Remote Sensing of Environment 212 (2018) 161–175

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images

José Alexandre Melo Demattê[a,*], Caio Troula Fongaro[a], Rodnei Rizzo[b], José Lucas Safanelli[a]

[a] Department of Soil Science, Luiz de Queiroz College of Agriculture, University of São Paulo, Ave. Pádua Dias, 11, Postal code 09, Piracicaba, São Paulo, Code 13416-900, Brazil

[b] Center of Nuclear Energy in Agriculture, University of São Paulo, Centenário Avenue, 303, Piracicaba, São Paulo State, Code 13416-000, Brazil

ARTICLE INFO

Keywords:
Bare soil
Satellite image
Landsat
Data mining
Soil monitoring
Topsoil
Soil security
VIS-NIR-SWIR spectroscopy
Digital mapping
Precision agriculture
Environment

ABSTRACT

Soil mapping has been identified as key to environmental issues. The determination of soil attributes to achieve the best decision making on land use planning is crucial. The use of remote sensing (satellite images) can improve understanding of the surface, since it collects a spectral reflectance fingerprint related to soil properties. However, methodologies still gather spatially fragmented information on bare soil in a single image; thus, there is still room to improve information as a continuous surface. This work has the purpose of developing a procedure using multi-temporal satellite image information, aiming to construct a single synthetic image which would represent soils. The work was carried out in the state of São Paulo, Brazil, on a site covering 14,614 km^2. The procedure, designated as Geospatial Soil Sensing System (GEOS3), is based on the following steps: a) creation of a database with Landsat 5 legacy data.; b) filtering of the database to provide images only from the dry season in the region; c) insertion of a set of rules into the system to filter other objects besides soils; d) Each bare soil occurrence for each location along the time-series was used to calculate a Temporal Synthetic Spectral Reflectance (TESS) of the soil surface; e) aggregation of all TESS composes the Synthetic Soil Image (SYSI); f) quantitative and qualitative validation of the SYSI through the correlation between laboratory and TESS, soil line assessment and the principal component analysis (PCA). GEOS3 was able to provide the best representative reflectance of soils for each band during the historical period. Thus, TESS is not the 'true' but a synthetic spectral reflectance. The canonical correlation between laboratory and satellite data reached 0.93. A value of up to 0.88 in the Pearson's correlation between laboratory and TESS was also achieved. In a single scene, only 0.5% of area was available as isolated bare soil for spatial analysis. However, SYSI reached 68%. Considering only the sugarcane agricultural areas, a value of 92% was achieved. Our study indicates that a multi-temporal data mining procedure can retrieve soil surface representation. The key to the results was calculating the median spectral reflectance from the bare soil pixels along the period of the time series. GEOS3 products can aid soil evaluation by assisting in digital soil mapping, soil security, precision agriculture, soil attribute quantification, soil conservation, environment monitoring and soil sample allocation, among others.

1. Introduction

Soil provides several ecosystem services, playing a major role in food production, climate regulation and water and element cycles. Despite its environmental services, soil is one of the most impacted natural resources on the planet (Lal, 2004). Moreover, the world is undergoing climatic changes caused by anthropogenic activities that, in many cases, are linked to soil degradation (Kalnay and Cai, 2003). With world population growth being projected to reach 9.6–12.3 billion people by 2100 (Gerland et al., 2014), there is pressure to expand agricultural areas and consequently increase food production. However, sustainable use of agricultural lands requires knowledge of soil spatial variation to assist in land management. This brings us to the recent concept of soil security, which involves its mapping and understanding to provide quality food, fresh water as to maintain climate and environmental quality (McBratney et al., 2014). As such, the extensive unmapped areas raise concerns, requiring the development and application of new tools for soil mapping (Nolasco de Carvalho et al., 2015). The use of digital approaches for soil mapping (McBratney et al., 2003) enhances the applicability of soil assessments (Hengl et al., 2015). Therefore, remote sensing (RS) has gained great relevance in spatial modeling and digital soil mapping, due to the strong background in

* Corresponding author.
E-mail addresses: jamdemat@usp.br (J.A.M. Demattê), caio.fongaro@usp.br (C.T. Fongaro), jose.lucas.safanelli@usp.br (J.L. Safanelli).

https://doi.org/10.1016/j.rse.2018.04.047
Received 22 August 2017; Received in revised form 24 April 2018; Accepted 26 April 2018
Available online 08 May 2018

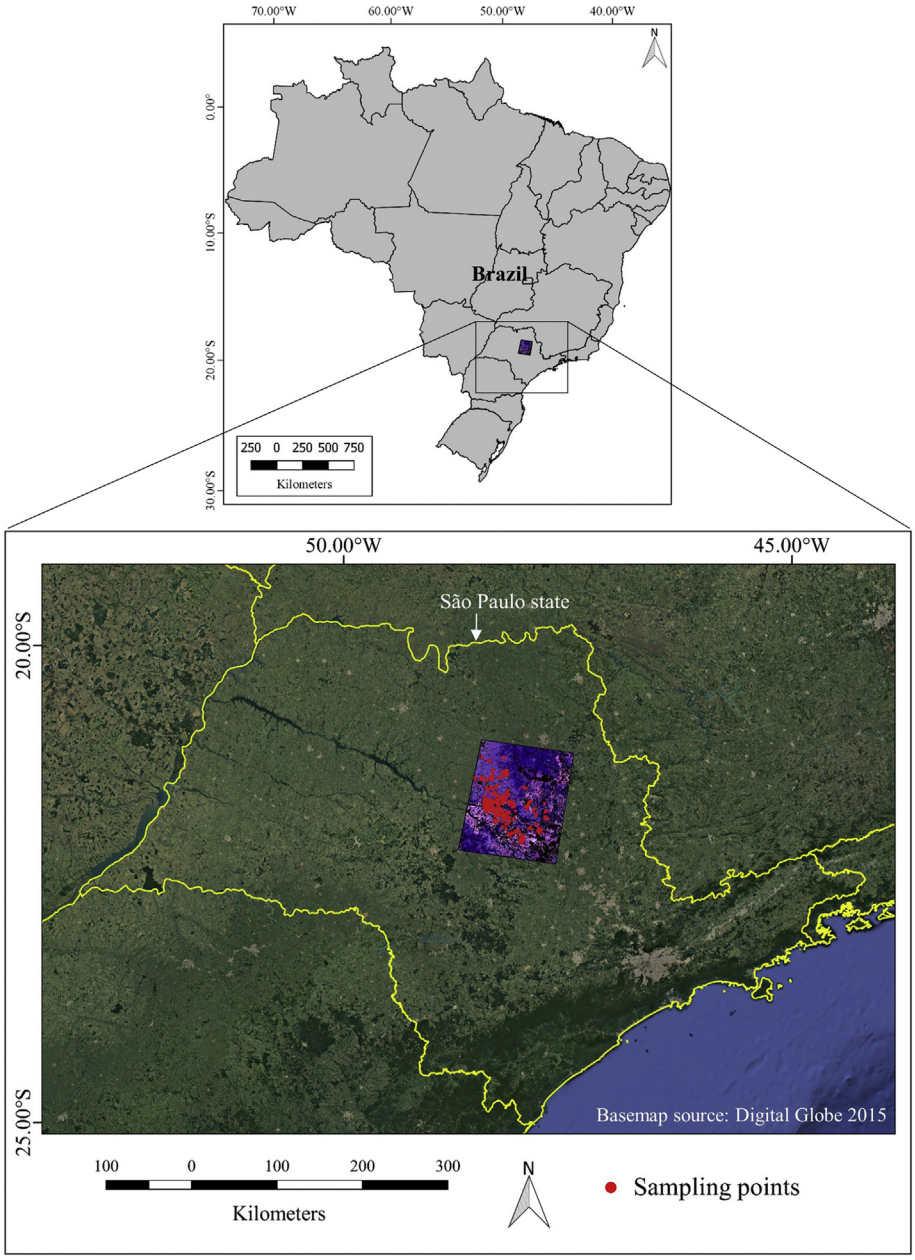

Fig. 1. Study site in the state of São Paulo, southeastern Brazil.

interactions between soil and energy (Stenberg et al., 2010). In addition, studies that correlate satellites with ground sensors are important to the application of remote sensing data in spatial variability of soil (Nanni et al., 2012).

One of the benefits of satellite images is to obtain soil information from large areas (Ben-Dor et al., 2009). However, an image obtained from a single period will not have bare soil for the entire area, which hinders evaluation as a continuous surface. In this case, there are two ways to analyze soils in an image; in spots or spatially. Spot information is related to a single pixel, whereas spatial information is related to bodies formed by contiguous pixels with similar characteristics. If a pedologist or another user is only interested in a specific bare soil pixel for a spot analysis, a single representative image is enough to provide soil surface reflectance. However, if the user wishes to understand the soil as a continuous surface, for spatial modeling, digital soil mapping or other uses, a different approach is required.

The use of satellite images in digital soil mapping is limited by the high vegetation cover of landscapes (Dobos et al., 2006). This topic raises the question: “How can these influences be reduced to acquire topsoil spectral reflectance?” Papers, such as Demattе et al. (2009) have sought to answer this question by developing methods to detect bare soil in a single image. However, a second important question remains - “How can the soil be expressed as a continuous surface, since it is usually covered by vegetation?” This led to the necessity for a second strategy using time series images, which encouraged different approaches, as in the work of Nanni and Demattê (2006), among others (Diek et al., 2017; Rogge et al., 2018).

Continuous representation enables users to apply bare soil satellite composites to their specific purposes, such as mapping clay content, organic matter, cation exchangeable capacity and salinity, or for assistance in soil management zones and precision agriculture. Thus, Diek et al. (2016) proposed the use of a multi-temporal composite image from the Airborne Prism Experiment (APEX) for soil monitoring. Based on stacked images, the authors doubled the amount of bare soil pixels and improved spatial representation of the soil surface. Müller et al. (2016) presented a method to derive high-resolution topsoil texture

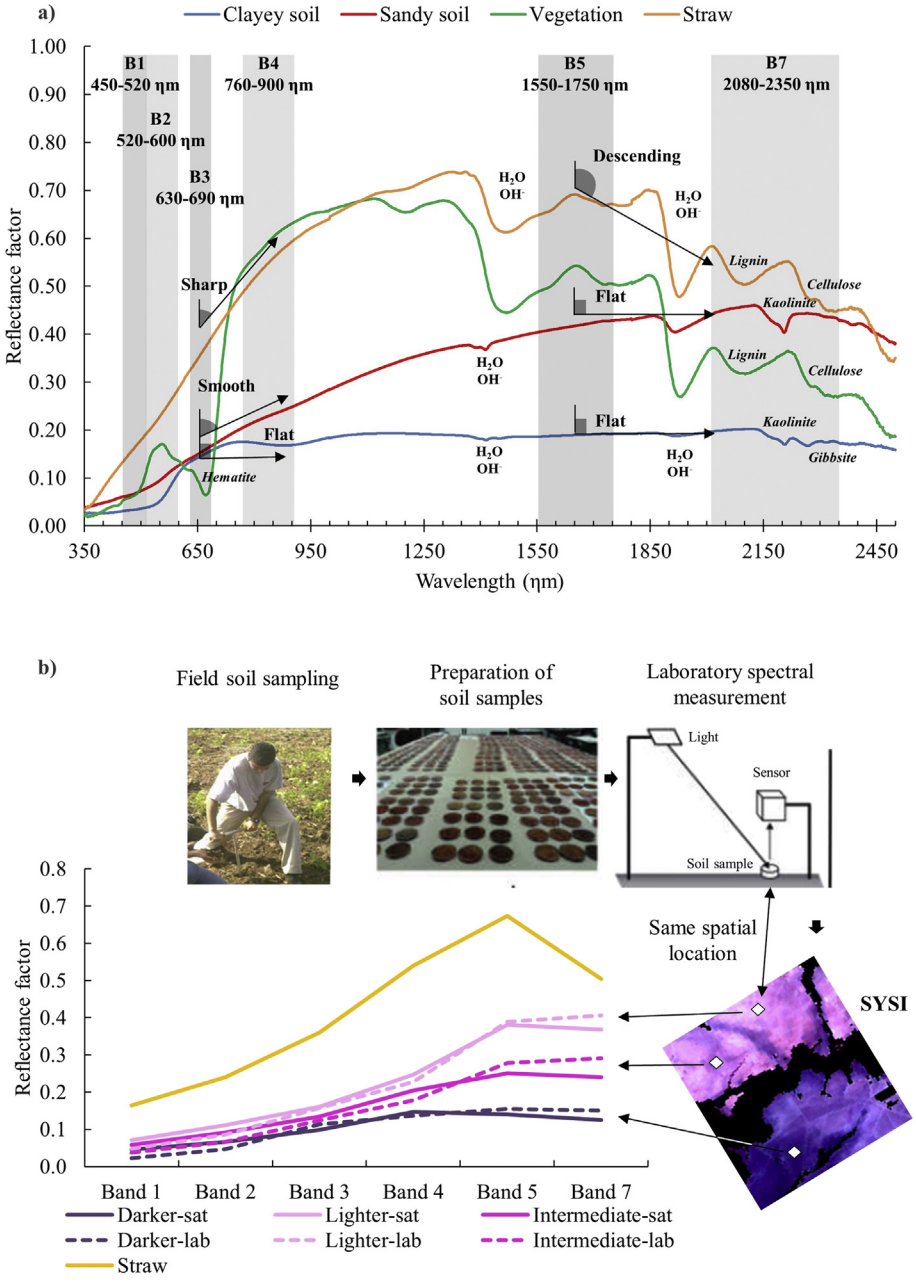

Fig. 2. a) Standard reflectance spectra patterns for straw, vegetation, clayey and sandy soils with respective shading of multispectral Landsat 5 Thematic Mapper. b) Acquisition of soil spectra collection by field soil sampling, sample preparation and laboratory analysis, and from the Synthetic Soil Image (SYSI) for the same georeferenced points. H_2O and OH^- are water and hydroxyl absorption groups. SYSI: synthetic soil image. The sat and lab in figure *b* are, respectively, satellite (SYSI) and laboratory (convolved) acronyms for their soil spectral reflectance.

from 28 images of the Advanced Space-borne Thermal Emission and Reflection Radiometer (ASTER), while Shabou et al. (2015) produced a clay content map over the Kairouan plain (Tunisia) based on a time series of Landsat images.

Furthermore, recent works have proposed novel methodologies to generate bare soil composites based on satellite time series (Diek et al., 2017; Rogge et al., 2018). Alterations in topsoil organic matter were assessed using remote sensing techniques (Pan et al., 2004), which have also been used to assist in land use planning (Marsh, 2010). Although recent studies conducted in temperate climate regions have had the same goal of producing a soil surface composite, there is still room to improve the methods, especially for tropical regions (Rogge et al., 2018). This indicates the necessity to continue improving strategies for bare soil detection, collaborating with the community in several aspects.

Thus, this study aims to develop a bare soil composite method by performing data mining of satellite time series, which could represent soil reflectance on an unfragmented spatially continuous surface. Our hypothesis is that the soil surface has been exposed to satellite measurements at least once and each occurrence could therefore be aggregated into a single representation. The method proposed here is denominated Geospatial Soil Sensing System (GEOS3). This product may be of assistance to users in several areas, such as soil surveys, DSM, quantification of soil attributes (clay content, organic carbon, calcium carbonates, etc.), soil monitoring, precision agriculture and any scientific area that requires topsoil information.

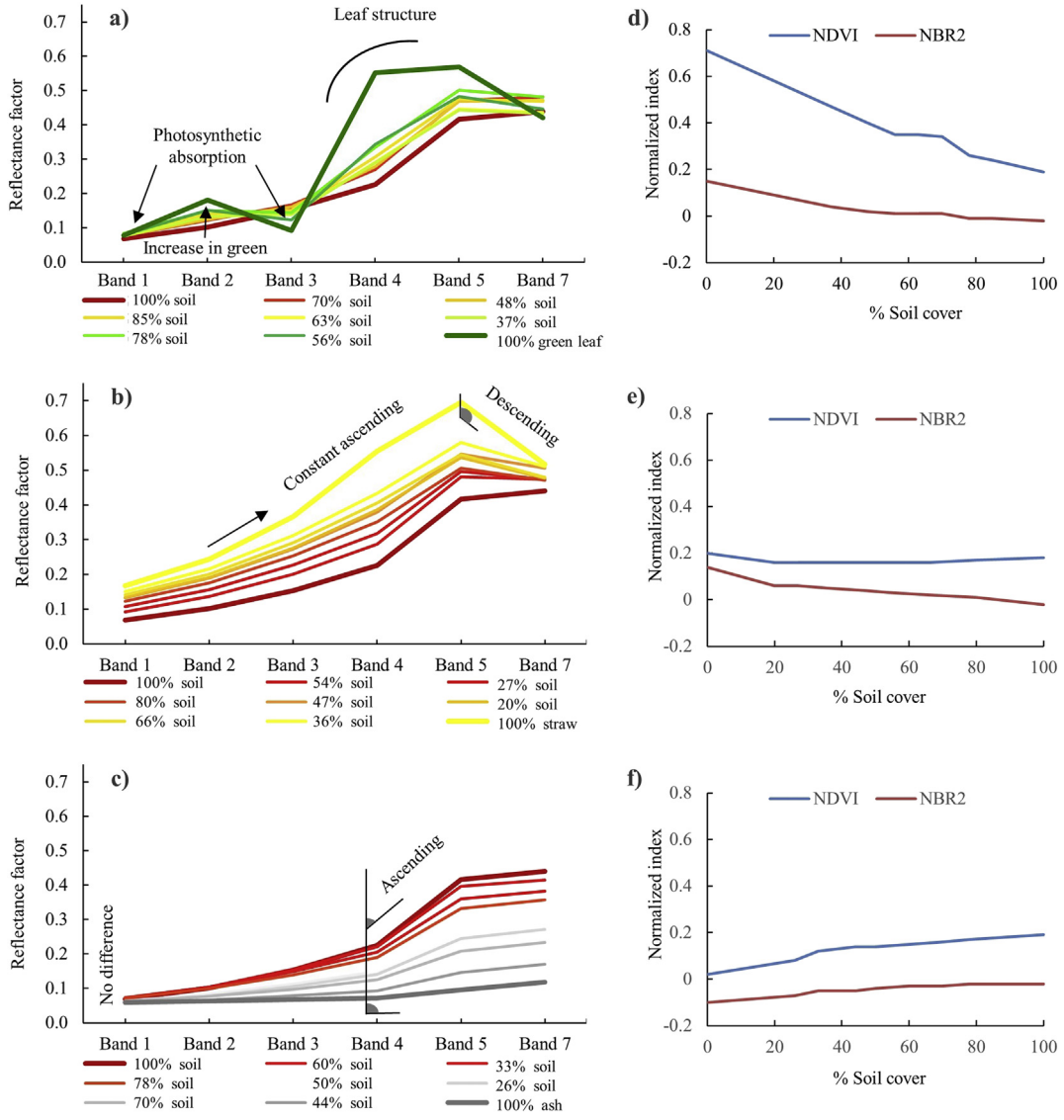

Fig. 3. Spectral reflectance for levels of mixtures of a sandy soil with sugarcane green leaves (a), sugarcane straw (b), and sugarcane straw ash (c). The figures d e and f are the respective Normalized Difference Vegetation Index (NDVI) and Normalized Burn Ratio 2 (NBR2) of the spectral reflectance from figures a, b and c. The spectral reflectance is equivalent to Landsat 5 TM bands.

2. Material and methods

2.1. Study site

The study site covers an area of 14,614 km^2 in the state of São Paulo, southeastern Brazil (Fig. 1). The climate is classified as humid subtropical (Cwa) on the Köppen classification with dry winters and hot summers (Alvares et al., 2013). The dry and wet seasons correspond to April–September and October–March, respectively. The geology is mainly composed of sandstone and basalt (Perrotta et al., 2005). The soil classes (IUSS, 2015) are primarily Ferralsols, Arenosols, Alfisols, Lixisols and Nitisols. Quartz is the main mineral of the sandy soils, while kaolinite and oxide (low-activity clay), with different iron contents, dominate the clayey soils. The relief is diversified, and the main crops are sugarcane, pasture, eucalyptus and citrus.

2.2. Field sampling and topsoil laboratory spectral measurements

A total of 919 points were allocated within the study site, using the *catena* sampling method. This technique is used to detect soil-landscape variations based on relief and geology, increasing the representativeness of soils within the area (Sommer and Schlichting, 1997). Soil samples were collected using an auger at 0–20 cm depth. In the laboratory, the samples were dried (45 °C, 24 h), ground and sieved through 2-mm mesh. Spectra were acquired in the range from 350 to 2500 nm, i.e. visible, near and shortwave infrared (VIS-NIR-SWIR). Samples were conditioned in *Petri* dishes, 8 cm from a sensor and 35 cm from a halogen lamp (50 W). Measurements were performed with a FieldSpec 3 sensor (Analytical Spectral Devices, Boulder, CO, USA), set under laboratory conditions. A spectralon plate was used as white reference. The spectra obtained in the laboratory were convolved to the Landsat Thematic Mapper (TM) spectral bands using a Gaussian function.

2.3. Empirical and theoretical bases of spectral reflectance patterns

The understanding of satellite information is supported by field reference data. The laboratory spectral reflectance patterns help understand the main differences between different landscape objects observed in satellite images (Fig. 2a). This section was added to the current paper as we have previously performed a reflectance spectroscopy experiment to improve understanding of spectral targets, proposing

potential thresholds for masking soil in satellite images. From the results, it could be observed that soil reflectance is highly related to soil particle size distribution (Fig. 2a). Clayey soils have flat spectral behavior in contrast to sandy soils, varying similarly as intensity increases. As the soil granulometry class changes from clayey to sandy, soil reflectance increases its intensity, mainly in band (B) 5 (1550–1750 nm) and B7 (2080–2350 nm). There are significant differences between these two kinds of soil in absorption and reflectance intensities at SWIR range, i.e. between 1400 and 2500 nm (Fig. 2a), and in agreement with Nanni and Demattê (2006), for both laboratory spectra and Landsat 5 TM images.

For a standard vegetation spectrum, energy absorption occurs in B1 (450–520 nm) and B3 (630–690 nm) due to photosynthetic activities (Fig. 2a). Moreover, higher reflection in B4 (760–900 nm) can also be observed due to leaf structure (spongy mesophylls). There is also an absorption peak in B7 (2080–2350 nm) related to cellulose and lignin (Jensen, 2013), which is similar for straw. These features are important to differentiate soils from vegetation, which can be highlighted using spectral indices. The reflectance spectra of straw are rather similar to that observed for senescent vegetation and, in some cases, for sandy soils. An important feature of soils is the absorption peak between 2100 and 2200 nm (Fig. 2a), where the lignin band is present for plant structures (straw and green vegetation) and the kaolinite peak appears for soils (Pizarro et al., 2001).

We have also performed a previous reflectance spectroscopy experiment on the most common targets for agricultural areas. This preliminary experiment was important as the multispectral patterns of sandy soils demonstrated as being similar to straw cover, presenting only relative differences in reflectance (Fig. 3). Increasing levels of finely ground green vegetation, straw, and burned straw (ash) were applied over the soil surface (see Demattê et al., 2016 for details) and spectral measurements (400–2500 nm) were acquired for each treatment using the same laboratory settings explained in the previous subsection, with the reflectance spectra being subsequently convolved to Landsat 5 TM bands. Coverage gradients ranged from bare soil to complete coverage with the residues. The quantification of the relative coverage was performed using supervised classification of images acquired from a digital camera (Fig. 3).

The laboratory experiment shows that sandy soils and straw have similar patterns for the Landsat 5 TM bands (Fig. 3). This behavior can confuse the targets in multi-temporal satellite images, since the latter cannot be discriminated by the lignin and cellulose absorptions at 2100 nm and 2200 nm, as shown in Fig. 2a. On the other hand, the relative differences between the multispectral bands can be used to differentiate vegetative structures from soils. Therefore, the discrimination between straw and sandy soil can be performed through the difference between intensities of B5 (1550–1750 nm) and B7 (2080–2350 nm) (Fig. 3b).

The use of spectral indices, which are univariate values calculated from multiple band reflectance, highlights the differences between spectral targets (Fig. 3d, e, f). Vegetation is quite different from soils regarding the Normalized Difference Vegetation Index (NDVI, Fig. 3d). The soil samples (100% bare soil) have a very low NDVI value (0.19), and the increment of green leaves increases the NDVI. On the other hand, the increment of straw on the soil surface decreases the NDVI values, which makes it difficult to use only the NDVI for spectral differentiation. Thus, the addition of an ancillary spectral index (Normalized Burn Ratio 2, NBR2) enables the differentiation of soils from straw (Fig. 3e).

The differentiation of straw from soil is a difficult task for remote soil sensing in tropical regions. When 100% soil is present in the FOV of the laboratory sensor, NDVI and NBR2 index values are respectively 0.18 and −0.02 (Fig. 3e), but the addition of straw increases the NBR2 to higher positive values. Some areas may also have burned straw or vegetation on surface soil during agriculture or forest management, which deposit ash on the soil surface. The presence of ash decreases soil reflectance, resulting in negative values, or values close to zero, for both the indices (Fig. 3f).

As previously mentioned, spectral indexes highlight differences between landscape objects, which are important for development of the technique, as it allows automatization of data mining in the Landsat 5 TM time series. Discriminating vegetation from soils in satellite images can be achieved by using the 0.25 threshold for NDVI. The laboratory experiment indicates the potential to use the 0.25 threshold for NDVI to discriminate pixels containing at least 80% bare soil and vegetation (Fig. 3d). The NDVI for bare soil was 0.19 and increased gradually with the addition of green leaves, reaching 0.71 for 100% vegetation. This increase is due to the spectral reflectance of vegetation, which has high absorption in the red spectral region and strong reflectance in the NIR spectra. However, dry vegetation (straw) loses this pattern and the NDVI did not show significant changes for its discrimination. As a result, the NBR2 was used to mask straw in the images.

NBR2 showed an increase with the addition of straw, varying gradually from −0.2 (100% soil) to 0.14 (100% straw) (Fig. 3e). The observed NBR2 values were slightly different, as afterwards it was not possible to fix a threshold to use for remotely sensed data. Soil surface reflectance measured by satellite sensor is highly influenced by soil water content, surface roughness and other factors, which can increase the NBR2 values. On the other hand, the results from Fig. 3e give a potential range of NBR2 thresholds. As a result of these observations, different NBR2 thresholds have been tested to support the selection of representative bare soil areas. The NBR2 thresholds also aimed to omit burned areas, the influence of field-soil water content and straw presence in the pixels, and the optimum choice was validated by the correlation statistics between topsoil laboratory spectra and the surface reflectance derived from the synthetic soil image.

2.4. The Geospatial Soil Sensing System (GEOS3)

Based on the indicated findings we went on to develop the technique described as GEOS3, which presented the following steps: a) Acquisition and creation of a time series of satellite images; b) Selection of bare soil areas in each image; c) Calculation of the spectral reflectance for each soil pixel, defined as the Temporal Soil Spectral Reflectance (TESS); d) All TESS represent the Synthetic Soil Image (SYSI); e) Validation and verification of the GEOS3 products.

A database was created using Level 2 data products of Landsat 5 TM legacy data (Masek et al., 2006). The products were acquired from the EROS (Earth Resource Observation and Science Center) Science Processing Architecture (ESPA) ordering interface of the United States Geological Survey (USGS). The Level 2 products consist of atmospherically corrected surface reflectance bands, processed by the Landsat Ecosystem Disturbance Adaptive Processing System - LEDAPS of ESPA (Schmidt et al., 2013). The LEDAPS produces top-of-atmosphere reflectance and applies atmospheric corrections to generate surface reflectance products. The corrections are based on the Second Simulation of a Satellite Signal in the Solar Spectrum - 6S (Schmidt et al., 2013; Vermote et al., 1997). The Landsat 5 TM Level 2 data used in the study corresponded to orbit 220 and path 75. We have selected images with up to 10% maximum cloud cover, which were collected between May and September from 1984 to 2011. A total of 151 images were used to create the time series processed by the GEOS3.

Level 2 products used in this study corresponded to the surface reflectance (SR) TM bands in the VIS-NIR-SWIR region: B1 (450–520 nm), B2 (520–600 nm), B3 (630–690 nm), B4 (760–900 nm), B5 (1550–1750 nm), and B7 (2080–2350 nm); two SR-derived spectral indices that are provided by the ESPA ordering interface: NDVI and NBR2; and Landsat SR quality assessment bands, also provided by the ESPA ordering interface: QA (quality assessment bands, cfmask and cfmask confidence). Additional information on surface reflectance data, spectral indices calculations and quality assessment bands can be found in the online product guide of Landsat 4–7 Surface Reflectance

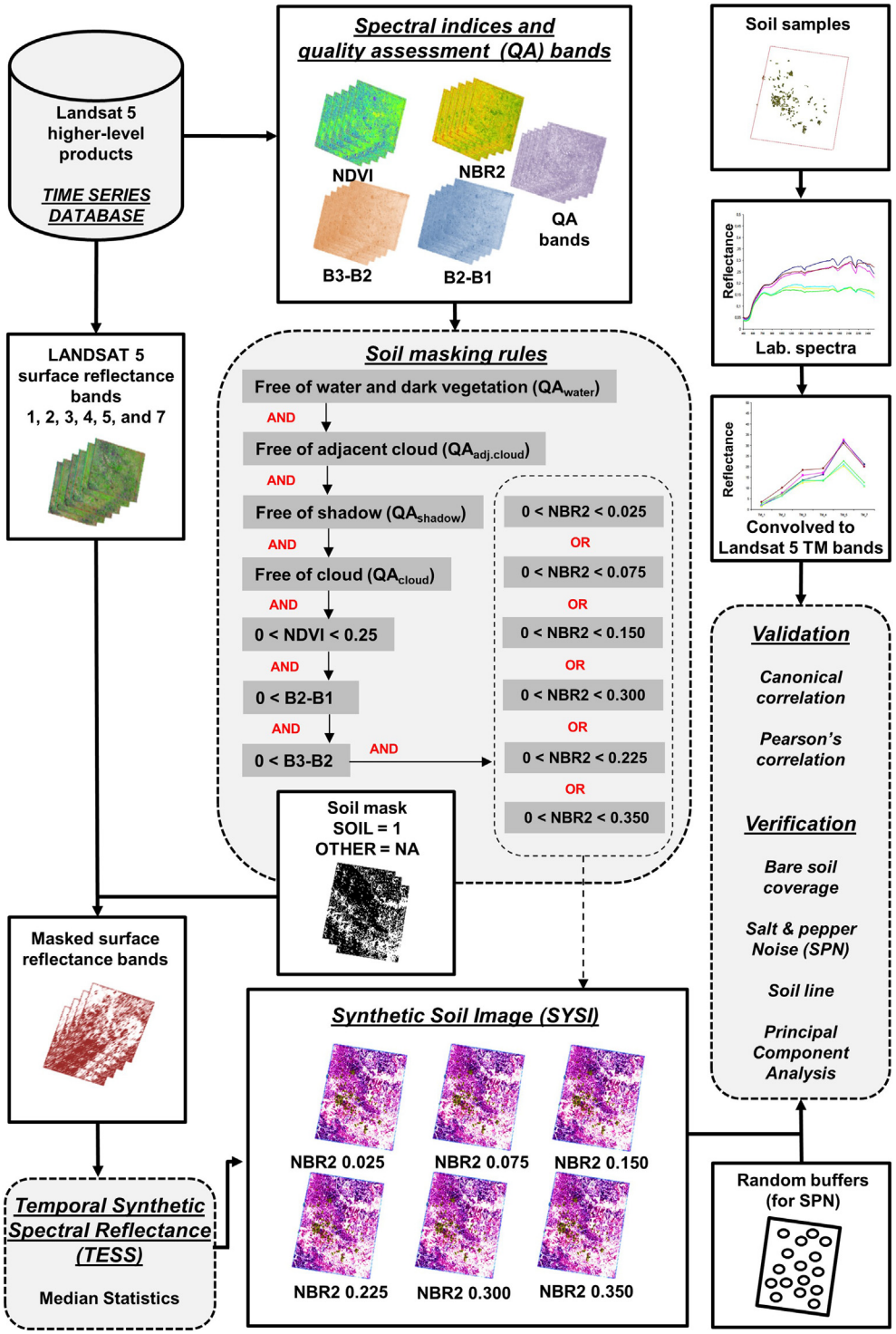

Fig. 4. Flowchart of the Geospatial Soil Sensing System (GEOS3) and its products: Temporal Synthetic Spectral Reflectance (TESS) and Synthetic Soil Image (SYSI). NDVI: Normalized Difference Vegetation Index. NBR2: Normalized Burn Ratio 2. The terms AND and OR are Boolean operators used in the script of GEOS3. The AND operator requires all the rules, narrowing the search for bare soil pixels. The OR operator requires either one or all the other rules, broadening the results to allow the sensitivity test. Not available (NA) assigns a missing value (a value of NA in the R software) for further calculation.

(LEDAPS) product (*https://landsat.usgs.gov/landsat-surface-reflectance-data-products*).

2.5. Selection of potential bare soil pixels and calculation of Temporal Synthetic Spectral Reflectance (TESS)

The bare soil pixels are masked by combining quality assessment bands and spectral indices. The term mask is related to the process of extracting all potential soil pixels, assigning the remaining pixels with a missing value (NA value in the R software), indicating not available for calculation. Pixels containing vegetation, straw, water and other are flagged NA. This process occurs when a pixel value exceeds at least one threshold of the masking products (QA bands and spectral indices), and only the remaining bare soil pixels are kept for further calculations (Fig. 4).

The QA bands were used to mask pixels influenced by clouds, cloud

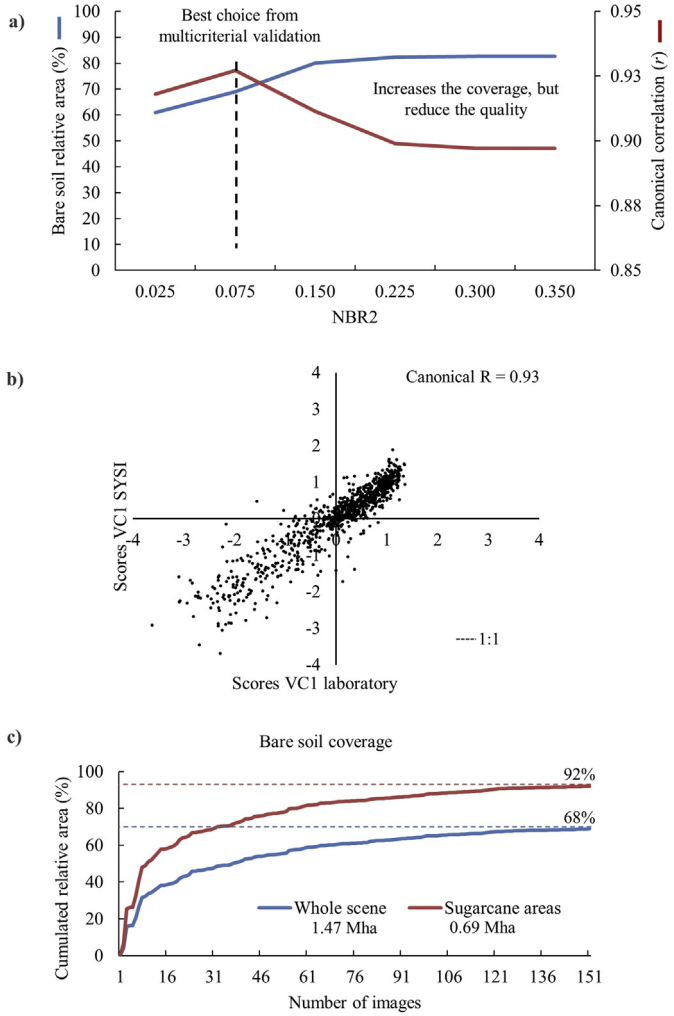

Fig. 5. Multiple-criteria evaluation of Normalized Burn Ratio 2 (NBR2) thresholds: a) bare soil coverage and canonical correlation of bands; b) Scores dispersion of the first and second canonical variables for the 0.075 NBR2 threshold, calculated from the canonical correlation analysis; c) Bare soil coverage for the whole study area and sugarcane crops (Rudorff et al., 2010) as a function of the number of images used for data mining in the Landsat 5 TM time series, using the 0.075 NBR2 threshold.

shadows, dark vegetation or water (Zhu and Woodcock, 2012). The NDVI was used to distinguish vegetation from bare soil areas and the spectral mixture of soil and vegetation (Chagas et al., 2016; Shabou et al., 2015). The NBR2 not only detects straw, but it has also been used as an indicator of soil moisture (Musick and Pelletier, 1988). NBR2 has also been used to map burned areas (Escuin et al., 2008) and to differentiate clay texture of the soil surface (Madeira Netto, 1996; Shabou et al., 2015). The difference between B3 and B2, as well as between B2 and B1 were also calculated to improve soil masking. Some studies indicate that the discrimination of bare soil in tropical regions may be improved with the use of B1, B2 and B3 (Fiorio and Demattê, 2009; Nanni and Demattê, 2006).

A set of combined rules was created considering the products derived from Landsat TM level 2 products, i.e., QA assessment bands, NDVI, NBR2 and the calculated indices B2-B1 and B3-B2 (Fig. 4). Pixels failing the QA tests were flagged as NA. The negative values of B3-B3 and B2-B1 were flagged as NA. We have considered the range between 0 and 0.25 for NDVI, and values outside this range were also flagged as NA. Regarding the use of NBR2, six threshold ranges were tested: 0 to 0.025, 0 to 0.075, 0 to 0.150, 0 to 0.225, 0 to 0.300 and 0 to 0.350, and values outside the previous ranges were also flagged as NA.

After masking the time series, each remaining pixel (potential soil) had its synthetic reflectance calculated through the median statistics. The image keeps the original Landsat 5 TM spectral bands, creating the TESS. Since TESS is related to the pixel spectral fingerprint, the final continuous surface of spatial information is SYSI. We have also generated a raster that retrieved the cumulative sum of the times that each pixel was assigned as bare soil during the time series.

2.6. Validation and verification of the results

Three criteria were considered for evaluation of the best NBR2 threshold, which was integrated with the other masking rules, to produce the final synthetic soil image. The canonical correlation between the SYSI bands and the convolved laboratory spectra, as well as Pearson's correlation analysis of paired bands (SYSI bands with the equivalent laboratory convolved spectra), were used to validate the results. The relative bare soil area obtained for each threshold and the level of salt and pepper noise (SPN) in the SYSI were used to verify the reasonableness of the results.

Considering the entire scene, the relative proportion with bare soil was calculated for the six NBR2 thresholds. The proportion of bare soil in agricultural areas (sugar-cane fields; Rudorff et al., 2010) was also calculated. The other verification step, related to the SPN, was defined by standard deviation statistics of a hundred random buffer polygons placed over the study area. The SPN is caused by sharp and sudden disturbances of reflectance in the SYSI, which prejudices spatial continuity and creates artifacts in the image (Gonzalez, 2009). As the noise increases, a higher dispersion in reflectance is expected. The boxplot of the standard deviation was plotted for each band and for each NBR2 threshold, and their median values were compared using an approximated 95% confidence interval expressed by notches in the boxplots (Krzywinski and Altman, 2014).

The canonical correlation was used to evaluate the similarity between TESS and the laboratory convolved spectra. In canonical correlation, all the bands for each group were reduced to separate canonical variables through multivariate linear transformation (Webster, 1977). The correlation between the scores for the first canonical variable of each group was evaluated by implementation of the candisc procedure in the R package candisc (Friendly and Fox, 2016). The Pearson correlation was also calculated for each band between SYSI and convolved spectra. Furthermore, the principal component analysis (PCA) and soil line plots were also used to verify the results. The PC1 and PC2 were plotted in PC space to check the congruence between laboratory and TESS.

3. Results and discussion

3.1. GEOS3 processing

The multi-criteria evaluation of the NBR2 threshold produced different bare soil coverage over the study site (Fig. 5a). Maximum coverage was acquired by using the 0.150 NBR2 threshold, with around 80% of the total area having the soil surface represented in SYSI (Fig. 5a). On the other hand, as a high amount of soil coverage is acquired when the NBR2 is less restrictive (greater values of NBR2), SYSI quality decreases as a result of confusion with other soil-similar land covers.

Canonical correlations between the laboratory and the spectra generated by different NBR2 thresholds are also presented (Fig. 5a). The use of the 0.075 NBR2 threshold provides the best association between the laboratory and satellite spectra with a canonical correlation of 0.93 (Fig. 5b). The highest threshold had a slight reduction in canonical correlation (to about 0.9), although the bare soil area increased. Despite there being small differences between canonical correlations among the thresholds, we believe that 0.075 NBR2 was the optimal choice, since it reduced the presence of artifacts in the image

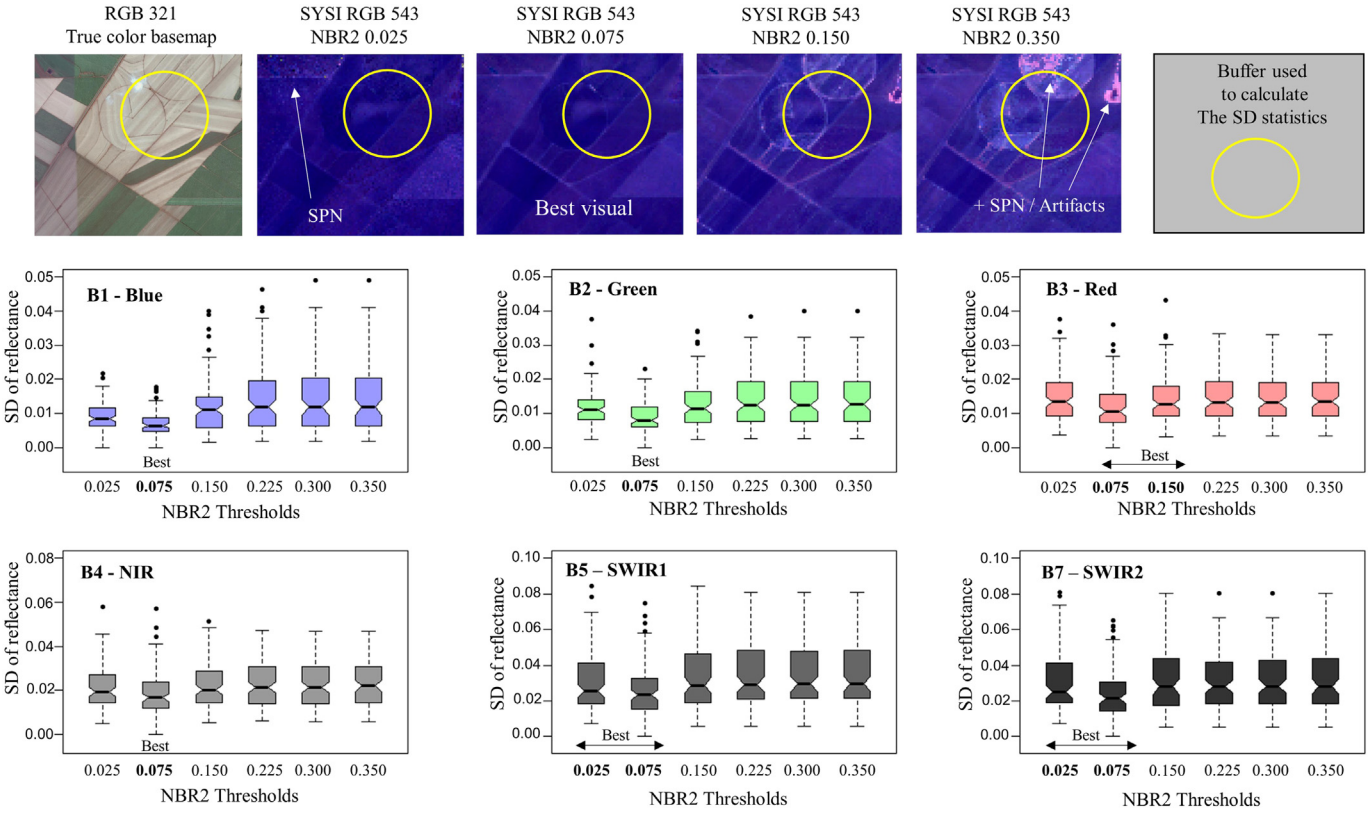

Fig. 6. Salt-and-pepper noise (SPN) and artifacts, which are caused by sudden changes in the spectral reflectance of the Synthetic Soil Image (SYSI). A RGB 321 true color composite from an agricultural landscape is presented at the top, whereas the other equivalent images show the increase of salt-and-pepper noise as the NBR2 threshold gets higher. The images made with 0.225 and 0.300 NBR2 thresholds were not shown. Boxplots of the standard deviation of spectral reflectance, calculated by a hundred buffers (300 m radius), for all Landsat 5 TM VIS, NIR and SWIR bands, as a function of the increase of NBR2 thresholds. The boxplots have their medians with approximate confidence intervals of 95% (notches).

(Fig. 6a). Additionally, good bare soil coverage (68%) was acquired with the 0.075 NBR2 threshold. The increase in canonical correlation and spatial consistency (low artifacts) compensates the decrease of around 10% total coverage in SYSI produced by higher NBR2 thresholds, since poor-quality spectral reflectance impacts its further use for soil modeling.

The SPN level was higher as the NBR2 thresholds increased (Fig. 6). The purple shades of Fig. 6 are related to reflectance variations for the RGB 543 composite. If the masking process does not select appropriate occurrences of bare soil, spatial artifacts can appear in the SYSI, increasing its SPN levels. For instance, the spatial artifacts are represented as magenta colored pixels within a region where the purple color fully represents the soil surface (Fig. 6). The purple shades and magenta are specific to the SYSI RGB 543 composite, and these artifacts are different from natural soil variations. The spatial artifacts are caused by high frequency variations in reflectance.

The SPN verification by deviation statistics confirmed that a restrictive NBR2 improves the spatial consistency of SYSI, as seen in the boxplots of Fig. 6. Standard deviation statistics were evaluated for a hundred samples (buffers with 300 m - equivalent to 10 pixels of radius), as represented by the circles. The increase of SPN in the SYSI resulted in higher standard deviations of surface reflectance. For the visible spectrum (B1, B2 and B3), the SD of the 0.075 NBR2 threshold was the lowest, with the exception of the red band, which was equivalent to 0.150 (Fig. 6). Furthermore, the 0.075 NBR2 had the lowest SD for all the infrared bands (B4, B5 and B7), or was at least equivalent to other similar thresholds. However, a minimum was evident for the 0.075 value. The SPN causes higher values of SD reflectance, which are related to mixes of straw and other covers with the soil, creating artifacts in the SYSI. The higher NBR2 values were not efficient to mask mixed straw and soil surface covers, since the vegetation or straw have higher intensities than soils in the visible and infrared spectrum (Figs. 2 and 3). Increases in the local standard deviation were caused by the increase in artifacts (Fig. 6a), in some cases appearing as salt and pepper noise due to sudden disturbances in spectral reflectance.

The NBR2 is a normalized difference index calculated from Landsat 5 TM bands 5 and 7, which represent the shortwave infrared spectrum. As stated in the Material and Methods section, the use of NBR2 aimed to remove influential pixels representing straw. Although Landsat data has limited spectral information, the combination of NBR2 and NDVI stress the differences between the soils and other common spectral objects, since they use different reflectance bands for their calculation. This combination was important for improving soil discrimination. Indeed, this was recommended in a previous study (Rogge et al., 2018), which suggested the use of short infrared spectral bands and the combination of additional spectral indices to improve bare soil discrimination.

Furthermore, employing a wider date range in the time series enabled the use of much more restrictive masks, which benefited the selection of purer bare soil pixels with lower surface effects. Considering the entire study area, with the use of just one image, GEOS3 achieved only 0.5% bare soil coverage (Fig. 7a). By increasing the time series to 10 and 40 images, 35 and 55% bare soil coverage, were respectively achieved (Fig. 7b and c). The GEOS3 achieved its maximum coverage (68%) for the entire area using 151 images from the database (Figs. 5c and 7d), when considering the 0.25 NDVI and 0.075 NBR2 thresholds. This number represents agricultural and non-agricultural areas; or, areas where, at a certain time, the vegetation had been extracted by anthropic or natural activities. The areas in black had no bare soil occurrences during the time series, and are related to forests, pastures and

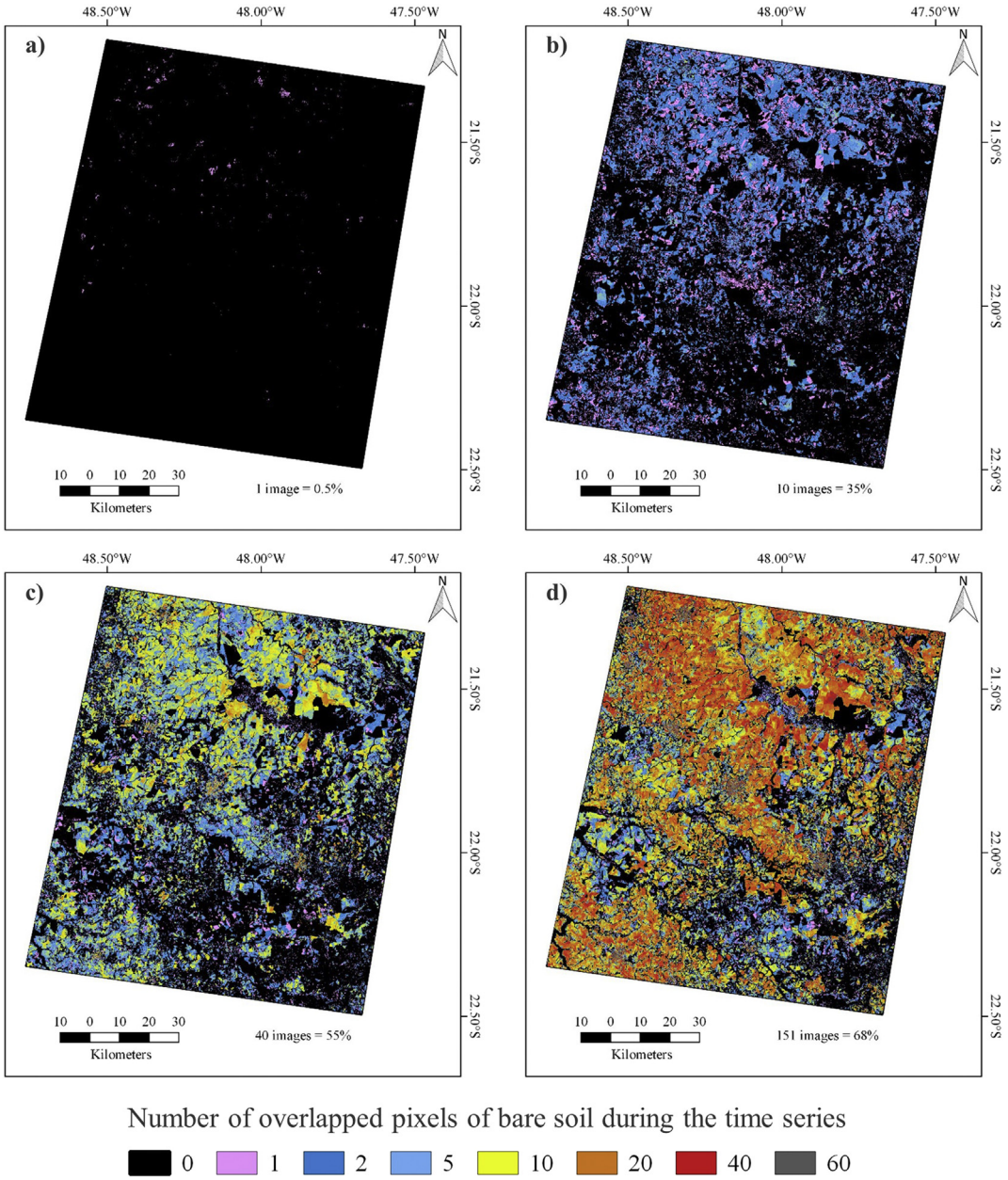

Fig. 7. The increase in bare soil area as a function of the number of images used in the Landsat 5 TM time series. a) Bare soil coverage and number of overlapped pixels using a single image in the time series. b) Bare soil coverage and number of overlapped pixels using 10 images in the time series. c) Bare soil coverage and number of overlapped pixels using 40 images in the time series. d) Bare soil coverage and number of overlapped pixels using the time series (151 images). The color represents the amount of overlapped bare soil pixels identified in the time series. The percentage of bare soil coverage for the whole scene is presented at the bottom of each figure.

other land uses without soil tilling (Fig. 7). When specifically considering the agricultural areas, from the sugarcane mapping (Rudorff et al., 2010), GEOS3 reached up to 92% bare soil coverage (Fig. 5c).

Another similar study has determined the barest soil composite for the Swiss Plateau (Diek et al., 2017), where researchers divided the Landsat time series into five-year intervals, achieving up to 43% bare soil for the total agricultural areas for each period. A similar approach was used in the work developed by Rogge et al. (2018). In both cases the temporal variability of bare soil occurrences in the short time periods was intended. In this study, besides the fact that it is possible to retrieve each bare soil occurrence at each time during the time series, the full date range database provided about 90% bare soil coverage in the agricultural areas, which represents about 68% for the total scene. These results stress the fact that a greater spatial representation of the soil surface can be achieved by increasing the number of images or widening the date range of the time series. Nevertheless, it is not only the number of images that will indicate the success of bare soil detection in time series, but also the correct identification of the pixel as 'soil' that will have a greater impact on the percentage of detected area.

Median filters are widely used to smooth data in image processing (Sun and Neuvo, 1994), therefore, median statistics were chosen to calculate spectral reflectance for a given pixel, since they provide a better averaged representation for highly skewed distributions of bare soil reflectance (Figs. 8a, b). We believe that the greater the number of locations with bare soil occurrences, the better the TESS estimation will be when using the median statistic. Increasing the occurrences by broadening the date range of the time series minimizes the influence of extreme values on the frequency distribution of reflectance.

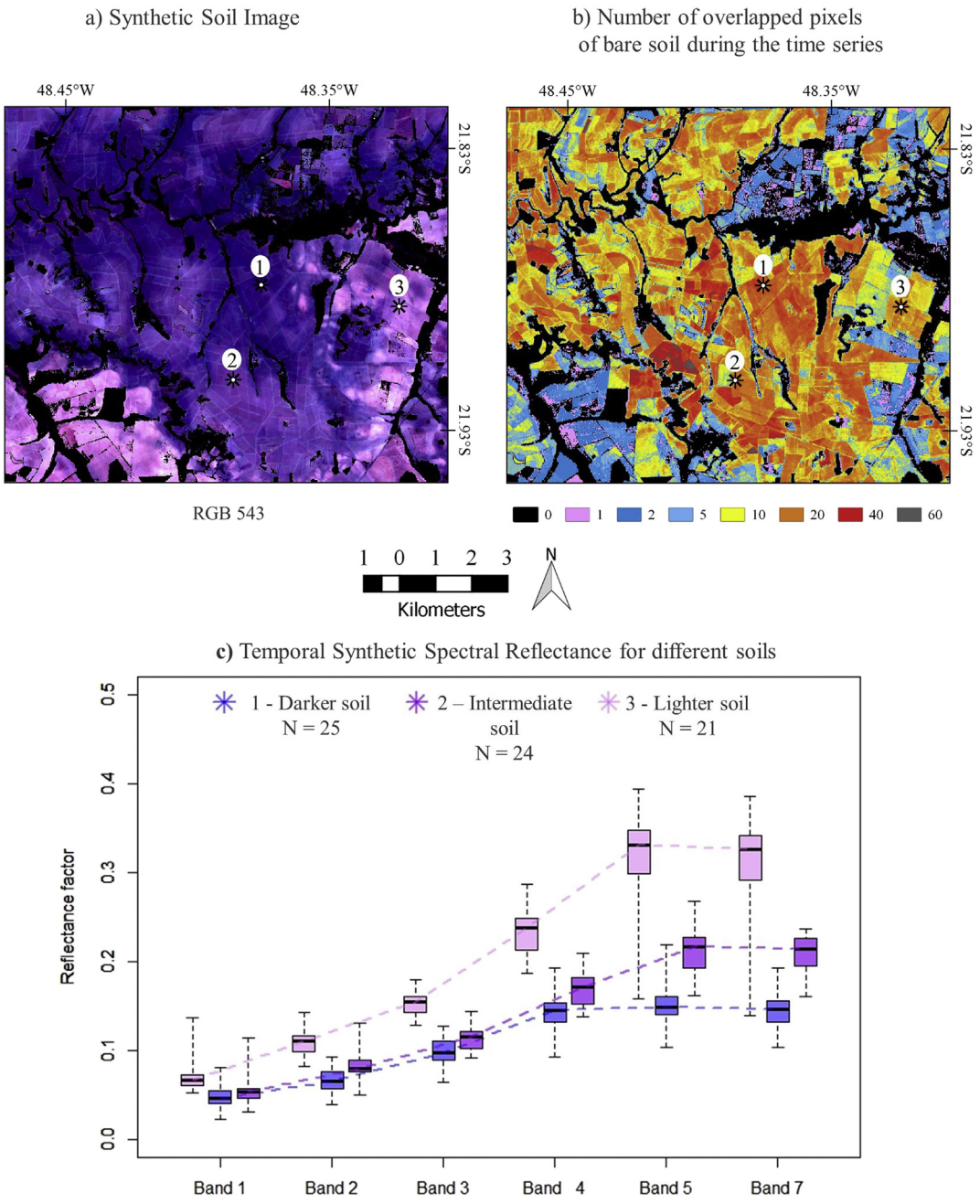

Fig. 8. a) The Synthetic Soil Image (RGB 543 with visual enhancement) created by GEOS3. Each pixel has a Temporal Synthetic Spectral Reflectance (TESS) calculated by the median statistics. b) Depending on the region, there were more or fewer overlapped pixels of bare soil, regardless of the soil type. c) Spectral reflectance dispersion for 1-darker, 2-intermediate and 3-lighter soils used to calculate the TESS.

Additionally, the median statistic was an averaging approach more robust than the previously published methodologies (Diek et al., 2017; Rogge et al., 2018).

The number of pixels used to calculate the median statistics were different depending on the region. For instance, at points 1, 2 and 3 (Fig. 8a), GEOS3 calculated the median as a function of 25, 24 and 21 stacked pixels (Fig. 8b). For these calculations, each location must have been bare soil at least once during the time series. Although the method relies mostly on arable areas, naturally exposed soil surfaces could also be detected. For instance, recently deforested areas and natural biomes that have a high fractional cover of soils (> 90%) in certain periods of the year, as in the case of savannas in tropical regions, can be detected by the masking process. Therefore, those areas are accounted for in the system to create the synthetic soil image.

The synthetic 543 composite contains only the spectral variations of soil (Fig. 8a). The purple shades are due to visual enhancement used for the SYSI and are related to soil differences (Fig. 8c). The false color composite RGB 543 was used because the red and infrared spectral range stresses the differences between tropical soils, which are related to their composition features (e.g. soil minerals, hematite, quartz, ilmenite, and others). To create the visual aspect of SYSI, a linear enhancement (stretched between 2 and 98% of the histogram) was applied in an original Landsat 5 TM containing many spectral patterns, such as water bodies, vegetation, urban patches, and soils. Due to this visual enhancement, soils had lower values in B4 relative to the B5 and B3, resulting in shades of purple. Then, the same enhancement was applied to SYSI. This approach was used only for visualization and to confirm that the SYSI image is a “true” soil composite. The original

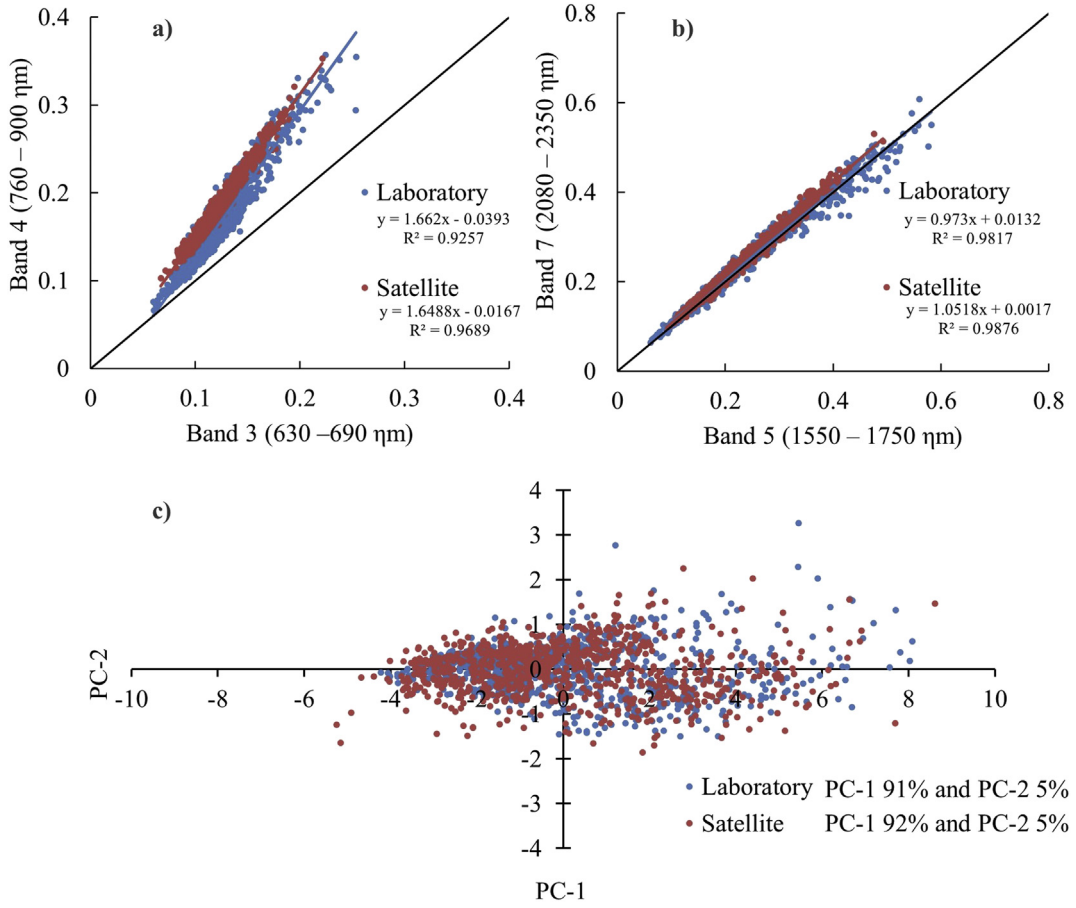

Fig. 9. Comparison between laboratory and Temporal Synthetic Spectral Reflectance (TESS) from the Synthetic Soil Image (SYSI) using: a) soil line (band (B) 3 and B4), b) dispersion between B5 and B7, and c) scores dispersion of the entire visible, near and shortwave infrared spectrum principal components (PC) for each spectral source (laboratory and TESS).

Table 1
Correlation between spectral reflectance from laboratory and Temporal Synthetic Spectra (TESS).

	Laboratory[1]					
TESS	B1	B2	B3	B4	B5	B7
B1	0.71	0.69	0.62	0.68	0.70	0.68
B2	0.75	0.74	0.67	0.75	0.78	0.76
B3	0.69	0.70	0.67	0.76	0.80	0.78
B4	0.72	0.72	0.68	0.78	0.83	0.81
B5	0.73	0.72	0.66	0.79	0.88	0.86
B7	0.72	0.71	0.65	0.79	0.88	0.87

[1] Landsat 5 TM simulated bands; B1: band 1 (450–520 ηm), B2: band 2 (520–600 ηm), B3: band 3 (630–690 ηm), B4: band 4 (760–900 ηm), B5: band 5 (1550–1750 ηm), B7: band 7 (2080–2350 ηm). $p < 0.0001$.

values of reflectance, however, is preserved for further analysis.

3.2. Validation and verification of the GEOS3 products

Different qualitative and quantitative approaches were performed in order to validate the GEOS3 products. The first qualitative evidence was evaluation of soil line information between B3 and B4 (Baret et al., 1993). The soil line was compared for both convolved laboratory and respective TESS from the SYSI (Fig. 9a). The comparison presented little difference between the sources, and the patterns were very close to the results from a soil spectral library of a Brazilian tropical region (Nanni and Demattê, 2006), since the same soil patterns occurred between B4 and B3, and between B5 and B7 (Fig. 9a, b). The information provided the spectral patterns is an important fingerprint and has been related to soil constituents and other environmental factors, as demonstrated by Viscarra Rossel et al. (2016).

In addition, similarities between the laboratory convolved spectra and the respective TESS were also evaluated using principal component analysis. The first 2 PCs of each dataset (laboratory and TESS) were compared according to their location in the principal component space (Fig. 9c). Both datasets present a similar pattern, with a dense cloud of points located on the right-hand side of the y axis, while a second group of points had a widely spread pattern on the left-hand side of the vertical axis. The slight differences between spectra dispersions are mostly due to environmental influences, such as soil moisture and surface cover mixtures. The similar dispersions of the 2 datasets in the PC space, besides the fact that PCs 1 and 2 explain most of the spectral data variability (~96%), shows GEOS3 to be a reliable source of soil information via satellite data mining.

Another important piece of evidence that supports the contention that SYSI represents surface topsoil is the high correlation between satellite and laboratory spectra (Table 1). Pearson's correlation (r) between spectral data for these two levels of acquisition varied between 0.67 (laboratory B3 with SYSI B3) to 0.88 (laboratory B5 with SYSI B5), with a mean of 0.77. The correlation for B5 and B7 had the highest r. Although the strongest correlation was for B7, this band from TESS presented a slightly lower mean reflectance due to the increased influence of water in this region of the spectrum for remotely sensed data, in agreement with Musick and Pelletier (1988) and Khanna et al. (2007). The lowest correlation for visible bands could be attributed to atmospheric scattering effects, which are more present in the visible spectrum region (Jensen, 2013). Furthermore, the statistical correlation

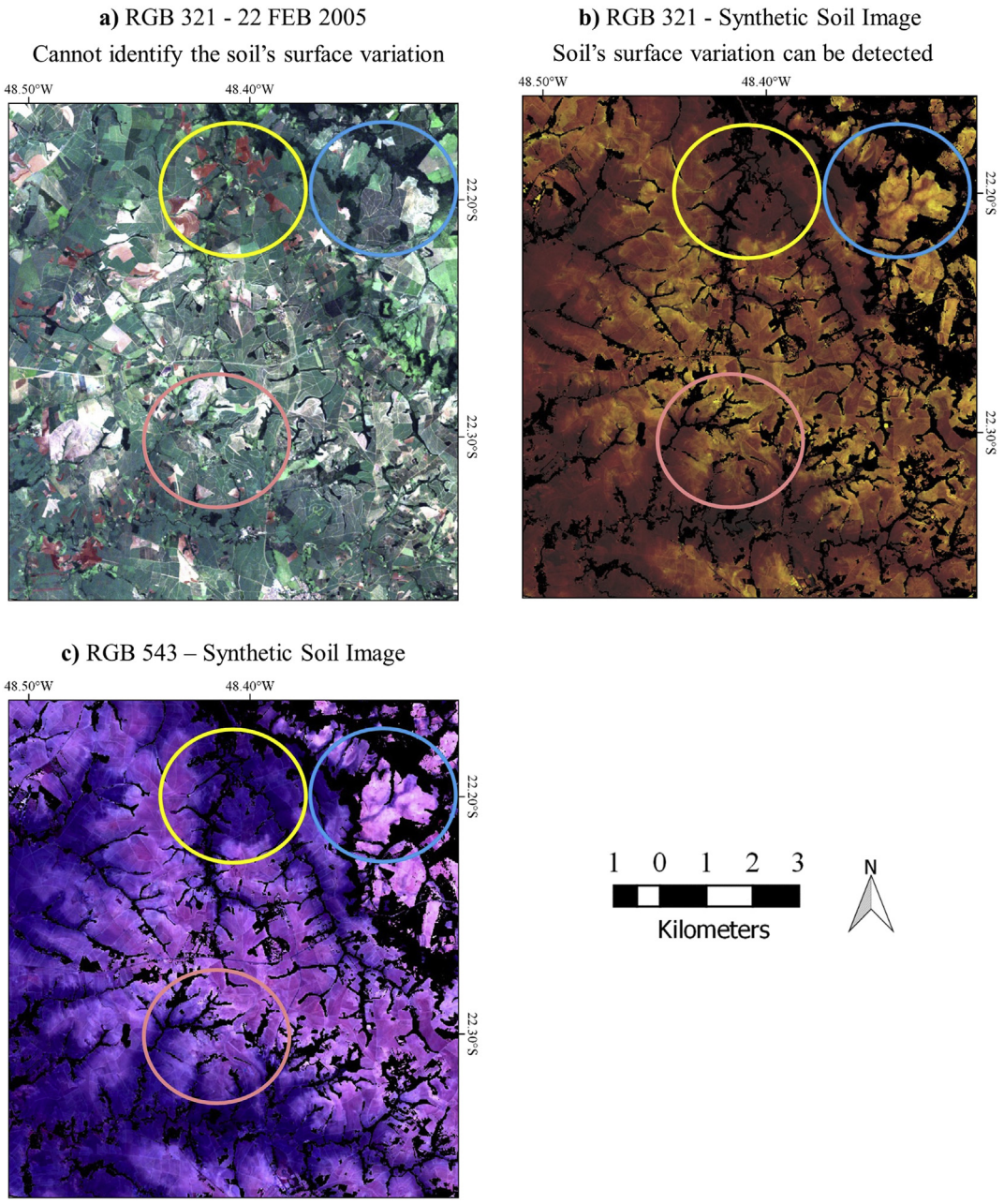

Fig. 10. Selected case inside the entire study site, illustrating a) a single Landsat 5 TM scene in RGB 321; b) Soil Synthetic Image (SYSI) in the RGB 321 composite with visual enhancement; c) SYSI in RGB 543 composite with visual enhancement.

between image spectral reflectance and topsoil spectra was a direct validation of the accuracy of GEOS3.

The method was supported by a robust direct validation using a total of 919 topsoil samples. TESS consistency was thus confirmed by the association with the laboratory spectroscopy reflectance analysis. For instance, an image subset (RGB 543 composite) was used to show the spectral similarity between the sources (Fig. 2b). It was observed that, in general, three main topsoils are related to the colors presented in the image - dark purple, intermediate purple and magenta. Each spectral reflectance curve had a different pattern regarding shape and intensity. The areas with dark purple are related to soils with high contents of opaque minerals (ilmenite and magnetite) and hematite. From purple to magenta shades, these minerals decrease, and the higher proportion of quartz causes an increase in reflectance.

Although the data obtained in the laboratory were related to dry sieved soil, the intensities and shapes were quite similar for both the sources (Fig. 2b). In the case of TESS, there is a clear difference between B5 and B7 due to the field condition factors, whereby its spectral reflectance was more affected by moisture and other environmental factors (Khanna et al., 2007). Moreover, although the laboratory spectra were taken from samples of a single survey, the convolved spectral reflectance is very similar to the TESS, confirming that it is an accurate spectral reflectance representation of topsoil samples (Fig. 2b).

3.3. Potential of synthetic soil image (SYSI)

The GEOS3 used a time series (1984–2011) from Landsat 5 TM to detect bare soil areas, especially, but not exclusively, for agricultural areas. It is similar to "removing" the land cover (Fig. 10a) and providing only the picture of the soil surface for an entire area (Fig. 10b and c). In areas with natural forest, GEOS3 cannot run correctly (see the black regions in Fig. 10b and c). The spatial variation of soils in the

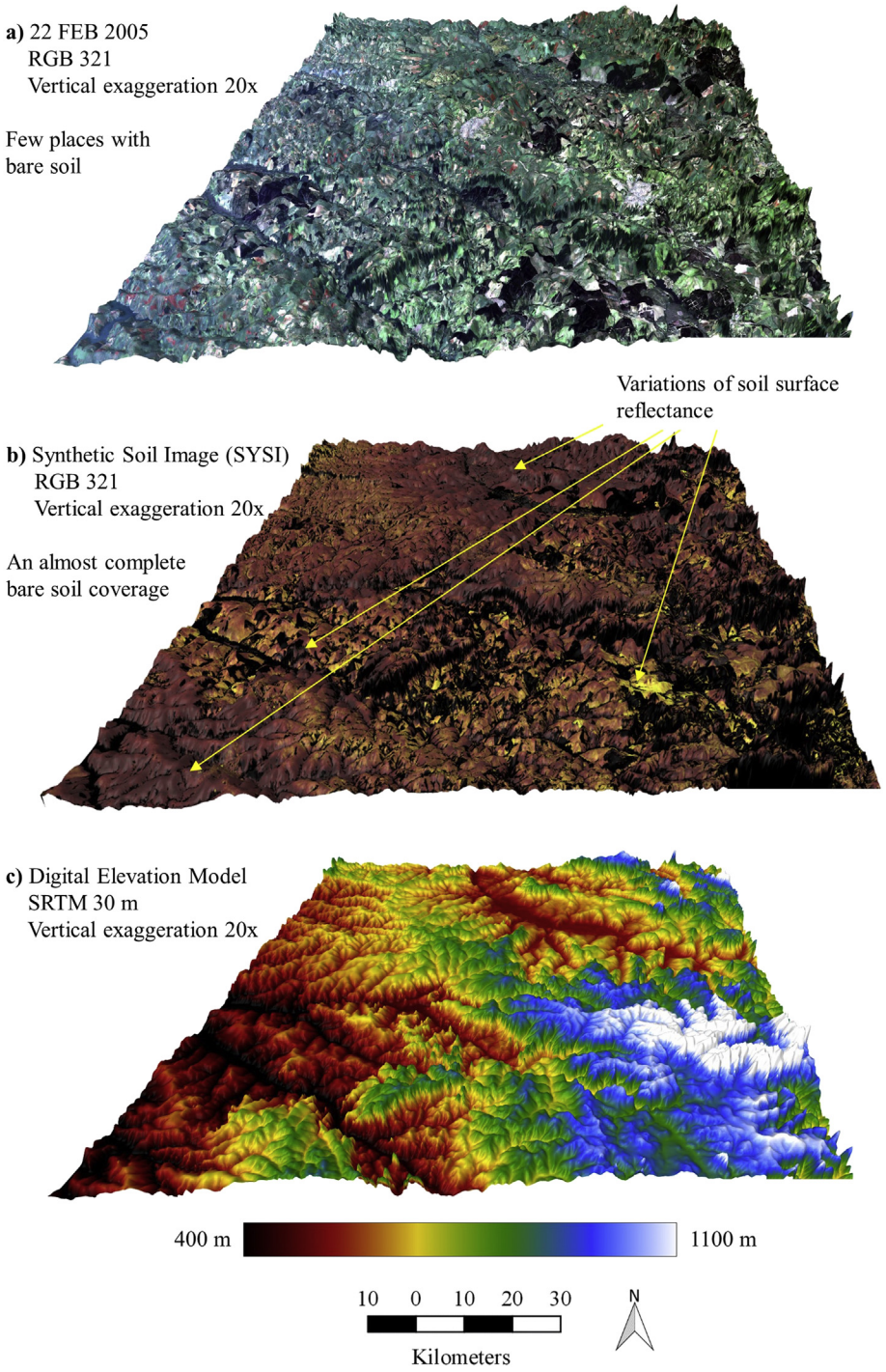

Fig. 11. 2.5-D representations of composites from the study site: a) a single Landsat 5 TM scene in the RGB 321; b) Synthetic Soil Image in the RGB 321; c) Digital Elevation Model.

study area can be seen in the true color composite of the SYSI (RGB 321, Fig. 11b). The different shades are related to the different topsoil compositions, where redder soils have higher hematite content than the yellow ones. These aspects of the image can reveal further information, which usual digital elevation models (DEM) and single images (Fig. 11a) are unable to provide. For instance (Fig. 11b), SYSI provides additional information on soil surface that was not evident in DEMs (Fig. 11c). Therefore, the SYSI can be integrated with other spatial data (i.e. DEM) to assist in spatial modeling and digital soil mapping.

Recently published papers have also proposed methodologies for the generation of a bare soil composite based on satellite time series (Demattê et al., 2016; Diek et al., 2017; Rogge et al., 2018). Demattê et al. (2016) combined 5 years of images and achieved 85% bare soil, albeit in a very specific small region. Rogge et al. (2018) developed the Soil Composite Mapping Processor (SCMaP), which is an automated process to overcome the limited availability of bare soil in satellite images. The researchers applied the SCMaP over Germany to produce spatial and temporal soil composites for further large-scale topsoil analyses in temperate zones. Additionally, Diek et al. (2017) developed a method to maximize bare soil coverage over a large agricultural area, creating the Barest Soil Composite for properties modeling.

Although the mentioned studies have the same objective, the

methodologies require distinct processes to generate and validate the bare soil composite. For example, Rogge et al. (2018) proposed the generation of a new index for soil masking, instead of using common spectral indices, e.g. NDVI. The GEOS3 masking process uses the NDVI and NBR2, both calculated and provided by the ESPA ordering interface for Landsat data. Additionally, the present study proposed the use of selected products of a full satellite time series, in this case the use of Landsat 5 TM, which includes surface reflectance bands, spectral indices and quality assessment bands. The GEOS3 also provides the temporal frequency of bare soil exposition, similar to the study of Rogge et al. (2018). This information can be extended to future soil degradation research, improving the understanding of the relationships between remote sensing indicators and soil properties (Shoshany et al., 2013). Furthermore, the strong correlation between topsoil spectra and synthetic soil reflectance stresses the validation of the remote sensing products and provides support to the use of the spectral reflectance from the SYSI for soil modeling in future studies.

The basis for understanding image spectral reflectance is the laboratory information. Thus, it is important to look at the shapes, intensities and the entire soil spectral signature, to guarantee the quality of the spectral measurement. In fact, field data may be subject to several alterations as a result of soil management (Demattê et al., 2016). However, it can be observed a similarity between spectra obtained from the SYSI and laboratory convolved spectra (Fig. 2b). The shape, intensity and tendency, agrees with our empirical and theoretical background (item 2.3).

4. Conclusion and final considerations

The study describes GEOS3, which successfully defined synthetic spectral reflectance of soil from a multi-temporal and multispectral remotely sensed database. Based on 27 years of Landsat products, it was possible to retrieve soil spectral reflectance represented as a spatially continuous surface in a single image. Although the data used in GEOS3 were obtained from different years, the method enabled calculation of median reflectance for each pixel from 151 images to create a TESS, which is a synthetic spectral fingerprint of the soil surface represented in the Synthetic Soil Image (SYSI).

Comparison of laboratory reflectance spectra from 919 soil samples in the same georeferenced location as TESS presented Pearson's correlation between 0.67 and 0.88. The canonical correlation reached a value of 0.93. TESS presented a very similar shape to the spectra obtained in the laboratory. Moreover, soil line technique and PC analysis demonstrated similar patterns between laboratory and SYSI information.

The system was able to identify areas with exposed soils during the time series, which opens doors for other research interests, such as soil degradation and land evaluation. Using one single image, the user can find 0.5% bare soil in the total area. With GEOS3 we achieved 68% in the total area of 14,614 km^2. When considering only areas with agriculture (sugarcane crops), the system reached 92%.

The functionality of GEOS3 is evident, as it could be used to support soil surveys, digital soil mapping (e.g. mapping of clay, sand and carbon contents), soil and agriculture monitoring, precision agriculture, soil sampling allocation, soil management zones, land use planning, soil conservation and for improvements to existing soil maps. The information provided by GEOS3 may thus assist in different decision-making processes.

Although the technique used only Landsat images, the method is able to employ time series data from any satellite. Additionally, the success of the procedure depends on the following key factors: (a) each location in the area must have had bare soil at some period in the time series, with the SYSI quality improving as the availability of bare soil occurrences increases; (b) the original images and their processing should have quality assessment indicators, (c) a database (laboratorial spectral library) is necessary to validate image processing, (d) evaluation of the temporal spectral reflectance (intensity and shape) and (e) the image database should be related to a season that minimizes environmental influences on the soil, especially in regard to soil water content.

Acknowledgements

The financial support, which came from different Brazilian institutions, is hereby acknowledged: grants numbers 2013/18769-9, 2013/20377-1, 2014/22262-0, 2016/01597-9 from the São Paulo Research Foundation (FAPESP), the grant numbers 305080/2007-5 and 130805/2014-9 from National Council for Scientific and Technological Development (CNPq), and the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES). The authors are grateful to Geotechnology in Soil Science Group, GEOSS (www.esalqgeocis.wixsite.com) and Marco Antonio Bortoletto for providing experimental data.

References

Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorol. Z. 22, 711–728. http://dx.doi.org/10.1127/0941-2948/2013/0507.

Baret, F., Jacquemoud, S., Hanocq, J.F., 1993. About the soil line concept in remote sensing. Adv. Space Res. 13, 281–284. http://dx.doi.org/10.1016/0273-1177(93)90560-X.

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. Remote Sens. Environ. 113, S38–S55. http://dx.doi.org/10.1016/j.rse.2008.09.019.

Chagas, C. da S., de Carvalho Junior, W., Bhering, S.B., Calderano Filho, B., 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. Catena 139, 232–240. http://dx.doi.org/10.1016/j.catena.2016.01.001.

Dematte, J.A.M., Huete, A.R., Ferreira Jr., L.G., Nanni, M.R., Alves, M.C., Fiorio, P.R., 2009. Methodology for bare soil detection and discrimination by Landsat TM image. Bentham Open - Open Remote Sens. J. 2, 24–35. http://dx.doi.org/10.2174/1875413900902010024.

Demattê, J.A.M., Alves, M.R., Terra, F. da S., Bosquilia, R.W.D., Fongaro, C.T., Barros, P.P. da S., 2016. Is it possible to classify topsoil texture using a sensor located 800 km away from the surface? Rev. Bras. Ciênc. Solo 40. http://dx.doi.org/10.1590/18069657rbcs20150335.

Diek, S., Schaepman, M., de Jong, R., 2016. Creating multi-temporal composites of airborne imaging spectroscopy data in support of digital soil mapping. Remote Sens. 8, 906. http://dx.doi.org/10.3390/rs8110906.

Diek, S., Fornallaz, F., Schaepman, M.E., de Jong, R., 2017. Barest pixel composite for agricultural areas using landsat time series. Remote Sens. 9, 1245. http://dx.doi.org/10.3390/rs9121245.

Dobos, E., Hengl, T., Reuter, H., 2006. Digital Soil Mapping as a Support to Production of Functional Maps. 68 Off. Off. Publ. Eur. Communitieshttp://dx.doi.org/10.1007/978-90-481-8863-5.

Escuin, S., Navarro, R., Fernandez, P., 2008. Fire severity assessment by using NBR (normalized burn ratio) and NDVI (normalized difference vegetation index) derived from LANDSAT TM/ETM images. Int. J. Remote Sens. 29, 1053–1073. http://dx.doi.org/10.1080/01431160701281072.

Fiorio, P.R., Dematte, J.A.M., 2009. Orbital and laboratory spectral data to optimize soil analysis. Sci. Agric. 66, 250–257. http://dx.doi.org/10.1590/S0103-90162009000200015.

Friendly, M., Fox, J., 2016. candisc: visualizing generalized canonical discriminant and canonical correlation analysis. In: R package version 0.7-2, . https://CRAN.R-project.org/package=candisc.

Gerland, P., Raftery, A.E., ev ikova, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B.K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G.K., Wilmoth, J., 2014. World population stabilization unlikely this century. Science 346 (80), 234–237. http://dx.doi.org/10.1126/science.1257469.

Gonzalez, R.C., 2009. Digital Image Processing. Pearson Education.

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., De Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10, e0125814. http://dx.doi.org/10.1371/journal.pone.0125814.

IUSS, 2015. World reference base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. In: World Soil Resources Reports No. 106. FAO, Rome.

Jensen, J.R., 2013. Remote Sensing of the Environment: an Earth Resource Perspective, 2nd revise ed. Pearson New International Edition.

Kalnay, E., Cai, M., 2003. Impact of urbanization and land-use change on climate. Nature 423, 528–531.

Khanna, S., Palacios-Orueta, A., Whiting, M.L., Ustin, S.L., Riao, D., Litago, J., 2007. Development of angle indexes for soil moisture estimation, dry matter detection and land-cover discrimination. Remote Sens. Environ. 109, 154–165. http://dx.doi.org/

10.1016/j.rse.2006.12.018.
Krzywinski, M., Altman, N., 2014. Points of significance: visualizing samples with box plots. Nat. Methods 11, 119–120. http://dx.doi.org/10.1038/nmeth.2813.
Lal, R., 2004. Soil carbon sequestration to mitigate climate change. Geoderma 123, 1–22. http://dx.doi.org/10.1016/j.geoderma.2004.01.032.
Madeira Netto, J.D.S., 1996. Spectral reflectance properties of soils. Photo Interpret. 34, 59–76.
Marsh, W.M., 2010. Landscape Planning: Environmental Applications. Wiley.
Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A landsat surface reflectance dataset for North America, 1990-2000. IEEE Geosci. Remote Sens. Lett. 3, 68–72. http://dx.doi.org/10.1109/LGRS.2005.857030.
McBratney, A., Mendonça Santos, M., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. http://dx.doi.org/10.1016/S0016-7061(03)00223-4.
McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. http://dx.doi.org/10.1016/j.geoderma.2013.08.013.
Müller, B., Bernhardt, M., Jackisch, C., Schulz, K., 2016. Estimating spatially distributed soil texture using time series of thermal remote sensing: a case study in central Europe. Hydrol. Earth Syst. Sci. 20, 3765–3775. http://dx.doi.org/10.5194/hess-20-3765-2016.
Musick, H.B., Pelletier, R.E., 1988. Response to soil moisture of spectral indexes derived from bidirectional reflectance in thematic mapper wavebands. Remote Sens. Environ. 25, 167–184. http://dx.doi.org/10.1016/0034-4257(88)90099-5.
Nanni, M.R., Demattê, J.A.M., 2006. Spectral reflectance methodology in comparison to traditional soil analysis. Soil Sci. Soc. Am. J. 70, 393–407. http://dx.doi.org/10.2136/SSSAJ2003.0285.
Nanni, M.R., Demattê, J.A.M., Chicati, M.L., Fiorio, P.R., Cézar, E., Oliveira, R.B., 2012. Soil surface spectral data from Landsat imagery for soil class discrimination. Acta Sci. Agron. 34. http://dx.doi.org/10.4025/actasciagron.v34i1.12204.
Nolasco de Carvalho, C.C., Nunes, F.C., Homem Antunes, M.A., Nolasco, M.C., 2015. Soil surveys in Brazil and perspectives in digital soil mapping. Soil Horiz. 56. http://dx.doi.org/10.2136/sh14-01-0002.
Pan, G., Li, L., Wu, L., Zhang, X., 2004. Storage and sequestration potential of topsoil organic carbon in China's paddy soils. Glob. Chang. Biol. 10, 79–92. http://dx.doi.org/10.1111/j.1365-2486.2003.00717.x.
Perrotta, M.M., Salvador, E.D., da Lopes, R.C., D'Agostino, L.Z., Peruffo, N., Gomes, S.D., Sachs, L.L.B., Meira, V.T., Garcia, M.G.M., Lacerda Filho, J.V., 2005. Mapa geológico do Estado de São Paulo escala 1:750.000. CPRM, São Paulo.
Pizarro, M.A., Epiphanio, J.C.N., Galvão, L.S., 2001. Caracterização mineralógica de solos tropicais por sensoriamento remoto hiperespectral. Pesq. Agrop. Brasileira 36, 1277–1286. http://dx.doi.org/10.1590/S0100-204X2001001000010.
Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., Heiden, U., 2018. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). Remote Sens. Environ. 205, 1–17. http://dx.doi.org/10.1016/J.RSE.2017.11.004.
Rudorff, B.F.T., Aguiar, D.A., Silva, W.F., Sugawara, L.M., Adami, M., Moreira, M.A., 2010. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using Landsat data. Remote Sens. 2, 1057–1076. http://dx.doi.org/10.3390/rs2041057.
Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., Gao, F., 2013. Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description. US Geological Survey.
Shabou, M., Mougenot, B., Chabaane, Z., Walter, C., Boulet, G., Aissa, N., Zribi, M., 2015. Soil clay content mapping using a time series of Landsat TM data in semi-arid lands. Remote Sens. 7, 6059–6078. http://dx.doi.org/10.3390/rs70506059.
Shoshany, M., Goldshleger, N., Chudnovsky, A., 2013. Monitoring of agricultural soil degradation by remote-sensing methods: a review. Int. J. Remote Sens. 34, 6152–6181. http://dx.doi.org/10.1080/01431161.2013.793872.
Sommer, M., Schlichting, E., 1997. Archetypes of catenas in respect to matter - a concept for structuring and grouping catenas. Geoderma 76, 1–33. http://dx.doi.org/10.1016/S0016-7061(96)00095-X.
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 163–215. http://dx.doi.org/10.1016/S0065-2113(10)07005-7.
Sun, T., Neuvo, Y., 1994. Detail-preserving median based filters in image processing. Pattern Recogn. Lett. 15, 341–347. http://dx.doi.org/10.1016/0167-8655(94)90082-5.
Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M., Morcette, J.-J., 1997. Second simulation of the satellite signal in the solar Spectrum, 6S: an overview. IEEE Trans. Geosci. Remote Sens. 35, 675–686. http://dx.doi.org/10.1109/36.581987.
Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth Sci. Rev. 155, 198–230. http://dx.doi.org/10.1016/j.earscirev.2016.01.012.
Webster, R., 1977. Canonical correlation in pedology: how useful? J. Soil Sci. 28, 196–221. http://dx.doi.org/10.1111/j.1365-2389.1977.tb02306.x.
Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94. http://dx.doi.org/10.1016/j.rse.2011.10.028.