna; they did not report initial or final values, so annual proportion rates cannot be computed. In Masvingo (Zimbabwe), Zingore et al. (2005) modeled long-term change rate of about −0.26 ton C ha^{-1} $year^{-1}$ (about 1.1% annually) for woodlands clearance for smallholder subsistence farming; this may also represent similar conversion in our study area given the proximity and similarity in vegetation types between the two areas.

4. Summary and conclusion

We summarize our study with respect to our objectives.

(1) Our data mining or archaeology exercise was successful, and revealed many more relevant surveys than we had suspected when beginning the project. Similar diligent detective work in any area of the world is likely to uncover a large number of useful surveys.
(2) We were able to evaluate the maps, and discovered that the effective map scales as measured by the IMR and observation density were much smaller than publication scales.
(3) We were able to renew selected surveys to some extent. We were not able to solve problems of missing information, nor rectify map unit boundaries not defined according to physiography or land cover which could be identified on recent imagery and DEM. We were able to identify base maps and their original geometry, and thus georeference the soil maps to moderate accuracy. We documented each step, as well as what we could infer about the source maps, as systematic metadata in the renewed digital products.
(4) We were able to assess map quality for SOC mapping, and make semi-quantitative estimates of the spatial distribution of SOC stocks at the mapping dates.
(5) The Cornell adequacy criteria proved to be a useful framework for determining the fitness for use of legacy surveys. However, these criteria only deal with map units and their interpretations, and do not provide a method for evaluating the fitness of primary point data. The advances in GIS since the original publication (1982) allowed us to compute geodetic and scale adequacy exactly more efficiently than in the original proposal.

We conclude that our strategies for the renewal of legacy soil data can be applied in general, whether to get the maximum value out of legacy surveys or to identify spatial and thematic knowledge gaps to guide (partial) resurveys.

Acknowledgments

Most of the legacy reports are archived in various libraries within the National Agrarian Research Institute (IIAM) in Maputo, where we were granted permission to exercise the data mining or archeology. This allowed us the write-up of the present paper and therefore we are very grateful to all staff members who aided us in locating the various reports and maps.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version, at http://dx.doi.org/10.1016/j.catena.2014.10.019. These data include Google maps of the most important areas described in this article.

References

Ahmed, F.B., Dent, D.L., 1997. Resurrection of soil surveys: a case study of the acid sulphate soils of The Gambia. II. Added value from spatial statistics. Soil Use Manag. 13, 57–59.
Batjes, N.H., 2004. SOTER-based soil parameter estimates for Southern Africa (ver. 1.0). ISRIC Report 2004/04, Wageningen.
Batjes, N.H., et al., 2013. Development of global soil information facilities. Data Sci. J. 12, 70–74.
Baxter, S.J., Crawford, D.M., 2008. Incorporating legacy soil pH databases into digital soil maps. In: Hartemink, A., McBratney, A., Mendonça-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer, pp. 311–318.
Bui, E.N., Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. Geoderma 103, 79–94.
Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2012. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. Geoderma 183–184, 41–48.
Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., 2013. A methodology for digital soil mapping in poorly accessible areas. Geoderma 192, 341–351.
Cambule, A.H., Rossiter, D.G., Stoorvogel, J.J., Smaling, E.M.A., 2014. Soil organic carbon stocks in the Limpopo national park, Mozambique: amount, spatial distribution and uncertainty. Geoderma 213, 46–56.
Casimiro, J.F., Veloso, A.P., 1969. Levantamento de solos da margem esquerda do Rio dos Elefantes e sua aptidão para o regadio (Zona a montante da confluência com o Rio Chinguidzi). Grupo de trabalho do Limpopo.
Casimiro, J.F., Veloso, A.P., 1972. Levantamento de solos da margem direita do Rio dos Elefantes e sua aptidão para o regadio (Marrenguele - Banga). Grupo de trabalho do Limpopo, Lourenco Marques.
Cenacarta, 1997. Carta de Uso e Cobertura da Terra. Ministério da Agricultura, Maputo.
COBA Consultores, 1981. Aproveitamento hidroagrícola de Massingir-Chinhangane. Carta dos Solos, Carta de Aptidão para Regadio. SERLI, Maputo.
COBA Consultores, 1982. Aproveitamento hidroagrícola de Massingir-Chinhangane. SERLI, Maputo.
COBA Consultores, 1983a. Aproveitamento hidroagrícola de Chibotane-Machaul-Madingane. SERLI, Maputo.
COBA Consultores, 1983b. Desenvolvimento das Aldeias Comunais de Cubo, Paulo Samuel Kankomba, Massingir-velho e Mavodze. Carta dos Solos e do Potencial Agrícola, Carta das Pastagens e do Potencial Florestal. SERLI, Maputo.
Dent, D.L., Ahmed, F.B., 1995. Resurrection of soil surveys: a case study of the acid sulphate soils of The Gambia. I. Data validation, taxonomic and mapping units. Soil Use Manag. 11, 69–76.
Dijkshoorn, J.A., 2003. SOTER Database for Southern Africa. ISRIC — International Soil Reference and Information Centre, Wageningen.
FAO, 1974. FAO/Unesco Soil Map of the World 1:5 000 000, Unesco, Paris.

FAO, 1988. FAO/Unesco soil map of the world — revised legend with corrections and updates, FAO, Rome. Reprinted with Updates as Technical Paper 20. ISRIC, Wageningen (1997).

FAO, 2001. Soil Carbon Sequestration for Improved Land Management. FAO, Rome 92-5-104690-5 (ISSN 0532-0488).

Forbes, T.R., Rossiter, D.G., Van Wambeke, A., 1982. Guidelines for Evaluating the Adequacy of Soil Resource Inventories. Cornell University Department of Agronomy, New York.

Gallant, J.C., Wilson, J.P., 1996. TAPES-G: a grid-based terrain analysis program for the environmental sciences. Comput. Geosci. 22, 713–722.

Giller, K.E., et al., 2008. Competing claims on natural resources: what role for science? Ecol. Soc. 13, 18.

Goodchild, M.F., Hunter, G.J., 1997. A simple positional accuracy measure for linear features. Int. J. Geogr. Inf. Sci. 11, 299–306.

Gorkin, A.F., Schmidt, O.J., Motylev, V.E., Nikitin, M.V., Shaposhnikov, B.M., 1937. Great Soviet World Atlas (in Russian). Scientific Editorial Institute, Moscow, USSR.

Gouveia, D.H.G., Azevedo, Á.L., 1954. The provisional soil map of Moçambique. Abstract of the Proceedings of the 2nd Inter-African Soils Conference, Leopoldville, pp. 1459–1468.

Gouveia, D.H.G., Azevedo, Á.L., 1955a. Características e distribuição dos solos de Moçambique: I - Carta provisória dos solos do sul de Save, II - Esboço pedológico da colónia de Moçambique. Centro de investigação científica algodoeira.

Gouveia, D.H.G., Azevedo, Á.L., 1955b. Os Solos. In: Junta de Exportaçao do Algodao (Ed.), Esboço de reconhecimento ecológico-agrícola de Moçambique. Centro de Investigação Científica Algodoeira, Lourenço Marques, pp. 3–56.

Gouveia, D.H.G., Marques, A.S.e.M., 1973. Carta de solos de Moçambique (Esc. 1: 4 000 000). Agronomia Moçambicana, 7, 1–68.

Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L., 2008. Digital Soil Mapping with Limited Data. Springer.

Hendriks, P.H.J., Dessers, E., Hootegem, G.V., 2012. Reconsidering the definition of a spatial data infrastructure. Int. J. Geogr. Inf. Sci. 26, 1479–1494.

Hengl, T., Rossiter, D.G., 2003. Supervised landform classification to enhance and replace photo-interpretation in semi-detailed soil survey. Soil Sci. Soc. Am. J. 67, 1810–1822.

Hughes, M.L., McDowell, P.F., Marcus, W.A., 2006. Accuracy assessment of georectified aerial photographs: implications for measuring lateral channel movement in a GIS. Geomorphology 74, 1–16.

Iliffe, J., Lott, R., 2008. Datums and Map Projections for Remote Sensing, GIS, and Surveying. Whittles Pub.; CRC Press, Scotland, UK; Boca Raton, FL.

INIA, 1995. Legenda da Carta Nacional de Solos - escala 1:1.000.000. Instituto Nacional de Investigação Agronómica Maputo.

ISRIC, 2013. Africa Soil Property Maps at 1 KM Resolution. ISRIC — World Soil Information, Wageningen.

IUSS Working Group WRB, 2006. World Reference Base for Soil Resources 2006. FAO, Rome.

Kiiveri, H.T., 1997. Assessing, representing and transmitting positional uncertainty in maps. Int. J. Geogr. Inf. Sci. 11, 33–52.

Lagacherie, P., McBratney, A.B., Voltz, M., 2007. Digital Soil Mapping: An Introductory Perspective. Elsevier, Amsterdam etc.

MacMillan, R.A., Pettapiece, W.W., Nolan, S.C., Goddard, T.W., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. Fuzzy Sets Syst. 113, 81–109.

Mayr, T., et al., 2010. Two methods for using legacy data in digital soil mapping. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Springer Science + Business Media B.V., pp. 191–202.

McBratney, A.B., Mendonca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.

Ministerio do Turismo, 2003. Limpopo National Park: Management and Development Plan. Ministério do turismo, Maputo.

Roeper, C., 1984. Inventário de estudos de solos efectuados na República Popular de Moçambique. Instituto Nacional de Investigação Agronómica, Maputo.

Rossiter, D.G., 2008. Digital soil mapping as a component of data renewal for areas with sparse soil data infrastructures. In: Hartemink, A., McBratney, A., Mendonça-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer, pp. 69–80.

Rural Consult Lda, 2008. Estudo pedológico e de avaliação de capacidade de carga da região entre aldeias de Chinhangane e Banga, vale do Rio dos Elefantes. Relatório final, Ministério do Turismo, Sub-programa de reassentamento do Parque Nacional do Limpopo, Maputo.

Ryan, C.M., Williams, M., Grace, J., 2011. Above- and belowground carbon stocks in a Miombo woodland landscape of Mozambique. Biotropica 43, 423–432.

Schokalsky, Z.J., 1944. A new soil map of Africa (in Russian, English abstract). Pochvovedenie/Pedology 9, 419–425.

Selvaradjou, S.-K., Montanarella, L., Spaargaren, O., Dent, D., 2005. European Digital Archive of Soil Maps (EuDASM) — Soil Maps of Africa. Office for Official Publications of the European Communities, Luxembourg.

Shantz, H.L., Marbut, C.F., 1923. Vegetation and Soils of Africa. American Geographical Society, New York.

Stalmans, M., Gertenbach, W.P.D., Carvalho-Serfontein, F., 2004. Plant communities and landscapes of the Parque nacional do Limpopo, Mocambique. Koedoe 47, 61–81.

Thompson, J.A., Kolka, R.K., 2005. Soil carbon storage estimation in a forested watershed using quantitative soil-landscape modeling. Soil Sci. Soc. Am. J. 69, 1086–1093.

Vågen, T.-G., Lal, R., Singh, B.R., 2005. Soil carbon sequestration in sub-Saharan Africa: a review. Land Degrad. Dev. 16, 53–71.

Williams, M., et al., 2008. Carbon sequestration and biodiversity of re-growing miombo woodlands in Mozambique. For. Ecol. Manag. 254, 145–155.

Zingore, S., Manyame, C., Nyamugafata, P., Giller, K.E., 2005. Long-term changes in organic matter of woodland soils cleared for arable cropping in Zimbabwe. Eur. J. Soil Sci. 57, 727–736.