Remote Sensing of Environment 205 (2018) 1–17

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014)

Derek Rogge[a], Agnes Bauer[b], Julian Zeidler[a], Andreas Mueller[a], Thomas Esch[a], Uta Heiden[a,*]

[a] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen, 82234 Wessling, Germany
[b] Hochschule für nachhaltige Entwicklung Eberswalde, 16225 Eberswalde, Germany

ARTICLE INFO

Keywords:
Soil mapping
Image composites
Landsat
Automated processors

ABSTRACT

Soil information with high spatial and temporal resolution is crucial to assess potential soil degradation and to achieve sustainable productivity and ultimately food security. The spatial resolution of existing soil maps can commonly be too coarse to account for local soil variations and owing to the cost and resource needs required to update information these maps lack temporal information. With improved computational processing capabilities, increased data storage and most recently, the increasing amount of freely available data (e.g. Landsat, Sentinel-2A/B), remote sensing imagery can be integrated into existing soil mapping approaches to increase temporal and spatial resolution of soil information. Satellite multi-temporal data allows for generating cloud-free, radiometrically and phenologically consistent pixel based image composites of regional scale. Such data sets are of particular use for extracting soil information in areas of intermediate climate where soils are rarely exposed. The Soil Composite Mapping Processor (SCMaP) is a new approach designed to make use of per-pixel compositing to overcome the issue of limited soil exposure. The objective of this paper is to demonstrate the automated processors ability to handle large image databases to build multispectral reflectance composite base data layers that can support large scale top soil analyses. The functionality of the SCMaP is demonstrated using Landsat imagery over Germany from 1984 to 2014 applied over 5 year periods. Three primary product levels are generated that will allow for a long term assessment and distribution of soils that include the distribution of exposed soils, a statistical information related to soil use and intensity and the generation of exposed soil reflectance image composites. The resulting composite maps provide useful value-added information on soils with the exposed soil reflectance composites showing high spatial coverage that correlate well with existing soil maps and the underlying geological structural regions.

1. Introduction

Soils are essential for human existence as they carry out a number of key environmental functions that directly impact food, fibre and timber production, water storage and redistribution, pollutant filtering and carbon storage. The understanding and modelling of the physical, chemical and biological processes underlying the different environmental functions has been the subject of many scientific works (Sanchez et al., 2009). Soil information is required as input for modelling studies, such as climate change (e.g. Omuto et al., 2013) or estimation of carbon storage in soils (e.g. Wiesmeier et al., 2014). High spatial and temporal resolution soil information is also crucial for agriculture in order to maintain or increase agricultural productivity, to assess and avoid soil degradation and ultimately to achieve sustainable food security (Barnes et al., 2003; Mulder et al., 2011; Omuto et al., 2013; Valero et al., 2016). This includes information on water holding capacity, soil organic carbon, nutrient availability, soil texture and compaction, erosion potential and the impact of different site management practices on soil properties (Barnes et al., 2003).

Existing soil mapping products such as the Harmonized World Soil Database (1:5,000,000, Nachtergaele et al., 2009) and the European Soil Database (1:1,000,000, Panagos et al., 2012) already provide information about different soil properties on a coarse scale. More detailed information can be found in national or regional products, such as the BÜK200 (BGR, 2015) of Germany (Bodenübersichtskarte – Soil base map, 1:200,000, 2015) or the SBK25 (LfU Bayern, 2015) of Bavaria (Standortkundliche Bodenkarten – Local soil maps, 1:25,000, 2015). However, the spatial resolution of these products can commonly be too coarse to account for local soil variations at the scale of 100's of meters or less. Additionally, the existing products have low temporal resolution owing to the cost and resource needs required to acquire the necessary data (Eckelmann, 2004), primarily in the form of soil cores

* Corresponding author.
E-mail address: uta.heiden@dlr.de (U. Heiden).

https://doi.org/10.1016/j.rse.2017.11.004
Received 4 May 2017; Received in revised form 29 October 2017; Accepted 1 November 2017

sampled at regular grid intervals, which can be quite large. For example, at the European scale the LUCAS survey (Land use and land cover survey, European Commission) included 19,969 soil cores with 1 core per 199 km^2 resulting in grid cell 14 km (Hengl, 2006). At a more regional scale, such as for the state of Bavaria (70,550 km^2) in Germany, 1460 soil cores from different soil surveys and permanent soil observation sites of the Bavarian Environment Agency (LfU) and the Bavarian State Institute for Forestry (LWF) were screened and used to map soil organic carbon storage resulting in a sampling density of one location per 48 km^2, or a grid scale of 7 km (Wiesmeier et al., 2014). Thus, there is a need to further improve the spatial-temporal coverage and detail of soil maps.

A promising approach that can be integrated into existing soil mapping approaches to increase temporal and spatial resolution of soil information is the use of remote sensing information (Mulder et al., 2011). Generally the use of remotely sensed data has a long tradition in assisting the mapping of soil properties, as early as the use of air photography being used to segment the landscape into soil–landscape units for which soil composition was established by sampling (Mulder et al., 2011). Today more advanced remote sensing data, such as airborne hyperspectral data, can directly provide topsoil properties like texture, soil organic carbon and content of some minerals have now been successfully identified (Mulder et al., 2011; Barnes et al., 2003). Much of the knowledge of soil spectroscopy has been applied through remote sensing that makes use of these hyperspectral sensor systems comprising 100's of spectral channels in the VIS-NIR-SWIR spectral region as this allows for the capture of broad and narrow absorption features related to soil chemistry of the organic and inorganic composition of the soils that enable quantitative analysis of several soil properties (e.g. Malley et al., 2004; Ben-Dor et al., 2009; Shepherd and Walsh, 2002; Nanni and Demattê, 2006). These types of sensor systems are, unfortunately, limited to laboratory, field and airborne platforms, which do not have the regional spatial and repetitive temporal coverage of spaceborne multispectral systems, such as Landsat or Sentinel-2. On the other hand, multi-spectral satellite systems are limited to a few broad spectral bands (usually < 10) which are focused on spectral regions more appropriate for the analysis of phenological changes of vegetation over a given growing season. However, the spectroscopic information on soil properties available from multispectral systems does allow for a regional scale assessment of broad soil characteristics spatially and temporally (e.g. color, brightness). These characteristics are related to the genetic nature of the soils (soil horizons, surficial geology (e.g. Boettinger et al., 2008)) or agricultural practices and by linking them to ground based databases of soil cores they can be used to extend and/or enhance spatial detail, distribution and temporal changes of existing soil maps.

There are many challenges to using satellite remote sensing for assisting in the mapping of soils. Users need to understand the capabilities and limitations of the image data and derived products (e.g. access, timely delivery of data, standardization of imagery, pre-processing and derived products, spatial and spectral resolution of sensors, frequency and consistency of image acquisition, and cloud and haze cover). Most importantly in the context of soil mapping is the fact that in many temperate and continental climates, such as central Europe, the use of remote sensing is hindered by layers of photosynthetic active and non-photosynthetic active vegetation (branches, stem, senescent foliage; Asner et al., 1998, Zhang et al., 2005), covering the soil either permanently or, like in the case of agricultural regions, during a dominant percentage of the year. Thus, for soil mapping in these regions it is necessary to have usable cloud-free imagery collected often enough and at consistent intervals over the growing season to capture exposed soils.

Until more recently high costs, accessibility and low computational capacity restricted the use of large collections of remotely sensed, high resolution data sets. However, with improved computational processing capabilities, increased data storage and most recently, the increasing amount of freely available data, these issues are being overcome. In terms of data accessibility the opening of the Landsat archive in 2008 (Woodcock et al., 2008) has provided global-scale high spatial resolution Earth observation data from 1982 till today. A number of large scale products are already becoming available that use Landsat data, such as top of atmosphere reflectance, surface reflectance, BRDF corrected, and terrain normalized (e.g. Hansen and Loveland, 2012; Wulder et al., 2012; Turner et al., 2015). In a review given by Wulder et al. (2012) the application of the newly available Landsat data includes many thematic applications such as, forests (change, fragmentation, status, burnt areas), rangeland (productivity, modification), urban (impervious surfaces, landscape patterns), coastal (shoreline, reefs), water quality, glaciers, mining, phenology, wetlands, and crops. In addition to the Landsat archive, image data from the Copernicus ESA Sentinel-2 mission for high resolution multi-spectral data is also open, in which the first of two satellites has been operational since November 2015 (ESA, 2015) and the second launched in March 2017. Moving towards the future there will also be the addition of planned hyperspectral satellite data from EnMAP (Guanter et al., 2015), HISUI (Matsunaga et al., 2014), HyspIRI (Abrams and Hook, 2013), HYPXIM (Carrere et al., 2013), PRISMA (Lopinto and Ananasso, 2013), and SHALOM (Bussoletti, 2012), which will provide increased spectral information related to soils. Having such data sources provides a unique opportunity to have a large number of images available spatially and temporally that would allow for building an "archive" of exposed soil.

The addition of the increased availability of data brings forth a new paradigm in remote sensing comprising pixel based image compositing that allows for the generation of cloud-free, radiometrically and phenologically consistent regional scale images (e.g. Roy et al., 2010; Hansen et al., 2011; Zhu and Woodcock, 2014; Yan and Roy, 2014; White et al., 2014; Hermosilla et al., 2015). Notably is the Web-Enabled Landsat Data (WELD) approach (Roy et al., 2010; Hansen et al., 2011) that builds weekly, monthly and seasonal composites that are generated using a set of criteria such as maximum Normalized Difference Vegetation Index (NDVI) (Tucker, 1979), maximum brightness temperature, maximum apparent surface temperature, maximum difference between the red and NIR channels, minimum scan angle or some combinations of these. Such composites have been used to build a vegetation continuous field with percent tree cover, other vegetation cover, bare ground and water. In a study by Yan and Roy (2014), WELD composite NDVI products that consider temporal variations in the agricultural cover were used as input into a variation regional-based geometric active contour segmentation method (VRGAC) to extract crop field segments with 90.1% overall accuracy. However, to improve omissions, more sophisticated understanding of the phenology of crops, forests, and grasslands is required to differentiate these classes. Landsat per-pixel composite rules were also used for mapping regional-scale boreal forest cover and change for European Russia (Potapov et al., 2011). Pixel-based annual, multi-year and proxy-value image composites were also used by White et al. (2014) to generate information products characterizing land cover, change and forest structural information. Best available pixel (BAP) composites have been generated for all of Canada for the year 2010 and for 15 years (1998–2012) for the provinces of Saskatchewan and Newfoundland (Hermosilla et al., 2015).

Although many studies are using this valuable data for monitoring vegetation, crops, and land cover use and change, its use for extracting exposed soil information is limited, particularly in areas of temperate and continental climates where soils are commonly covered by vegetation most of the year. One study that considered this question was presented by Demattê et al. (2016), in which they developed a strategy to identify continuous areas of soil exposure using an approach of "fusing" multiple scenes acquired at specific times of the year when soil exposure was high. The methodology presented uses a linear spectral mixture model approach to discriminate soils from vegetation (Demattê et al., 2009) which is subsequently followed by a series of image-processing procedures using NDVI and band combinations to indicate the true nature of the pixel. This process is performed independently on the

available scenes with the respective portions of bare soils obtained by "fusing" the results of each scene providing a mosaic of exposed soils. In this study a total area of 65,000 ha was assessed over a five-year cycle allowing for exposure of soil through tilling practices.

A new fully automated approach designed to make use of per-pixel compositing to overcome the issue of limited soil exposure is presented in this study and is called the Soil Composite Mapping Processor (SCMaP), which is presently being developed within the Operational Platform for Provision and Processing of Sentinel data for Global Monitoring for Environment and Security (DLR, 2011). The approach developed here makes use of multiple scenes to build a more continuous map of exposed soils. Such a conceptual approach is applicable in many temperate regions, such as central Europe, where extensive crop-based agriculture provides for the exposure of soil for short periods of time. However, the method presented here differs to Demattê et al. (2009) in a number of aspects. First we make use of all available scenes over a given time period, rather than at a specific time of the year, which increase the potential of extracting bare soil exposure that is variable across a given year. Secondly, we extract bare soils using information from a stack of input scenes instead of from individual scenes which are subsequently "fused". Third, using this stack of scenes, periods of maximum and minimum photosynthetic active green vegetation can be assessed on a per-pixel basis to derive two composite images, which are subsequently combined (intersected) to separate exposed soil from non-vegetated areas like urban and water, and permanently vegetated areas like forest and grasslands. The result is a soil mask which contains all pixels that include exposed soil at some point over a given time period. Using this mask, surface reflectance composites, that average all available bare soils for given pixel, can be produced. Such reflectance composites then form the basis for deriving additional information on soil use and properties.

The objective of this research is to develop an automated processor to handle large image databases and build multispectral reflectance composite data layers that will support and enable large scale top soil analyses applied to temperate regions across the globe. To demonstrate the functionality of the processor we start first with Germany using Landsat imagery from 1984 to 2014 applied over five year periods. The SCMaP will provide three primary product levels that will allow for a long term assessment and distribution of soils, which comprise 1) the spatial and temporal distribution of exposed soils; 2) a statistical assessment of the temporal frequency of exposed soils that can be linked to soil use and intensity; and, 3) generation of soil reflectance image composites. Comparing the resulting composites with existing soil and geological maps of Germany, we demonstrate how the SCMaP composites can highlight the spatial distribution at a 30 m pixel scale of broad soil groups even with the limited spectroscopic information available for soils. Subsequent work will focus on linking the soil composites to existing soil information (maps and cores) to develop higher order products that can assist in mapping specific soil parameters.

2. Study area

The study area for the SCMaP demonstration is focused on Germany (357,168 km^2; latitudes 47° and 55° N and longitudes 5° and 16° E), but also includes the border areas around the country. Germany has elevation ranges from the highest point (2962 m above sea level – a.s.l.) in the mountains of the Alps to sea-level at the North and Baltic Seas. The majority of the country has a temperate seasonal climate, ranging from temperate sub-oceanic to temperate sub-continental climate. The land comprises approximately 35% arable land, 14% pastures, 30% forest and woodland, 3% water bodies, 12% covered by human structures, and the remaining 8% listed as other (Statistische Ämter der Länder, 2015).

The regional surface geology of Germany (BGR, 2016) can be divided into three physiographic regions. In the northern third of the country is the Central European Depression comprising late Palaeozoic and Cenozoic sedimentary rock sequences. The Central European Blocks, which are a product of the Alpine Orogeny, include the central uplands and south German Scarplands, the upper Rhine plain and foothills of the Alps. The Alpine-Carpathian Arc comprises the mountainous terrain of the Alps and is limited to the very south of Germany, but has a high geological diversity (Plant et al., 2003). Quaternary glaciation has resulted in extensive glacial cover in the northern third of the country. Impacts from older glacial periods in the central parts of Germany are less pronounced. Pleistocene glacial periods in the Alps have also resulted terminal moraines that extend well into the foreland region.

The soil regions of Germany are grouped into 12 main areas defined by different soil forming factors (BGR, 2015). The northern third of Germany (Central European Depression) includes the Holocene Coastal Plains along the North Sea comprising mostly tidal-flats and marshland soils (regosols) and an expansive region of glacial drift comprising luvisols, podzoluvisols, gleysols, cambisols, podzols and histosols. The central part of Germany (Central European Blocks) includes six main areas comprising of loess and sandy loess areas (luvisols, chernozems, gleysols, podzoluvisols and regosols) and mountain and hilly areas with soils derived from various types of non-metamorphic sedimentary, igneous and metamophic rocks (luvisols, leptosols, cambisols, gleysols, regosols and podzols). In the very south of Germany (Alpine-Carpathian Arc) the Alpine foreland includes quaternary gravel plateaus and hills (cambisols, luvisols, gleysols and regosols) and in the Apls soil formation is mainly influenced by the relief which results in immature soils (regosols, cambisols and gleysols). In addition a number of major fluvial plains across Germany (fluvisols and gleysols) associated with the major river systems of the Oder, Danube, Rhine, Weser and Elbe that form the last main soil region of Germany.

3. Landsat data

For this study all available Landsat 4, 5 and 7 (Scan Line Corrector (SLC) on and off) were downloaded for 36 path/row combinations (Paths 192–197, Rows 22–27) that cover the entirety of Germany from 1984 to 2014 (Fig. 1). Landsat Thematic Mapper (Landsat 4 and 5) has a scene size of 170 km north-south by 183 km east-west, comprises six broad spectral bands ranging from 0.45 to 2.35 μm with a spatial resolution of 30 m and one additional thermal band (10.4–12.5 μm) at 120 m spatial resolution, resampled to 30 m. Landsat Enhanced Thematic Mapper Plus (Landsat 7) has a very similar configuration, but with the addition of a panchromatic band having 15 m spatial resolution. In May 31, 2003, the SLC on Landsat 7 failed resulting in usable data, but with linear gaps in the data. The downloaded data is compressed GeoTiff formatted data, with a Universal Transverse Mercator projection (WGS 84 datum) and includes standard terrain correction (L1T) and systematic correction (L1G) (Tucker et al., 2004), the latter being filtered owing to less accurate geocorrection.

4. Methodology

Generation of the soil composite maps comprises three different automated modules which include: 1) download and ingestion; 2) preprocessing and data filtering; and, 3) the SCMaP (Fig. 2). Modules 1 and 2 are part of the TimeScan framework (Esch et al., 2017 submitted) that are integrated into a processing infrastructure that includes a Calvalus cluster based on Apache Hadoop (Apache Foundation, 2017a), an Apache Mesos-orchestrated private cloud (Apache Foundation, 2017b), and a High Performance Cluster (IT4Innovation, 2017). Preprocessing steps are applied to individual Landsat path/row images to build an image database that can be used for multiple tasks and research development. SCMaP is applied using 1 × 1° tiles to take advantage of overlapping portions of the Landsat scenes and for efficient data handling and processing. The following sections focus on the SCMaP methodology (Section 4.2) with the download and

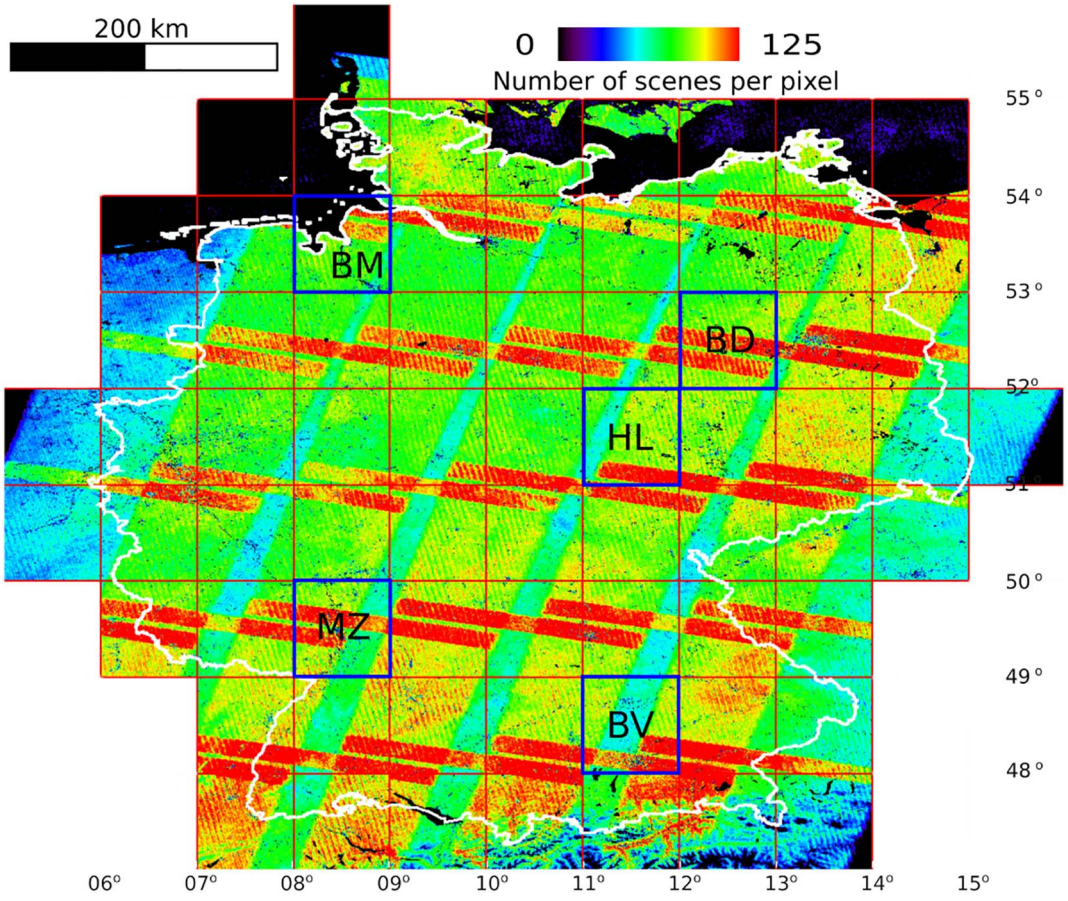

Fig. 1. Study area showing the overlapping Landsat scenes (path and row overlaps) used in the per pixel compositing for the 36 path/row combinations used (Geographic Lat-Long projection, WGS84). The number of scenes per pixel shown here is an example for the 2000–2004 time period after all data filtering and processing steps (refer to Sections 4.1 and 4.2 for details). The outline of Germany is shown in white. Red squares represent 1 × 1° tiles used in the SCMaP processing (refer to Section 4.2.1 for details). Blue squares are the tiles used for the threshold determination (refer to Section 4.2.3 for details) (BM-Bremen, BD-Brandenburg, HL-Halle, MZ-Mainz, BV-Bavaria). Scale bar calculated based on 51.5° north. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

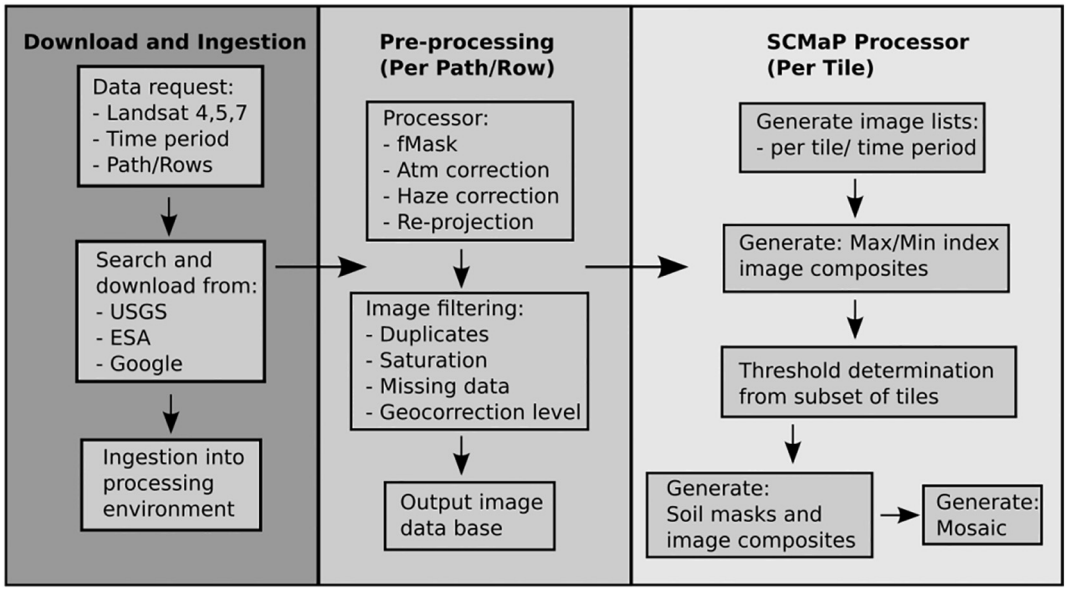

Fig. 2. Overview of the 3 main modules required to generate soil composites.

Table 1
Number of scenes retained after preprocessing and filtering steps.

Number of scenes	Process	Removed
38,033	Downloaded and ingested files for pre-processing	
25,193	Fmask, Atmospheric correction, haze removal, projection	12,840
19,657	Check for correct geocorrection level (L1T)	5536
19,435	Check for error files: e.g. saturation, missing data	222 (LT5 only)
15,588 (41% of original files)	Check for duplicates/triplicates: scenes with more than one file 3690 (only LT5 and LE7); order for retaining scenes USGS > GOOGLE > ESA	3847 (645 LT5, 3202 LE7)

Average of 520 scenes per year for all of the 36 path/rows; 14 scenes per year, per individual Path/Row combination.

preprocessing (Section 4.1) described only briefly.

4.1. Download, ingestion and preprocessing

The data download allows for multi-parameter queries (e.g. sensor, time period, path/row, archive). In this study a total of 38,033 scenes downloaded from USGS, Google and ESA archives and then ingested into the processing environment (Table 1). Pre-processing begins with the Fmask algorithm (Zhu and Woodcock, 2012; Zhu et al., 2015) for generating cloud, shadow, and snow masks. Next, atmospheric correction and haze removal is applied to each Landsat scene using ATCOR (Richter et al., 2006; Richter, 2010) using an automated processor that applies scene specific parameters (e.g. water vapor, aerosols) and image metadata information (e.g. date of acquisition, solar zenith angle, solar elevation, Lat-Long). The use of at surface reflectance allows for a direct comparison with existing laboratory and field spectral libraries with the composite pixel reflectance spectra to be generated. Processing is excluded for the first 30 and last 30 days of the year (winter scenes) removing approximately 17% of the available scenes. Additionally, 17% of scenes were not properly processed due to errors during the Fmask process. The resulting at surface reflectance data and Fmasks are projected to a geographic Lat-Long projection (WGS84). The preprocessed scenes were then passed through a series of filters to remove duplicates, incorrect geocorrection (L1G), and images with saturated and/or missing data (assessed per band) resulting in 15,588 usable scenes comprising the input database for SCMaP (Table 1).

4.2. SCMaP

Naturally exposed soils in temperate and continental climates such as Germany occur rarely. Some exceptions are in the Alps above the tree line, as shore lines or along rivers. Agricultural regions can provide areas of exposed soil for short periods of time owing to farming practices. These areas change from exposed soil to photosynthetically active vegetation (PV) cover, potentially more than once per year. The SCMaP makes use of this characteristic to build an exposed soil mask, which is accomplished using a vegetation index, such as NDVI, which can distinguish clearly between PV (high NDVI values) and exposed soils (low NDVI values). However, other materials, such as water, many urban materials and non-photosynthetic vegetation also have low NDVI values, such that deriving a set of thresholds to separate soil from these other materials is not possible. By using temporal data it is possible to separate exposed soil from other materials by building composite images that consider the maximum and minimum occurrences of PV on a per pixel basis over a stack of images. In this case an agricultural field with mature crops will have a high PV value in the maximum composite, whereas a low PV value after tilling in the minimum composite. By intersecting all pixels having high PV values within the maximum composite and low values in the PV minimum composite a mask of exposed soils can be generated. This mask can then subsequently be used to provide a statistical assessment of exposed soils that can be linked to soil use and intensity and also build a soil reflectance image composite that can be used to assist in mapping soil parameters.

Key to the success of the SCMaP composites is the identification of appropriate maximum and minimum PV thresholds. These thresholds are derived from a detailed analysis of PV maximum and minimum composites for key land cover classes based on existing land cover data and image information. Five different regions (Fig. 1) were chosen in different parts of Germany to derive these thresholds, which should also be transferable across the full study region and over the different investigation periods.

4.2.1. Image tile lists

The SCMaP produces composites for a pre-defined tile size, in this case 1 × 1° geographical tiles. Thus, the first step is to generate a file list of all intersecting path/row scenes and their associated Fmasks for a given tile over a set time period. This allows each tile to make use of overlapping path/row combinations increasing the number of images in the stack with potentially clear imagery and exposed soil available at a per-pixel basis. Note, overlapping regions of adjacent images along the orbital path represent the same source of data. As such, subsequent calculations take this into account and use only one of the images available for the overlap area for a given day. In this study we divide the data into six time periods each making use of five years of data. An exception is made for the first time period which has six years, 1984–1989.

4.2.2. Building PV maximum and minimum composites

The second step is to build two composite images showing maximum and minimum PV on a per-pixel scale, where $p_{(x,y),t} = (p_{(x,y),1}, p_{(x,y),2}, \ldots, p_{(x,y),\ Tm})$ at time t across the total number of time steps $Tm_{(x,y)}$. For the given pixel only those time steps that are cloud/shadow/snow free based on Fmask results are included. To build the maximum and minimum PV composites we use a modification of the normalized difference vegetation index,

$$PV = ((NIR–RED)/(NIR + RED)) + ((NIR–BLUE)/(NIR + BLUE)) \quad (1)$$

The addition of the second normalized difference index using the *BLUE* channel is to minimize the impact and selection of reflectance images that still contain thin haze that were not adequately corrected for in the preprocessing steps. This is important when generating the soil reflectance composites to remove increased reflectance in the *BLUE* channel caused by thin haze. For each pixel $p_{(x,y)}$ the $PVmax_{(x,y)}$ and $PVmin_{(x,y)}$ are calculated as,

$$PVmax_{(x,y)} = max_{t\in[1,Tm(x,y)]}(PV_{(x,y),t}) \quad (2)$$

$$PVmin_{(x,y)} = min_{t\in[1,Tm(x,y)]}(PV_{(x,y),t}) \quad (3)$$

4.2.3. Determination of thresholds

From the *PVmax* and *PVmin* composite images, thresholds can be applied to generate the soil mask. Determining the *PVmax* and *PVmin* composite image thresholds can be done for each tile region. However, given that in this study 70 tiles are required to cover the full extent of Germany, we use a subset of five tiles to determine optimal thresholds that can be applied across the full study area and over the six time-periods assessed.

Table 2
Tile extents and number of sample pixels per tile for the threshold determination.

Tile name	Tile extent		Total number of sample pixels					
	Extent east (degrees)	Extent north (degrees)	Coniferous	Deciduous	Fields	Grassland	Urban	Water
Mainz	8–9	49–50	5000	4999	5000	5000	4938	453
Brandenburg	12–13	52–53	5000	4999	5000	5000	4538	230
Bremen	8–9	53–54	5000	5000	5000	4999	3811	1166
Halle	11–12	51–52	4998	5000	4984	5000	4684	650
Bavaria 1984–89	11–12	48–49	4979	4981	5000	4998	4832	781
Bavaria 1990–94	11–12	48–49	4979	4981	5000	4998	4932	1098
Bavaria 1995–99	11–12	48–49	4979	4981	5000	4998	4852	217
Bavaria 2000–04	11–12	48–49	4979	4981	5000	4998	4890	201
Bavaria 2005–09	11–12	48–49	4979	4981	4994	4998	4654	855
Bavaria 2010–14	11–12	48–49	4979	4981	4971	4998	4222	5

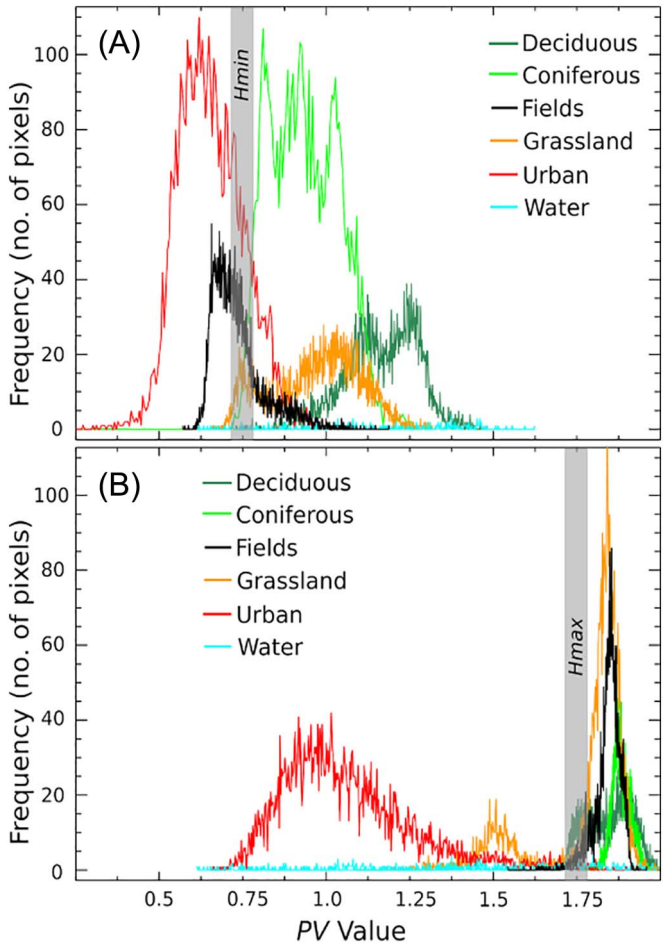

Fig. 3. Example of histograms for different classes derived from *PVmin* and *PVmax* composite images. (A) *PVmin* composite shows overlap between exposed soil fields and urban materials, with partial overlap with NPV found on grassland and in deciduous forests. (B) *PVmax* composite shows overlap between fields and PV grasslands and forests. Grey vertical bars show range of potential *Hmax* and *Hmin* thresholds that can be used the extract fields from all other classes.

A crucial point is to set the thresholds such that the classification of the soil mask is temporally and spatially transferable and that false positives are avoided. To achieve this goal the thresholds were empirically determined by analyzing the *PV* characteristics of six land cover classes (coniferous forest, deciduous forest, agricultural fields, grassland, dense urban areas and water bodies). As there was no ground data taken during this study, 5000 sample pixels per class and tile were collected based on CORINE Land Cover data (EEA, 2007). These sample pixels were subsequently checked with Landsat images to ensure unvarying land use over the investigation period and spatially homogeneity, as the CORINE Land Cover data does not include variations smaller than 25 ha.

For spatial transferability the five tiles were chosen as investigation areas such that they cover a range of climate regions and vegetation diversity (Table 2). Additionally, effort was made to cover the range of soil parent materials and topsoil carbon content as they are known to influence vegetation indices (Huete, 1988). For all investigation tiles the *PVmax* and *PVmin* composite images were calculated for the five year period from 2000 to 2004 (chosen as it has the greatest overlap of data between Landsat 5 and 7, with minimum SCL failure). In order to assess the effect of temporal variations all six investigation periods were tested for the Bavarian tile (Table 2). From each investigation area the total number of sample pixels may be < 5000 owing to masking. *Hmax* and *Hmin* thresholds are determined for Germany as a whole by integrating the data from the five investigation areas.

In the analyses of the plots of *PVmin* versus *PVmax* (see Section 5.1) some key points should be considered when setting the thresholds. For the *PVmin* composite, fields are commonly exposed soil and result in low *PV* values that overlap with urban materials, which does not allow for good separation between these materials. There is also difficulty in the separation of exposed soils from non-photosynthetic vegetation (NPV) material found on fields (e.g. stubble), grasslands and within forests (e.g. litter), specifically deciduous forests (e.g. Elvidge, 1990; Asner and Heidebrecht, 2002; Okin, 2007). As such, *Hmin* is determined based on highlighting the minimum *PV* values for non-photosynthetic active vegetation (NPV) materials. This approach minimizes the potential false positives of NPV materials included in the soil composites. The resulting *Hmin* threshold then can be used to separate fields and urban materials from grassland, deciduous and coniferous forests and water (Fig. 3A). For the *PVmax* composite, fields are commonly covered with PV and have values that overlap with grasslands, deciduous and coniferous forest regions. However, for the *PVmax* composite urban materials and water are separable from fields. In this case the *Hmax* threshold can be set using the maximum *PV* values for urban materials (Fig. 3B).

4.2.4. Soil mask, exposed soil and change frequency, and reflectance composites

Once the *PVmax* and *PVmin* composite image thresholds *Hmax* and *Hmin* have been determined, they are used to derive two image masks, *A* and *B*. *Hmax* is applied to the *PVmax* image to derive our *A* mask, whereas *Hmin* is applied to the *PVmin* image to derive our *B* mask,

$$A_{(x,y)} = \{1 \mid PVmax_{(x,y)} > Hmax\} \quad (4)$$

$$B_{(x,y)} = \{1 \mid PVmin_{(x,y)} < Hmin\} \quad (5)$$

To separate our exposed soil fields from the other classes in *PVmax* and *PVmin* we take the intersection of these two masks to derive our soil mask, *SM*.

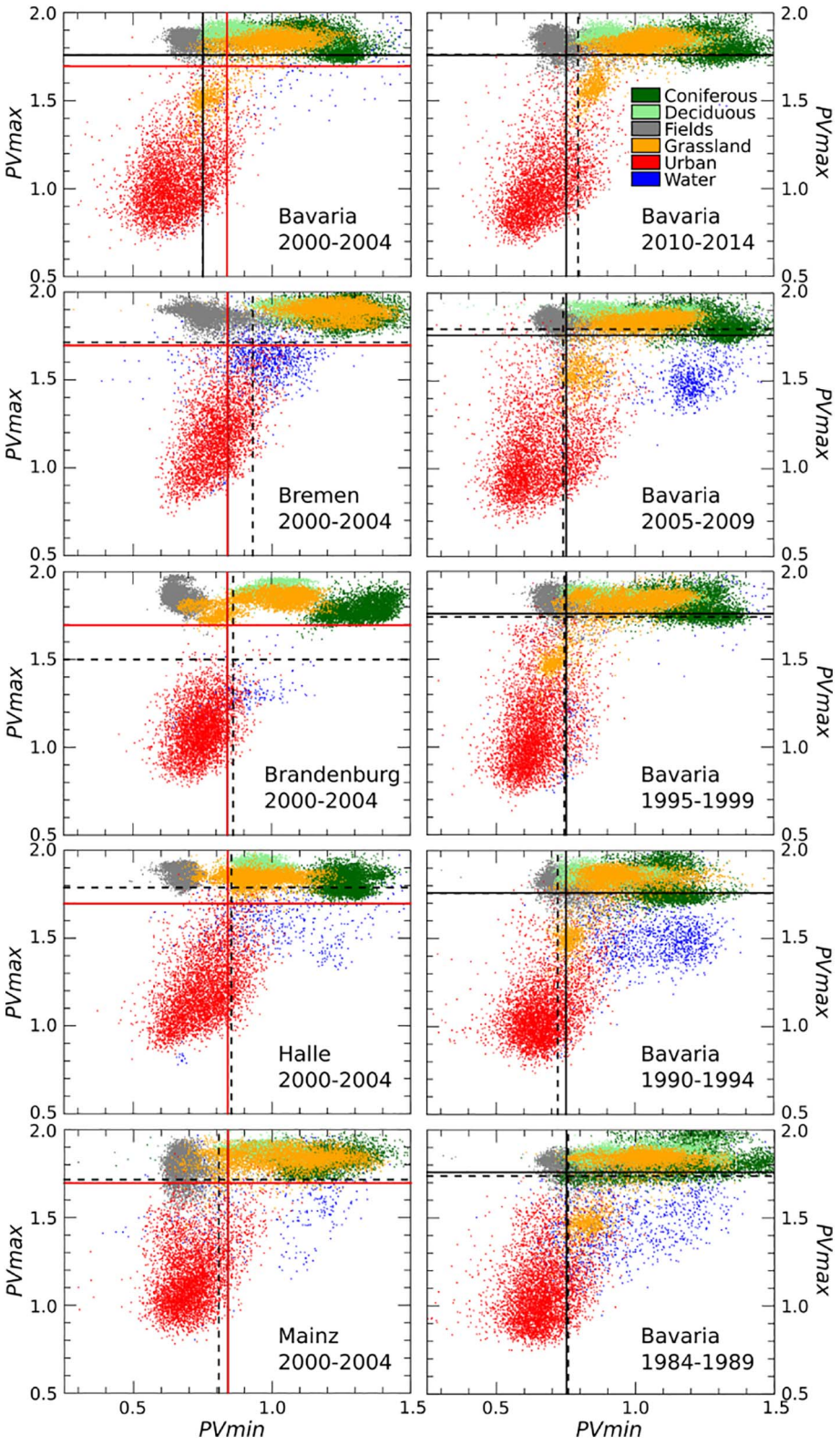

Fig. 4. Scatter plots showing the *PVmin* and *PVmax* characteristics of six land cover types and average *Hmax* and *Hmin* thresholds for the five investigation areas (2000–2004) (red solid line) and for all time periods (Bavaria) (solid black line). The calculated *Hmax* and *Hmin* thresholds for each individual tile are shown as dashed black lines. Note for Bavaria 2000–2004 both sets of thresholds are shown with tile specific *Hmax* and *Hmin* approximately equal to the *Hmax* and *Hmin* for all Bavaria time periods (see Table 3). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$SM_{(x,y)} = A_{(x,y)} \cap B_{(x,y)} \quad (6)$$

With the soil mask generated, per-pixel statistical assessments of exposed soils can be calculated using the stack of data. In this study two factors are calculated. The first is exposed soil frequency (*Sf*), given as a percentage of the number of times a pixel occurs as exposed soil in the data stack. The second statistical assessment is the number of times there is a change (*Cf*) from photosynthetic active vegetation to exposed soil for a given pixel. Unlike *Sf*, *Cf* is not given as a percentage as the number of times a field changes is independent of the number of images in the stack.

The last output generated by the SCMaP is the soil reflectance composite, *SRC*, which is calculated at each $p_{(x,y)}$ as the average reflectance at band *b* of all exposed soils in the data stack for a given pixel, where $PVmin_{(x,y)} < Hmin$ and $SM_{(x,y)} = 1$,

Table 3
Calculated *Hmax* and *Hmin* threshold values across time (Bavaria) and investigation areas.

Time	*Hmin*	*Hmax*	Areas	*Hmin*	*Hmax*
BV1984–89	0.7579	1.7381	BV 2000–04	0.7508	1.7598
BV1990–94	0.7215	1.7572	BD 2000–04	0.8613	1.5004
BV1995–99	0.7443	1.7412	BM 2000–04	0.9307	1.7136
BV2000–04	0.7508	1.7598	HL 2000–04	0.8537	1.7880
BV2005–09	0.7410	1.7947	MZ 2000–04	0.8078	1.7160
BV2010–14	0.7943	1.7632			
Avg BV	0.7516	1.7590	Avg Areas	0.8409	1.6956
Min Bv	0.7215	1.7381	Min Areas	0.7508	1.5004
Max Bv	0.7943	1.7947	Max Areas	0.9307	1.7880
Std BV	0.0243	0.0203	Std Areas	0.0668	0.1135

$$SRC_{(x,y),b} = (SUM_{t\in[1,Tm]|(Pvmin(x,y) < Hmin)t\&SM=1}(p_{(x,y),b,t}))/n \quad (7)$$

where,

$$n = (SUM_{t\in[1,Tm]|(Pvmin(x,y) < Hmin)t\&SM=1}\ (P_{(x,y),t})) \quad (8)$$

Averaging all $p_{(x,y),t}$ in the stack of data allows for the reduction of both natural variability in the exposed soils, such as moisture and texture, but also differences cause by farming practices, such as tilling or spreading of natural fertilizer. In addition to the SRC image, the mean soil reflectance across all bands (*SRCmean*) and the normalized mean reflectance composite (*SRCnorm*) images can be derived. The latter is calculated to highlight spectral shape while reducing amplitude variation. The mean reflectance composite $SRCmean_{(x,y)}$ and the $SRCnorm_{(x,y)b}$ are calculated by the following, where *Nb* is the total number of bands.

$$SRCmean_{(x,y)} = (SUM_{b\in[1,Nb]}(SRC_{(x,y),b}))/Nb \quad (9)$$

$$SRCnorm_{(x,y),b} = SRC_{(x,y),b}/M_{(x,y)} \quad (10)$$

5. Results and assessment

5.1. PVmax and PVmin composites and threshold determination

Fig. 4 shows the *PVmax* and *PVmin* results as scatter plots for the six classes in the five investigation areas and across the six time periods for the Bavaria tile. The *Hmin* threshold was set using the lower 0.005 percentile of deciduous forests for *PVmin* values, which can be assumed to be dominated by NPV. Grassland was not included in the *Hmin* determination as this class showed multiple clusters for some investigation areas. For each area the primary grassland cluster correlates with the deciduous class. However, in some cases such as the Brandenburg area, grassland shows an additional cluster of pixels that overlap with the field class. These pixels were found to be related to flooding in the Brandenburg area. For Bavaria a second grassland cluster was found overlapping partially within the urban class and is associated with variable sub-pixel mixing of vegetation and exposed soil owing to a large military exercise test area. The *PVmax* values are defined using the upper 0.995 percentile for urban areas. Included in each plot are the *Hmax* and *Hmin* calculated for individual tiles/periods (dotted black line) with average values included on the plots for the five investigation areas (solid red line) and across the six time periods for the Bavaria tile (solid black line). Table 3 shows all *Hmax* and *Hmin* calculated values.

Observing first the results for Bavaria over the six time periods the distribution of the class clusters are broadly consistent over time. The

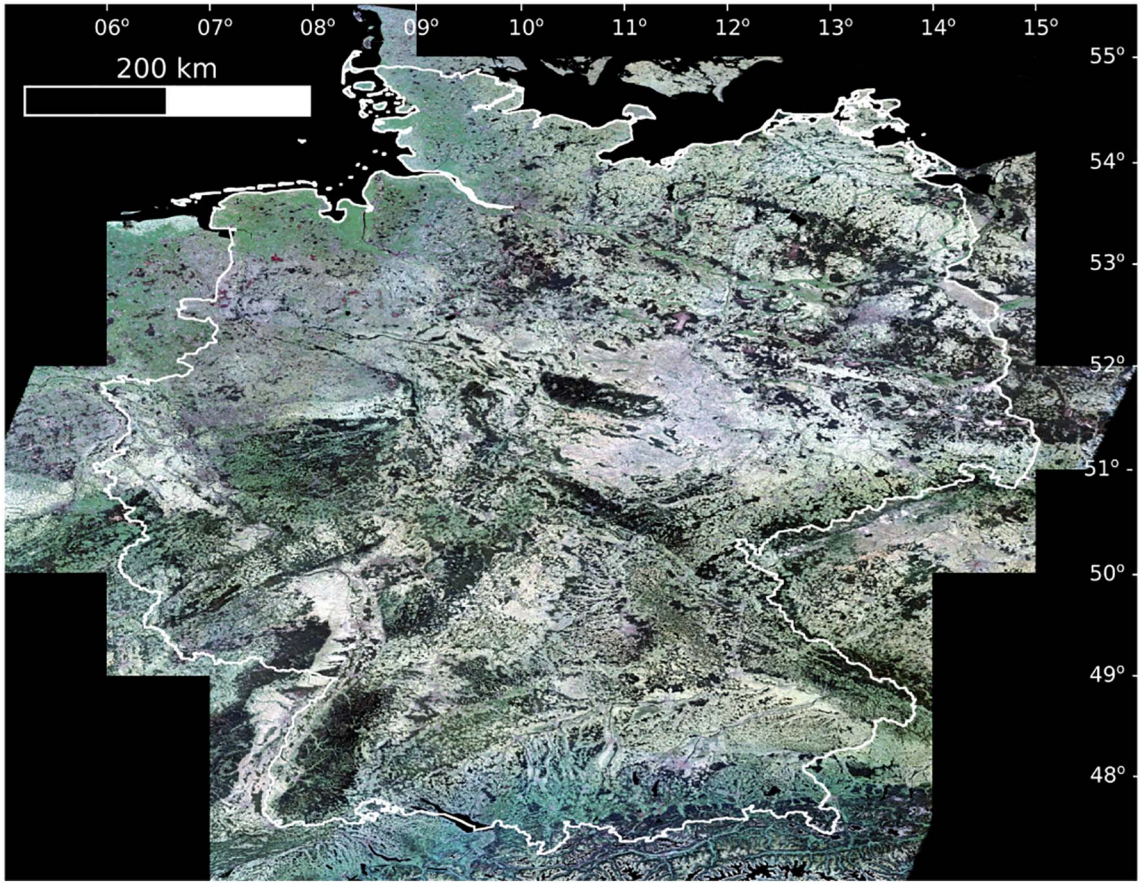

Fig. 5. True color RGB per-pixel average reflectance composite image (Landsat bands 3-2-1) for the study area derived from 2000 to 2004 Landsat imagery (Geographic Lat-Long projection, WGS84). Outline of Germany shown in white. Refer to Fig. 11 for zoom of investigation areas. Scale bar calculated based on 51.5° north. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

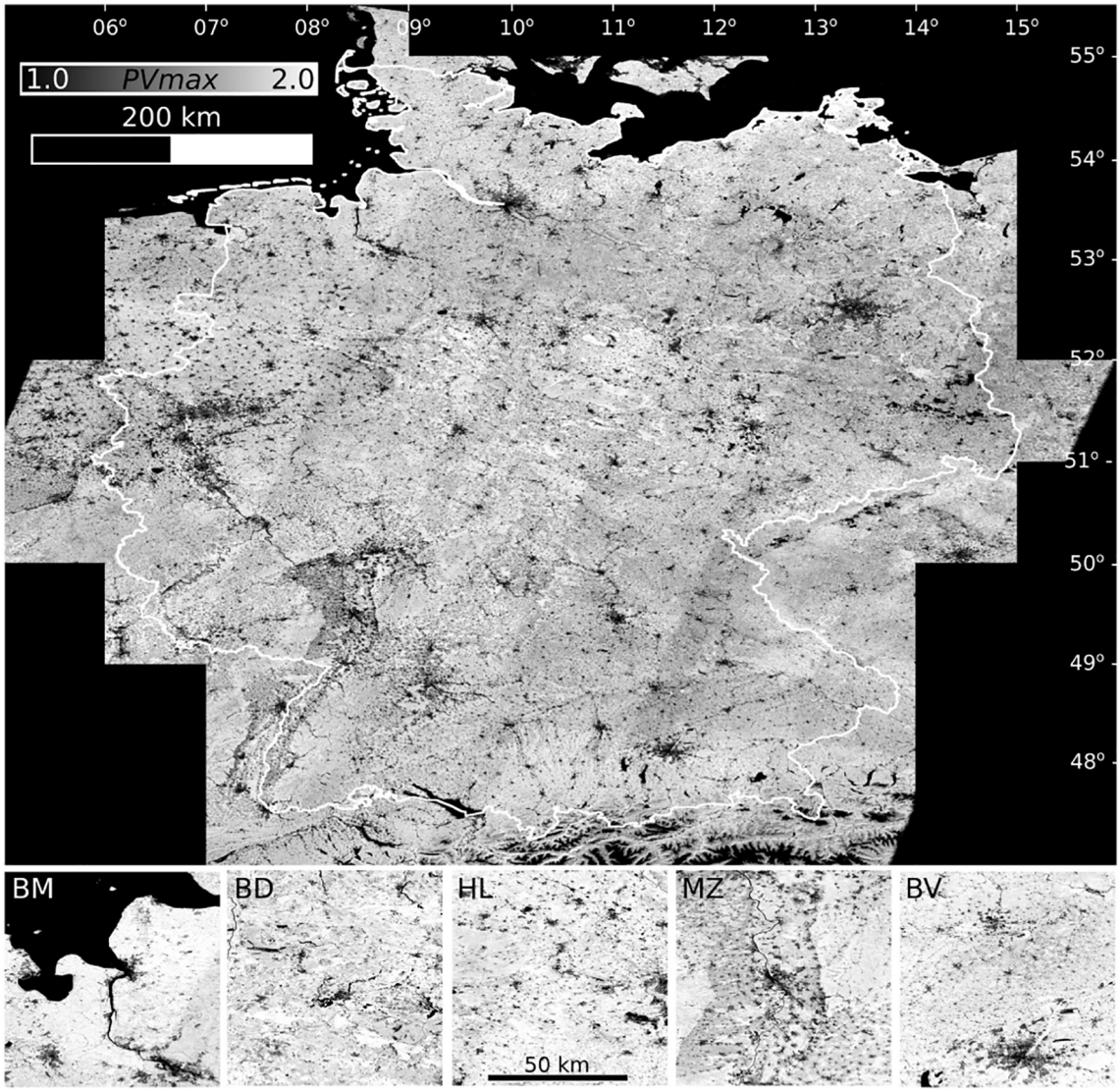

Fig. 6. *PVmax* composite image for the study area and for the five investigation areas derived from 2000 to 2004 Landsat imagery (Geographic Lat-Long projection, WGS84). Outline of Germany shown in white. Scale bar calculated based on 51.5° north.

calculated *Hmax* and *Hmin* show low overall variation with standard deviations of 0.0203 and 0.0243, respectively. On the other hand the distribution of the classes for the five different investigation areas show a higher variability with *Hmax* and *Hmin* standard deviations of 0.0668 and 0.114, respectively. However, for *Hmax* this value is skewed by a much lower Brandenburg value of 1.50 with the other four sites over 1.713. The lower Brandenburg values may be explained by the choice of urban sample pixels containing few pixels that were mixed with vegetation compared to the other four areas. The results shown in Fig. 4 and Table 3 demonstrate that overall the *Hmax* thresholds are most consistent over time, but less so for the different sites. *Hmin* values show higher variability, specifically across the investigation areas. The results demonstrate that regional climate, vegetation and geography/geology differences appear to have a greater impact on the resulting *Hmax* and *Hmin* values compared with time variation. Different agricultural practices (e.g. crop rotation and tilling) may also impact the resulting *Hmax* and *Hmin*, where this change may occur across regions or over time as practices may change. The average *Hmax* and *Hmin* values derived from the five investigation areas are subsequently used for all of Germany for all time periods to generate our *SM*, *Sf*, *Cf*, *SRC*, *SRCmean*, and *SRCnorm* outputs.

Fig. 5 shows a per-pixel cloud-free average reflectance composite (true color) for the full study area. This is a byproduct of the processing chain and is included for reference as it highlights the different geographic regions across the study area. Minimum and maximum reflectance composites are also generated, but not illustrated here. Fig. 6 shows the *PVmax* composite and highlights the large spatial distribution of photosynthetically active vegetation that exists across Germany (bright areas). Dark areas represent the many urban areas, rivers, lakes, high alpine regions and open pit coal mining regions (e.g. eastern Germany). Some BRDF effects are visible. Fig. 7 shows the *PVmin* composite which highlights the distribution of agricultural regions (dark areas) compared with more permanently vegetated regions, specifically coniferous dominated forested regions (bright areas).

5.2. Soil composite results for Germany 1984–2014

The following composite images include the *Sf* (Fig. 8), *Cf* (Fig. 9), the Germany-wide *SRCnorm* (Fig. 10) and temporal changes over three time periods for the five investigation areas (Fig. 11). Each of these results include only *SM* pixels (Eq. (6)) so that grasslands, forests, urban areas and water are excluded. To assess the quality of the *SM* output, the mask was compared to the field class sample pixels derived from the CORINE land cover classification (Table 2) to evaluate the *Hmax* and

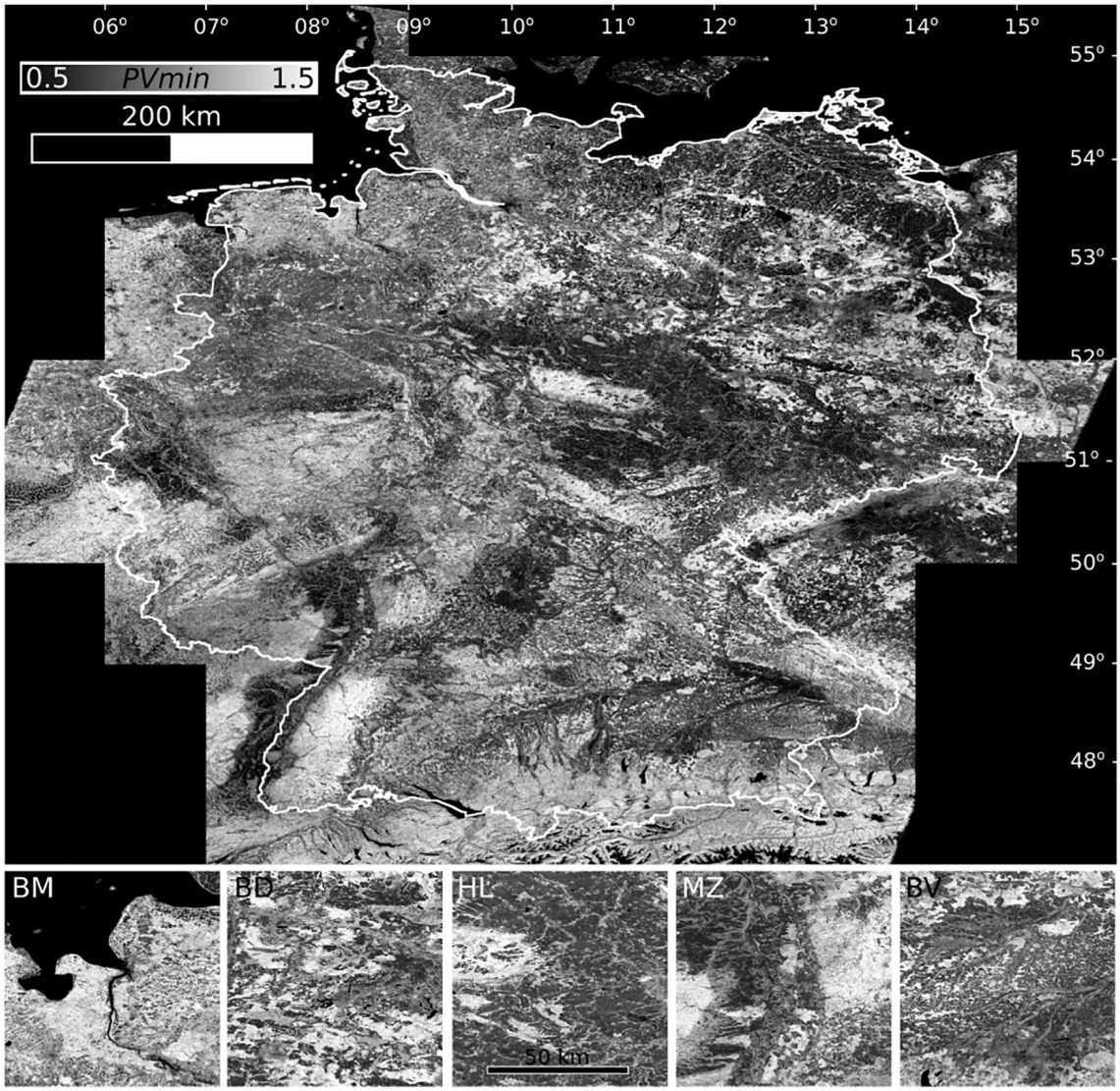

Fig. 7. *PVmin* composite image for the study area and for the five investigation areas derived from 2000 to 2004 Landsat imagery (Geographic Lat-Long projection, WGS84). Outline of Germany shown in white. Scale bar calculated based on 51.5° north.

Hmin thresholds used for both the local and global thresholds (Table 4). Pixels which did not lie within the boundaries of the land cover class "non-irrigated arable land" in 2000 or 2006 were marked as potentially false positives. Differences in the number of soil pixels can be quite variable between using the local versus global thresholds. However, the overall errors are low with the average error for the global threshold at 2.24%, whereas the local threshold is slightly higher at 5.89%. The investigation area Bremen showed much higher error compared with the other areas. This can be explained by observing the scatter plots in Fig. 4, which show Bremen having a much higher local *Hmin*. This can be attributed to limited deciduous forest in the area and, in turn, the degree on NPV that the *Hmin* threshold is defined.

The *Sf* composite highlights the agricultural regions across the study area with the frequency of soil exposure highly variable. A similar distribution is found in the *Cf* composite, although regions with high soil exposure frequency do not necessarily correlate with regions with high relative change. This is an important distinction as it likely reflects different agricultural practices and/or crops. This is clearly shown for the investigation areas that have a high degree of variability in characteristics of each region, such as the spatial extent, field size, distribution, soil exposure frequency and change. For example, the Halle area is known to be one of the most active cropland agricultural regions in Germany. Comparing the *Sf* and *Cf* composites the Halle area experiences both moderate soil exposure and change frequency. However, in the case of the Mainz area, which includes large areas of vineyards, *Sf* is high, but *Cf* is low. In the Bremen area there is both low soil exposure and change, which is consistent with the area having a mixture of cropland and permanent pasture land. Together the *Sf* and *Cf* composites can provide useful information about land use and intensity for individual, and to some extent within fields, for a given region.

The *SRCnorm* composite shown in Fig. 10 highlights spectral shape differences between the red and the two short-wave bands. Other band combinations are left out owing to space restrictions. The color variations shown can be linked to broad differences in soils related to such factors as clay content, iron oxidation, calcium carbonates and soil organic carbon. The various patterns observed in the *SRCnorm* composite for the three Landsat bands can be directly correlated with many of the structural regions shown in the regional geology of Germany (available online; BGR, 2015) and with the soil map of Germany (available online, BGR, 2016). These correlations include large scale regional variations, such as the division of extensive deposition of glacial deposits in the northern part of Germany, the complex metamorphic terrains throughout central Germany and the foreland alpine region in the south. On a more local scale a high level of correlation and

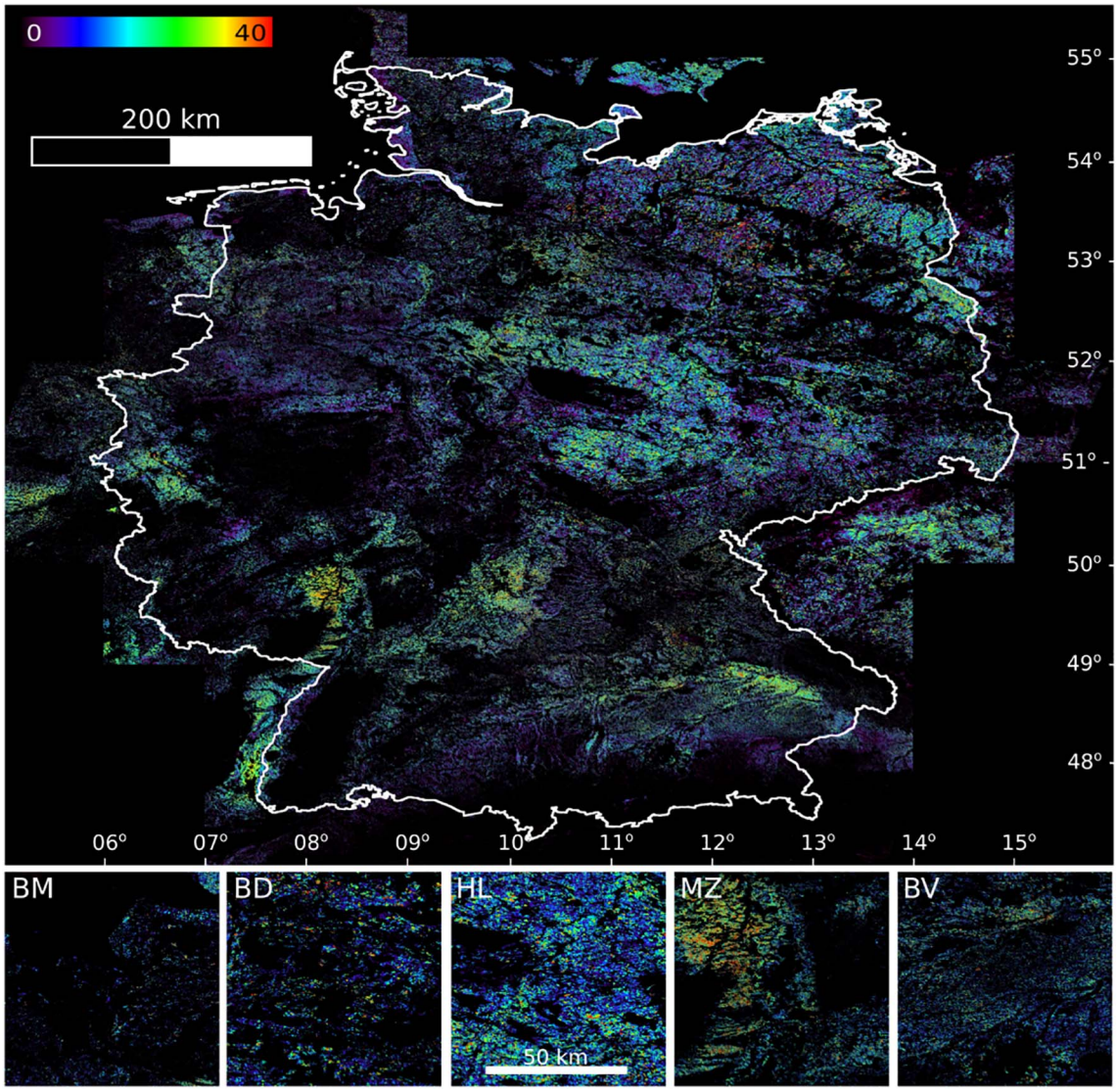

Fig. 8. *Sf* composite image for the study area and for the five investigation areas derived from 2000 to 2004 Landsat imagery (Geographic Lat-Long projection, WGS84). Outline of Germany shown in white. Scale bar calculated based on 51.5° north.

detail can be seen along major river systems (e.g. Rhine valley in the Mainz area), large fluvial deposits (e.g. southern Bavaria area), variation in glacial cover (e.g. moraines in the Brandenburg area), coastal plains along the North Sea (e.g. tidal-flats and marshlands in the Bremen area), and the highly fertile "black" soils that form the Magdeburg Boerde (e.g. Halle area).

The broad structural patterns observed in Fig. 10 for the 2000–2004 period are consistent across all other time periods (not shown owing to space restrictions). This is expected as the underlying geology and primary soil regions of Germany have not changed during the 30 year period analyzed. However, local differences in both color and spatial distribution are visible in the 1 × 1 tile areas. Fig. 11 highlights some examples of temporal variations that can be observed in the *SRCnorm* composites across three time periods. The color histograms for each time period have been matched as closely as possible between each time period to best highlight differences and to minimize differences caused by default histogram stretching. Close examination shows that some areas show an overall brightening or darkening, subtle color change, or the spatial structure itself. This would indicate that there have been some changes to the soil parameters that may be correlated with both natural and/or anthropogenic causes (e.g. soil erosion, degradation). A detailed comparison with the existing geological and soil maps, and identifying temporal changes and their causes is unfortunately beyond the scope of the description of the SCMaP given here. Such detailed analysis is to be assessed in subsequent studies.

6. Discussion

6.1. Pre-processing and image quality

During the pre-processing of data a number of issues have been highlighted related to the Fmask process, atmospheric correction and data filtering impact on the quality and quantity of data available for the SCMaP. The results of the cloud/shadow/snow mask generated by Fmask can be considered of good quality. However, a number of problems have been noted. First is the high number of scenes that are not processed owing to Fmask implementation failures that are under review (refer to Table 1). Second is the occurrence of thin cloud and varying degrees of haze in the imagery that impacts the accuracy of masking accurately cloud boundaries. This is minimized to some extent using the haze removal option available in ATCOR. However, imperfect haze correction and cloud/shadow/snow masking results is some scenes having to be removed manually, which particularly affects the output for the *PVmin* composites where snow and/or clouds are incorrectly

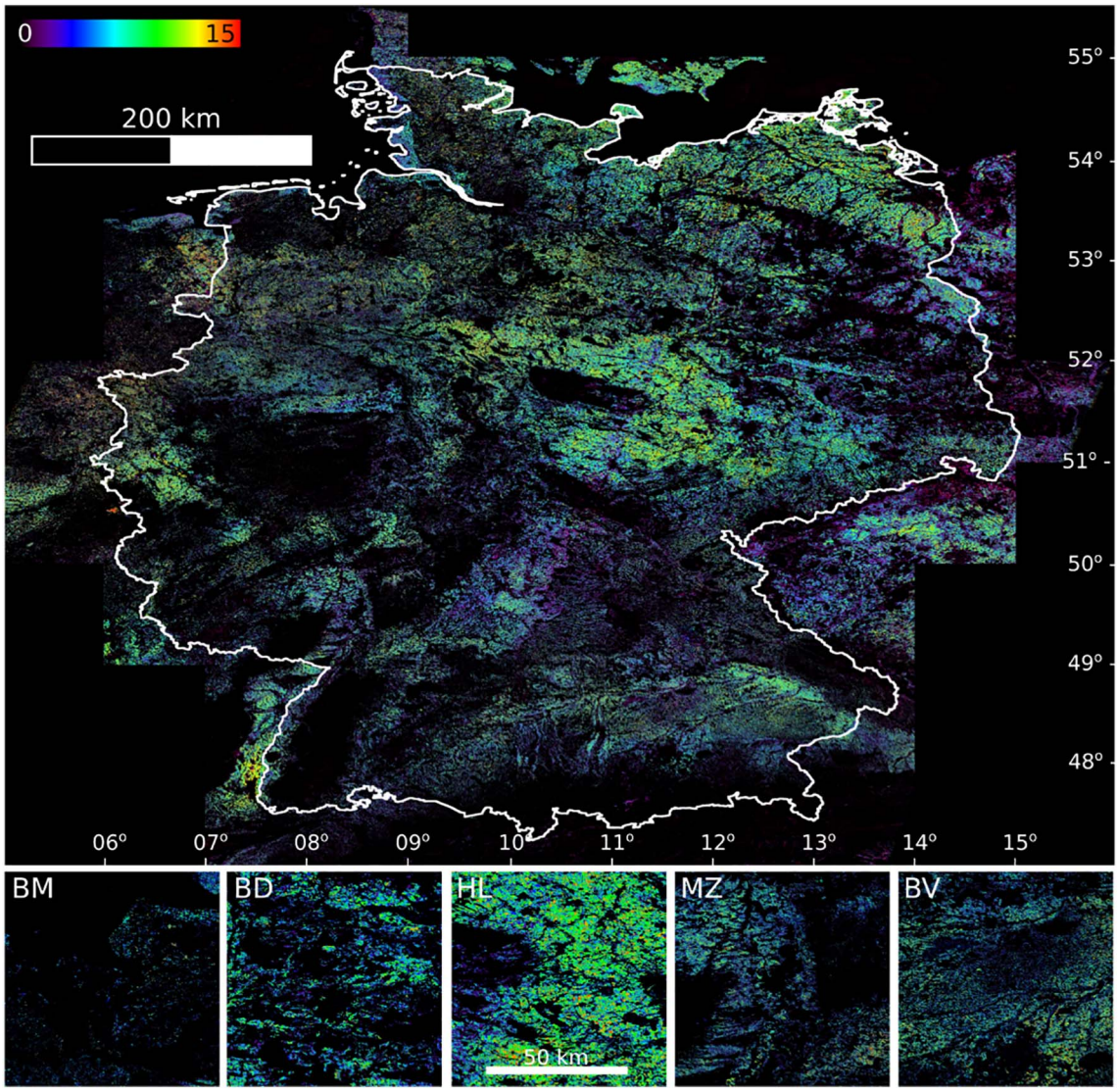

Fig. 9. *Cf* composite image for the study area and for the five investigation areas derived from 2000 to 2004 Landsat imagery (Geographic Lat-Long projection, WGS84). Outline of Germany shown in white. Scale bar calculated based on 51.5° north.

included. For *PVmax* the primary problem is the thin haze that causes increased reflectance in the *BLUE* channel, which becomes visible in the reflectance composite products. The *PV* index used here was able to minimize this problem and provide mostly haze free *PVmax* and *PVmin* reflectance composites.

Data filtering was used to remove images with either missing data or saturation, which was checked per band. However, many scenes have artifacts that cannot be systematically corrected and/or removed. Index maps are generated during the SCMaP processing that highlight the images used in the various output composites. This allows images with artifacts to be removed from the input tile file lists before re-running the process. This requires iterative processing for some tiles to provide the best quality output. Work is ongoing to improve the overall quality of data filtering to provide to best possible output data base available for the SCMaP and other processors being developed.

6.2. SCMaP methodology

The ability of the SCMaP to extract pixels with exposed soil starts first with the *PV* index. This modified NDVI index uses the *BLUE*, *RED* and *NIR* bands of Landsat, which are useful in effectively separating PV from most other materials. On the other hand this index does not provide unique spectral information that can be used to separate soils from all other materials. As such, the SCMaP methodology overcomes this limitation by deriving *Hmax* and *Hmin* thresholds that can be used in combination (intersection) to extract exposed soil pixels. The latter being key to accurately derive the *SM*. The selection of *Hmax* and *Hmin* as shown in Section 5.1 highlights difficulties in determining a set of thresholds that are transferable across both time and space. In this work five different investigation areas were selected that were considered to be representative of the different regions of Germany. The results showed that threshold variability was less for time compared with space, with the key factor being regional changes in the climate, vegetation, and geography/geology. Although the results show that for a specific area locally derived thresholds would be optimal, the results also showed that it is possible to scale up and use average thresholds derived from a subset of the full study area. This was demonstrated for Germany, but would not necessarily apply across larger regions or in areas with highly variable climate, vegetation and geography/geology. As such, using the SCMaP approach it would be necessary to optimize the *Hmax* and *Hmin* for a given region.

In deriving the thresholds for Germany a set of broad classes were chosen with sample locations within each investigation area. Care was taken in making sure the class samples (pixels) were correctly chosen to

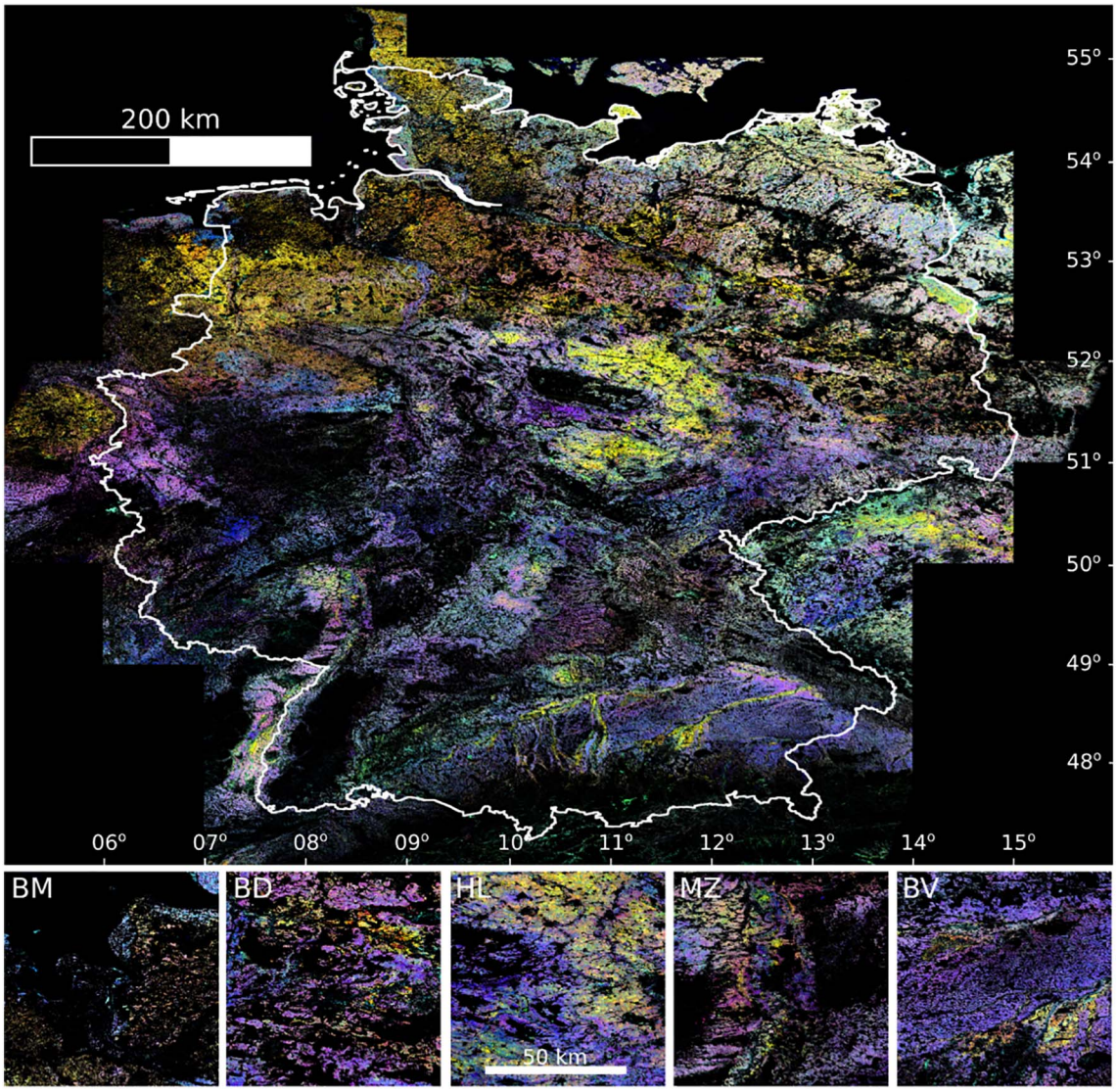

Fig. 10. *SRCnorm* soil composite image (Landsat bands 7-5-3) for the study area and for the five investigation areas derived from 2000 to 2004 Landsat imagery (Geographic Lat-Long projection, WGS84). Outline of Germany shown in white. Scale bar calculated based on 51.5° north.

minimize mixing with other classes, have multiple sample polygons for each class, and that a statistically robust number of samples were selected. As noted above the regional variability of climate, vegetation, and geography/geology was likely the primary cause of differences found in the thresholds for each area. However, the data source for the class selection (CORINE) may also influence the quality of the samples. This was noted in the results where the one outlier for *Hmax* was Brandenburg, which may have been caused by differences in the amount of vegetation within the dense urban area class compared with other areas. This is evident in the scatter plots in Fig. 4 which show four of the five urban area clusters more distributed towards higher *PVmax* values likely indicating a higher degree of mixing with vegetation. This one example highlights the impact on the class sample selection on deriving accuracy and transferable thresholds.

A key issue in deriving the *Hmin* threshold is the spectral similarity between soils and NPV material using Landsat data. *PVmin* values in the Fig. 4 scatter plots show that NPV material within the deciduous and grassland classes have variable overlap with the field class. This overlap is caused by the potential occurrence of NPV material intermixed with exposed soils on agricultural fields with either loose NPV material or stubble retained from the previous growing season. For deciduous forests the overlap is caused by predominantly NPV material mixed with limited exposed soil. For SCMaP, the extraction of exposed soils and the ability to remove false positives (pixels containing some NPV) is considered more important than the loss of some soil pixels. Grasslands showed a similar distribution to deciduous forests, but in some cases also extended into the field class cluster. This was attributed to difficulty in the sampling of permanent grasslands compared to pasture areas that occasionally are tilled. Additional disturbances, such as occasional flooding or grasslands used for periodic military exercises, are not detectable from the CORINE database and single Landsat scenes. As such, the 0.005 percentile of deciduous forests for *PVmin* was used, which can be assume to be constant over the given time periods.

Landsat has limited spectral information, with six broad reflectance bands, which makes the discrimination between many materials difficult. As noted above this is a problem specifically with soils and NPV. In this work the combination (intersection) of *PVmax* and *PVmin* was shown to provide good separation between soil and other materials in the case of Germany. However, other indices that make use of the shortwave infrared bands have the potential to improve the overall discrimination of exposed soils. Additionally, in other parts of the world other indices, or a combination of indices, might be more appropriate. These aspects of the SCMaP are presently being assessed.

The SCMaP implemented in this study made use of 5-year periods to

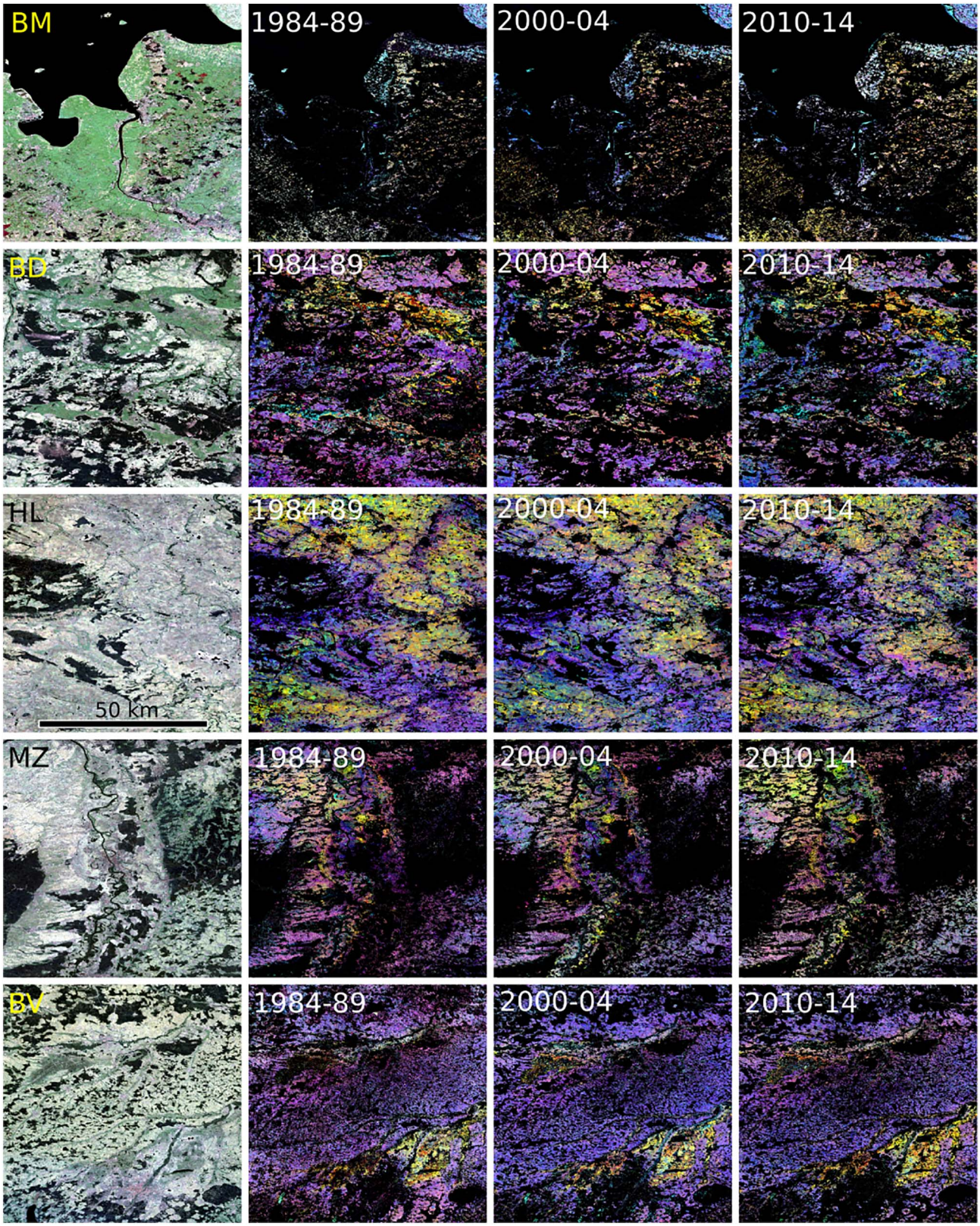

Fig. 11. *SRCnorm* soil composite image (Landsat bands 7-5-3) for the five investigation areas over three time periods showing subtle variations in localized regions that may be linked to changing soil parameters. Included in the far left column are the true color RGB images (Landsat bands 3-2-1) for reference (Geographic Lat-Long projection, WGS84). Scale bar calculated based on 51.5° north. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
Evaluation of *SM* results for all investigation tiles compared with CORINE database.

Tile thresholds	*SM* pixels	False positives	Error (%)
Bavaria global	3,259,796	133,735	4.10
Bavaria local	2,644,198	81,814	3.09
Brandenburg global	2,771,728	47,441	1.71
Brandenburg local	4,545,289	346,533	7.62
Bremen global	486,928	53,190	10.92
Bremen local	1,869,578	471,850	25.24
Halle global	7,140,623	34,446	0.48
Halle local	7,465,490	81,392	1.09
Mainz global	2,239,184	87,007	3.89
Mainz local	2,832,229	158,394	5.59
Total global	15,898,259	355,819	2.24
Total local	19,356,784	1,139,983	5.89

derive the various composite products. The choice of five years was based primarily on providing enough scenes to extract exposed soil pixels with high spatial coverage. Preliminary tests of the processor showed that spatial coverage increased significantly over one to three years and leveled off after five years. Additionally, five years also provided a sufficient number of scenes for statistical analysis. For this work a roughly even distribution of scenes across a given year was assumed. However, the distribution of scenes with respect to the season can be variable for different parts of Germany, which may impact the number of time soils are exposed. This is being assessed in more detail in subsequent work.

Another factor to consider when choosing the time period is the potential changes within soils. It can be assumed that the underlying geology will be consistent across the 30 year range of data, so that any changes in the soil within or across the 5-year periods will be caused by external factors that can be related to 1) small scale variability in moisture or agricultural practices (e.g. spreading of organic fertilizer, tilling); 2) land cover/land use changes (e.g. pasture to cropland); or, 3) soil degradation and/or soil erosion (or vice versa) caused by natural (e.g. flooding, long term large scale climatic changes) or anthropogenic changes (e.g. increasing/decreasing production). With a 5-year period the first factor (small scale changes) will be averaged resulting in the SRC reflecting soil properties related to baseline mineralogy and soil organic content. Thus, changes observed across the 5-year periods can be assumed to be related primarily to the second and third factors, which may provide useful information relating to key environmental functions that directly impact food, fibre and timber production, water storage and redistribution, pollutant filtering and carbon storage.

6.3. Conclusions and future work

The SCMaP was implemented with Landsat 4, 5, and 7 data for the entire area of Germany over 30 years divided into 5 year periods. The results have demonstrated the ability of the method to provide 1) useful value-added information on the spatial distribution and extent of exposed soils; 2) statistical information related to agricultural use and intensity (e.g. depending on cultivation cycle and pattern); and, 3) provide soil composite reflectance images of high spatial coverage that correlate well with existing soil and geological maps. The results provide a preliminary glimpse of the potential of the automated SCMaP products that can be used for numerous tasks related to the use and mapping of soils. Such composite products have not yet been produced at the spatial and temporal scale for a temperate region where soil exposure is periodic.

The practicality of using the limited spectral information available with Landsat for soils is known. However, the SCMaP results do show that broad variations in the spectra are visible and correlate well with both the soil and the underlying geological structural regions of Germany. As such, subsequent work will focus on linking the soil composites to existing soil information (maps and cores) to develop higher order products that can assist in mapping specific soil parameters. These next steps in the development of the SCMaP will focus on linking the SRC spectral information with existing information on soil parameters (e.g. texture, mineralogy and soil organic content). The first task will be to determine which parameters can be correlated with existing information and subsequently be used to enhance and/or improve the spatial and temporal soil information detail presently available. This includes potentially using the spatial continuous composites to improve spatial detail in existing maps or the merging of soil maps across state or country boundaries where internal surveys may differ. Next will be an assessment of the observed temporal variations related to the soil parameters and how they may link to important issues, such as land cover/land use changes and soil degradation. The latter can make more detailed use of the statistical information extracted (*Sf* and *Cf*) from the per-pixel data stacks where this information can be linked with productivity across a given region, local area or even within individual fields.

The SCMaP approach was demonstrated on Germany over a 30 year period, but it is applicable to other regions of the globe that also have limited soil exposure owing to vegetative cover. In particular, these are regions with extensive agricultural activities where soils become exposed periodically throughout the year, or over many years. As many of these areas include intensive agriculture, continued monitoring of the underlying soils is crucial to assess soil degradation and to achieve sustainable productivity and ultimately food security. Key to the SCMaP approach is the ability to define *PVmax* and *PVmin* thresholds that are applicable to a specific region. This was demonstrated for an area the size of Germany, where locally defined thresholds can vary from the regional threshold. Applying the SCMaP to other regions will require specific thresholds for the given region that take into account the characteristics of the regional climate, vegetation, and geography/geology. Additionally, the user must also consider different agricultural practices, such as whether or not tilling is used, which has a direct impact on the degree of NPV exposed at surface. As data from sensors such as Landsat and Sentinel-2 have global spatial and multi-temporal coverage and is freely available to the public, the SCMaP is particularly well suited to extract exposed soils for numerous regions around the globe.

Moving forwards with other sensors, such as Landsat 8 or Sentinel-2, will provide additional opportunities to improve products developed from the SCMaP. Sentinel-2A/B has an increased acquisition rate (e.g. 3–5 day repeat orbit) over the full year with increased spectral information (10 visible and near-infrared broad band sensors) at spatial resolutions of 10 and 20 m (four bands at 10 m and six bands at 20 m from 490 to 2190 nm). The increase in acquisition rate compared to Landsat 4, 5 and 7 will allow for a higher number of cloud-free scenes and for implementation of the SCMaP over shorter time periods (e.g. two to three years). The increase in the number of spectral bands and higher spatial resolution will also allow for a higher level of detail provided in the SCMaP products. Unfortunately, Sentinal-2A/B will still have limited information in the shortwave infrared that can be used to detect spectral features related to different mineral groups or distinguish between soils and NPV. Improving the spectral analysis of the soils can be accomplished using satellite hyperspectral data, such as the EnMAP sensor to be launched in 2019. The EnMAP sensor has a similar spatial resolution of 30 m, but contiguous spectral information from 420 to 2450 nm compared with Sentinel-2A/B. However, it has lower spatial coverage and acquisition rates which do not make it as applicable for the SCMaP. In combination with Sentinel 2A/B and/or Landsat 8, EnMAP data could be used to extract a higher level of soil information for a sample of the exposed soil in a given area, which can subsequently be used to extrapolate across the more continuous SCMaP soil composites.

Acknowledgements

The authors would like to thank the Bavarian Ministry of Economic Affairs and Media, Energy and Technology (FKZ:0703/89373/18/2013) for funding the OPUS-GMES project (Concept of an Operational Platform for the Provision and Processing of Sentinel Data in Support of Copernicus Geo-Information Services) which provided the basis for the technical developments of the TimeScan framework. The further developments of the SCMaP have been carried out within the AGRO-DE project that is funded by the Federal Agency for Agriculture and Food (BLE), FKZ: 2815704815.

References

Abrams, M.J., Hook, S.J., 2013. NASA's hyperspectral infrared imager (HyspIRI). In: Kuenzer, C., Dech, S. (Eds.), Thermal Infrared Remote Sensing. Springer, New York, NY, USA, pp. 117–130.

Apache Foundation, 2017a. Apache Hadoop, Version 2.7.3. Online: http://hadoop.apache.org, Accessed date: 23 August 2017.

Apache Foundation, 2017b. Apache Mesos, Version 1.2.1. Online: http://mesos.apache.org/, Accessed date: 23 August 2017.

Asner, G.P., Heidebrecht, K.B., 2002. Spectral unmixing of vegetation, soil and dry carbon in arid regions: comparing multi-spectral and hyperspectral observations. Int. J. Remote Sens. 23, 400–410.

Asner, G.P., Wessman, C.A., Archer, S., 1998. Scale dependence of absorption of photosynthetically active radiation in terrestrial ecosystems. Ecol. Appl. 8, 1003–1021.

Barnes, E.M., Sudduth, K.A., Hummel, J.W., Lesch, S.M., Corwin, D.L., Yang, C., Daughtry, C.S.T., Bausch, W.C., 2003. Remote- and ground-based sensor techniques to map soil properties. Photogramm. Eng. Remote. Sens. 69 (6), 619–630.

Ben-Dor, E., Chabrillat, S., Dematte, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using imaging spectroscopy for soil properties. Remote Sens. Environ. 113, S38–S55.

BGR, (2015). Bodenübersichtskarte 1:200000 (BÜK200). (Online: http://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/BUEK200/buek200_node.html (accessed April 28, 2017)).

BGR, (2016). Geologische Karte der Bundesrepublik Deutschland 1:1 000 000 (GK1000), (Online: https://www.bgr.bund.de/DE/Themen/Sammlungen-Grundlagen/GG_geol_Info/Karten/Deutschland/GK1000/gk1000_inhalt.html (accessed April 28, 2017)).

Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saunders, A.M., Stum, A.K., 2008. Landsat spectral data for digital soil mapping. In: Hartemink, A.E., McBratney, A.B., Mendonca-Santos, M.L. (Eds.), Digital Soil Mapping With Limited Data. Springer Science, Australia, pp. 193–203.

Bussoletti, E., 2012. Space observations for agriculture and food support. In: Paper Presented at 32nd Session of Inter-Agency Meeting on Outer Space Activities, Rome, March 7–9.

Carrere, V., Briottet, X., Jacquemoud, S., Marion, R., Bourguignon, A., Chami, M., Dumont, M., Minghelli-Roman, A., Weber, C., Lefevre-Fonollosa, M., Mandea, M., 2013. HYPXIM: a second generation high spatial resolution hyperspectral satellite for dual applications. In: Paper Presented at 5th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Gainsville, FL, June 25–28.

Demattê, J.A.M., Huete, A., Guimarães, L., Nanni, M.R., Alves, M.C., Fiorio, P.R., 2009. Methodology for bare soil detection and discrimination by Landsat-TM image. Open Rem Sensing J. 2, 24–35.

Demattê, J.A.M., Alves, M.R., Terra, F.S., Bosquilia, R.W.D., Fongaro, C.T., Barros, P.P.S., 2016. Is it possible to classify topsoil texture using a sensor located 800 km away from the surface? Rev. Bras. Ciênc. Solo 40, 1–13.

DLR, 2011. OPUS-GMES. Online: http://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-9978/17061_read-41279, Accessed date: 28 April 2017.

Eckelmann, W., 2004. Soil information for Germany: the 2004 position. In: European Soil Bureau - Research Report, No. 9, pp. 147–157.

EEA, 2007. CLC 2006 technical guidelines. In: EEA Technical Report, No 17/2007, Copenhagen, Online: http://www.eea.europa.eu/publications/technical_report_2007, Accessed date: 28 April 2017.

Elvidge, C.D., 1990. Visible and near infrared reflectance characteristics of dry plant materials. Int. J. Remote Sens. 11, 1775–1795.

ESA, 2015. News, Sentinel-2A Products Available in the Data Hub. Online: https://sentinel.esa.int/web/sentinel/news/-/asset_publisher/xR9e/content/sentinel-2a-products-available-in-the-data-hub, Accessed date: 28 April 2017.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., van der Linden, S., Leitão, P.J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The Environmental Mapping and Analysis Program (EnMAP) spaceborne imaging spectroscopy mission for earth observation. Remote Sens. 7 (7), 8830–8857.

Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ. 122, 66–74.

Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T.R., 2011. Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. Remote Sensing Letters 2, 279–288.

Hengl, T., 2006. Finding the right pixel size. Comput. Geosci. 32, 1283–1298.

Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., 2015. An integrated Landsat time series protocol for change detection and generation of annual gap-free surface reflectance composites. Remote Sens. Environ. 158, 220–234.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25 (3), 295–309.

IT4Innovation, 2017. Apache Mesos, Version 1.2.1. Online: https://docs.it4i.cz/anselm/introduction/, Accessed date: 23 August 2017.

LfU Bayern, 2015. Standortkundliche Bodenkarten. Online: http://www.lfu.bayern.de/boden/boden_daten/sbk/index.htm, Accessed date: 28 April 2017.

Lopinto, E., Ananasso, C., 2013. The Prisma hyperspectral mission. In: Proceedings of EARSeL Symposium, Matera, Italy. June 3–6.

Malley, D.F., Martin, P., Ben-Dor, E., 2004. Application in analysis of soils. Chapter 26. In: Craig, R., Windham, R., Workman, J. (Eds.), Near Infrared Spectroscopy in Agriculture, A Three Societies Monograph (ASA, SSSA, CSSA). vol. 44. pp. 729–784.

Marconcini, M., Üreyen, S., Esch, T., Metz, A., Zeidler, J., 14-16 March 2017. Towards a new baseline layer for global land-cover classification derived from multitemporal satellite optical imagery. In: WorldCover 2017. ESA, Frascati, Italy.

Matsunaga, T., Iwasaki, A., Tsuchida, S., Tanii, J., Kashimura, O., Nakamura, R., Yamamot, H., Tachikawa, T., Rokugawa, S., 2014. Current Status of Hyperspectral Imager Suite (HISUI). In: IGARSS, July 13–18, 2014, Québec, Canada.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayrc, T.R., 2011. The use of remote sensing in soil and terrain mapping — a review. Geoderma 162 (1–2), 1–19.

Nachtergaele, F., van Velthuizen, H., Verelst, L., Batjes, N., Dijkshoorn, K., van Engelen, V., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., Wiberg, D., Shi, X., 2009. Harmonized World Soil Database – Version 1.1, FAO.

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.

Okin, G.S., 2007. Relative spectral mixture analysis—a multitemporal index of total vegetation cover. Remote Sens. Environ. 106, 467–479 2007.

Omuto, C., Nachtergaele, F., Rojas, R., 2013. State of the Art Report on Global and Regional Soil Information: Where Are We? Where to Go? pp. 81. Online: http://library.wur.nl/isric/fulltext/isricu_i33761_001.pdf, Accessed date: 28 April 2017.

Panagos, P., Van Liedekerke, M., Jones, A., Montanarella, L., 2012. European Soil Data Centre: response to European policy support and public data requirements. Land Use Policy 29, 329–338.

Plant, J.A., Whittaker, A., Demetriades, A., De Vivo, B., Lexa, J., 2003. The geological and tectonic framework of Europe. In: Salminen, R. (Ed.), Geochemical Atlas of Europe, Part 1: Background Information, Methodology and Maps. Geological Survey of Finland, Espoo, Finland.

Potapov, P., Turubanova, S., Hansen, M.C., 2011. Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. Remote Sens. Environ. 115 (2), 548–561.

Richter, R., 2010. Atmospheric/topographical correction for airborne imagery. In: ATCOR4 User Guide. Wessling, Germany.

Richter, R., Schläpfer, D., Müller, A., 2006. An automatic atmospheric correction algorithm for visible/NIR imagery. Int. J. Remote Sens. 27, 2077–2085.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, Y., Hansen, M., Loveland, T.R., Vermote, E., Zhang, C., 2010. Web-enabled Landsat Data (WELD): Landsat ETM + composited mosaics of the conterminous United States. Remote Sens. Environ. 114 (2010), 35–49.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., de Lourdes Mendonça-Santos, M., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G., 2009. Digital soil map of the world. Science 325, 680–681.

Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Sci. Soc. Am. J. 66 (3), 988–998.

Statistische Ämter der Länder, 2015. Arbeitskreis Umweltökonomische Gesamtrechnungen der Länder – Flächenversiegelung 2000–2014 nach Bundesländern. Online: http://www.ugrdl.de/tab54.htm, Accessed date: 28 April 2017.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8 (2), 127–150.

Tucker, C.J., Grant, D.M., Dykstra, J.D., 2004. NASA's global orthorectified Landsat data set. Photogramm. Eng. Remote. Sens. 70, 313–322.

Turner, W., Rondinini, C., Pettorelli, N., Mora, B., Leidner, A.K., Szantoi, Z., Buchanan, G., Dech, S., Dwyer, J., Herold, M., Koh, L.P., Leimgruber, P., Taubenboeck, H., Wegmann, M., Wikelski, M., Woodcock, C., 2015. Free and open-access satellite data are key to biodiversity conservation. Biol. Conserv. 182, 173–176.

Valero, S., Morin, D., Inglada, J., Sepulcre, G., Arias, M., Hagolle, O., Dedieu, G., Bontemps, S., Defourny, P., Koetz, B., 2016. Production of a dynamic cropland mask by processing remote sensing image series at high temporal and spatial resolutions. Remote Sens. 8 (1), 55.

White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A., Guindon, L., 2014. Pixel-based image compositing for large-area dense time series applications and science. Can. J. Remote. Sens. 40, 192–212.

Wiesmeier, M., Barthold, F., Spörlein, P., Geuß, U., Hangen, E., Reischl, A., Schilling, B., Angst, G., von Lützow, M., Kögel-Knabner, I., 2014. Estimation of total organic carbon storage and its driving factors in soils of Bavaria (southeast Germany). Geoderma Regional 1, 67–78.

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W.B., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R., 2008. Free

access to Landsat imagery. Science 320, 1011.
Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., Woodcock, C.E., 2012. Opening the archive: how free data has enabled the science and monitoring promise of Landsat. Remote Sens. Environ. 122, 2–10.
Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. Remote Sens. Environ. 144, 42–64.
Zhang, Q., Xiao, X., Braswell, B., Linder, E., Baret, F., Moore, B.I.I.I., 2005. Estimating light absorption by chlorophyll, leaf and canopy in a deciduous broadleaf forest using MODIS data and a radiative transfer model. Remote Sens. Environ. 99, 357–371.
Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118 (15), 83–94.
Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171.
Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sens. Environ. 159, 269–277.