

# An Operational Radiometric Landsat Preprocessing Framework for Large-Area Time Series Applications

David Frantz<sup>1\*</sup>, Achim Röder<sup>1</sup>, Marion Stellmes<sup>1</sup> & Joachim Hill<sup>1</sup>

Affiliations:

<sup>1</sup>Environmental Remote Sensing & Geoinformatics, Faculty of Spatial and Environmental Sciences, Trier University, 54296 Trier, Germany.

\*Corresponding Author:

David Frantz, e-Mail: [frantz@uni-trier.de](mailto:frantz@uni-trier.de)

© 2016 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Reference:

D. Frantz; A. Röder; M. Stellmes; J. Hill, "An Operational Radiometric Landsat Preprocessing Framework for Large-Area Time Series Applications," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp.1-16.

DOI: 10.1109/TGRS.2016.2530856

Online available at: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7426833>

The PDF document is a copy of the final version of the accepted manuscript. The paper has been through peer review, but it has not been subject to copy-editing, proofreading and formatting added by the publisher (so it will look different from the final version of record, which may be accessed following the DOI above depending on your access situation).

**Abstract**—We developed a large area pre-processing framework for multi-sensor Landsat data, capable of processing large data volumes. Cloud and cloud shadow detection is performed by a modified Fmask code. Surface reflectance is inferred from Tanré’s formulation of the radiative transfer, including adjacency effect correction. A pre-compiled MODIS water vapor database provides daily or climatological fallback estimates. Aerosol optical depth (AOD) is estimated over dark objects that are identified in a combined database and image-based approach, where information on their temporal persistency is utilized. AOD is inferred with consideration of the actual target reflectance and background contamination effect. In case of absent dark objects in bright scenes, a fallback approach with a modelled AOD climatology is used instead. Topographic normalization is performed by a modified C-correction. The data are projected into a single coordinate system and are organized in a gridded data structure for simplified pixel-based access. We based the assessment of the produced dataset on an exhaustive analysis of overlapping pixels: 98.8% of the redundant overlaps are in the range of the expected  $\pm 2.5\%$  overall radiometric algorithm accuracy. AOD is in very good agreement with AERONET sunphotometer data ( $R^2$ : 0.72 to 0.79, low intercepts and slopes near unity). The uncertainty in using the water vapor fallback climatology is approximately  $\pm 2.8\%$  for the TM SWIR1 band in the wet season. The topographic correction was considered successful by an investigation of the non-relationship between the illumination angle and the corrected radiance.

**Index Terms**—Atmospheric correction, cloud detection, Landsat, large area, multitemporal, southern Africa, surface reflectance, topographic correction.

## I. Introduction

LANDSAT data are one of the most valuable resources for earth observation [1] due to their long term data continuity [2] and their optimal resolution to monitor processes at the landscape level [3]. With the advent of open data policy [4], accompanied by technical progress in terms of processing, storing and transmission infrastructure and the increasing availability of automated processing routines (e.g. [5]), the usage of Landsat data changed fundamentally [6]. Historically, analyses were either based on a few images and large areas [7] or on small areas and shorter time steps [8]. Now, it has become feasible to make use of the full depth of the Landsat archive, as well as covering very large areas at the same time. Nevertheless, Landsat data are still provided in the traditional Worldwide Reference System 2 (WRS-2) framework [9], where the image footprints vary from acquisition to acquisition. This involves several obstacles for the end user, e.g. the integrated usage of data from different paths requires the reprojection to a unique coordinate system, which simplifies the adequate usage of the data-rich orbital overlap area. Even in the case of using only one WRS-2 footprint, all images have to be cropped to the same extent. This is even mandatory if sophisticated follow up applications like time series analyses (TSA) or the derivation of pixel based composites (PBC) [10] are to be addressed. The usage of these dense time-series applications requires the data of a given location to be easily and quickly accessible, regardless of their initial path and row designation, projection, acquisition time or sensor. Therefore, pre-processed imagery in a gridded data structure represents a more suitable structure for TSA or PBC applications [9], as demonstrated by the Web-enabled Landsat Data (WELD) project [11].

Existing large area production systems [9] or similar pre-processing architectures [11] most often do not include a full and integrated radiometric treatment, i.e. accounting for atmospheric and topographic effects simultaneously. Topographic variation even has a greater impact on the remotely sensed data than atmospheric effects [12], thus topography should be accounted for if the area of interest is not merely flat. More sophisticated analyses of large amounts of data like spectral unmixing or the quantitative derivation of biomass indicators (e.g. in support for monitoring systems in a Reducing Emissions from Deforestation and forest Degradation (REDD) in developing countries context) for several or many time steps also requires more sophisticated radiometric corrections, i.e. surface reflectance products. [9] recently published an

overview of several Landsat data processing systems for large area monitoring, whereby the corrections range from Top-of-Atmosphere (TOA) reflectance to surface reflectance, some of them including topographic and/or directional effects. Most systems only correct the data to TOA reflectance [9]. The widely-used Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) [13] produces surface reflectance by utilizing radiative transfer modelling, though the system does not account for topography. Of note is the Eastern Australia pre-processing framework [14] that minimizes atmospheric, topographic and bi-directional effects, though their integrated BRDF correction might not allow the direct transfer to areas that suffer from low data availability.

Large area generation systems that apply a terrain normalization and provide gridded surface reflectance products of multi-sensor Landsat data are still scarce or require input data that may not be available in each part of the world. As such, we chose a sufficiently elaborate method set that minimizes the amount of input data, in order to provide a processing framework that may even be applied in areas where the general data availability is still low and where specific environmental settings preclude the usage of specific processing strategies, e.g. in our study area in southern Africa. We here present an operational approach that processes all available multi-sensor Level 1T (L1T) Landsat Digital Number (DN) data to surface reflectance and stores the processed data in a gridded tile structure as known from the MODIS land products. The generated products are tailored for applications that require rapid and easy data access, make use of a large amount of data across space and time and demand radiometrically normalized data. The processing scheme includes modules for cloud masking, atmospheric and topographic correction, reprojection and gridding.

## II. Study Area

The method was developed in southern Africa, entirely including the countries of Angola, Zambia, Zimbabwe, Botswana and Namibia (ca. 3.7 Mio. km<sup>2</sup>). The area was designated because nation-wide and cross-national wall-to-wall mapping of forest and ecosystem related parameters are to be targeted in the future, thus the inclusion of whole countries. For example, any REDD+ assessment should monitor deforestation and forest degradation rates at the national level to avoid leakage to unobserved spots [15]. In addition, the countries themselves were chosen because their national territory partially falls into the Kavango-Zambezi Transfrontier Conservation Area (KAZA TFCA), which is planned to be centered on the Caprivi-Chobe-Victoria Falls area. Thus, the area is a highly interesting spot for any kind of cross-boundary studies regarding a wide variety of ecological and social questions. The area is climatically diverse, especially owing to the latitudinal precipitation gradient, mainly as a consequence of the Intertropical Convergence Zone (ITCZ) [16]. The vegetation cover ranges from dense Miombo forests in Angola to sparse xerilic savanna ecosystems in the Kalahari, and also includes more extreme surface types like swamps, salt pans and deserts [17]. The seasonal cycle is closely tied to seasonal changes in large scale air movement and solar configuration, which results in three hygrothermal seasons that spatially and annually differ in timing and length due to the variability in precipitation [18]: (i) hot wet season (~ November to April), (ii) cool dry season (~ May to August) and (iii) hot dry season (~ September to October). The study area is displayed in Fig. 1. Nonetheless, the presented pre-processing scheme is not bound to this area and can be ported elsewhere.

Please place Fig. 1 approximately here.

### III. Data

#### A. Landsat data

All available intersecting Landsat images from 194 WRS-2 frames were acquired from the U.S. Geological Survey archive (USGS). We downloaded even very cloudy images of up to 70 % cloud coverage because the Automated Cloud Cover Assessment (ACCA) system occasionally fails to estimate the cloud coverage with sufficient precision [5]. We discarded images that were not corrected with the Level 1 Product Generation System (LPGS) to L1T precision, as a reliable co-registration among images was considered to be of major importance for TSA and PBC follow-up applications. At the time of writing, we acquired and processed 58,731 L1T images with a total data volume of nearly 15 TB (see table I for details).

Fig. 1 displays the total number of available images for each WRS-2 frame. The data availability is quite low when compared to other areas such as the United States. There is a pronounced North-South gradient where data availability is extremely low in northern Angola due to the characteristic regional rainfall patterns where the north-western part of the study area is already located in the tropics.

Fig. 2 displays all available Landsat images for each year (top) and each month (bottom). The data availability is unevenly distributed among the years, e.g. due to satellite (de-) commissioning, changes in acquisition strategies, technical failure, climatic reasons and so forth. There is a seasonality in data availability with more data during the dry season and less data during the wet season.

The Landsat Global Archive Consolidation effort [19] is currently still in progress, and we might expect up to 200 additional Landsat images per frame [20] which will significantly foster the applicability of subsequent analyses approaches— provided that each image is of sufficient quality to be processed to L1T precision.

Please place table I approximately here.

Please place Fig. 2 approximately here.

#### B. Digital Elevation Model

Recently, the USGS publicly released the 1-arc-Second (~30 m) Digital Elevation Model (DEM) derived by the Shuttle Radar Topography Mission (SRTM) [21]. We pre-compiled an elevation mosaic generously covering all of sub-equatorial Africa. DEM data are used for the topographic correction and for applying an elevation correction to the optical depths.

#### C. Precipitable Water Vapor

MODIS water vapor data are used for the physically based correction of water vapor absorption in the earth's atmosphere. We use the MOD05 and MYD05 products, as well as the MOD03 and MYD03 geolocation tables, which are automatically downloaded from the Level 1 and Atmosphere Archive and Distribution System (LAADS) at NASA's Goddard Space Flight Center. We use data that were derived from the near infrared water vapor algorithm [22], which relies on water vapor attenuation of the NIR radiation. The column water vapor amount is determined from radiative transfer theory on basis of ratios between water vapor absorbing and atmospheric window bands. The product is obtained at 1 km spatial resolution and the temporal resolution is up to one day if considering both Aqua and Terra observations.

### IV. Methods

The presented framework for processing Landsat data from L1 DN values to a gridded surface reflectance product is schematically shown in Fig. 3. The download of Landsat and SRTM data, as well as the mosaicking of the DEM are performed rather manually in advance. The core functionality of the framework is enclosed in the solid box, where the main modules are the identification of clouds and cloud shadows (B), the radiometric processing (C) and geometric modules for the finishing of the data (D). The radiometric correction combines methods for atmospheric (C.2) and topographic corrections (C.3), which rely on the

computation of the angular scene parameters, the derivation of topographic information and an elaborated scheme for retrieving AOD in a joint database-, image- and object-based approach (C.2-b). The gaseous transmission is accounted for by a MODIS derived water vapor database, that is scheduled operationally (C.2-a). Table II defines the mathematical symbols used in this paper.

#### A. Implementation

The code is able to process Landsat images from the Thematic Mapper (TM), Enhanced Thematic Mapper (ETM+) and Operational Land Imager (OLI) sensors. For the sake of dataset consistency we opted for processing only the traditional six reflective bands. Thus we omit the ultra blue and cirrus bands of the OLI sensor. We also do not make use of the Landsat-8 quality layer for the same reason as well as not using the cirrus band in Fmask, as opposed to [23].

Detailed processing information of the key modules are appended to the metadata in order to enable the user to gauge the quality of the processed images.

The framework is entirely implemented with Open Source software and is written in C. The Geospatial Data Abstraction Library [24] API is used for reading data in GeoTiff and HDF format, as well as for reprojection purposes. The cURL API is used for the automatic download of MODIS data. The GNU Scientific Library is used for optimization procedures. On-node and across-node parallelization is achieved by using GNU parallel [25] where one processor is fed with one image at a time. The processing chain is streamlined in RAM such that the data is only read once and only the very final output is written to disc.

Several modules can be disabled or enabled in any possible combination, e.g. the topographic correction, reprojection and/or tiling modules. The atmospheric correction can also be switched off, in which case TOA reflectance is produced. Instead of using the dark object database retrieval options, externally derived AOD values can be passed to the algorithm. More advanced processing options can also be modified, e.g. the environmental correction can be disabled and either the multiple scattering or the simpler single-scattering approximation can be employed.

Please place Fig. 3 approximately here (page-wide).

#### B. Cloud and cloud shadow detection

We integrated a modified version of the Fmask algorithm [5] and we implemented the modifications described by [26], i.e. the discarding of the termination criterion for shadow matching and the inclusion of an additional darkness filter, as well as most of the Fmask updates [23]. In addition, we modified the match similarity metric for matching the clouds with their shadows. In the original Fmask code [5], a cloud is shifted along a projected search path and a match between this projected shadow and potential cloud and cloud shadow layers is computed. The original cloud is excluded from the match. Nevertheless, if there is a big cloud in the search path, the match similarity is maxed out, because the projected shadow is completely contained in the bigger cloud. Thus, the shadow matching often “runs” into the next big cloud and the actual shadow is missed. Therefore, we only match the projected shadow with the potential shadow layer, which provides good results when combined with the disabling of the termination criterion [26].

Rather than buffering the cloud and cloud objects, we calculate the distance to the next cloud or cloud shadow [27] for each pixel. This approach increases the flexibility of this dataset regarding varying demands of different follow-up applications: the user can decide on how large the buffer should be for his specific application or can make use of the full distance information; e.g. [10] used the cloud distance in a PBC application to score the usability of a given observation by using a transfer function. We append the cloud/shadow pixel distance as 7<sup>th</sup> layer to the processed data (spectral information is stored in the first 6 bands).

In order to increase the computational performance, we implemented two full-stop criteria where the processing of an image is terminated after or within (before the costly shadow matching) the cloud detection module if the scene cloud coverage exceeds a given user-defined threshold. We set this threshold to 25% unbuffered cloud and cloud shadow cover.

### C. Radiometric Processing

Following Tanré's formulation for radiative transfer [28] and by adding a topographic correction factor  $A$  [12, 29], the surface reflectance  $\rho$  can be expressed as [30]:

$$\rho = A \cdot \frac{\rho^*/T_g(\mu_s, \mu_v) \cdot [1 - \langle \rho \rangle \cdot s] - \rho_p \cdot [1 - \langle \rho \rangle \cdot s] - T(\mu_s) \cdot t_s(\mu_v) \cdot \langle \rho \rangle}{T(\mu_s) \cdot t_d(\mu_v)} \quad (1)$$

where  $\rho^*$  is the at-satellite reflectance,  $\rho_p$  is the path reflectance and  $\langle \rho \rangle$  is the background contribution to the apparent target reflectance.  $\mu_s$  and  $\mu_v$  are the cosines of the sun and view zenith angles  $\Theta_s$  and  $\Theta_v$ .  $T(\mu_v)$ ,  $t_d(\mu_v)$  and  $t_s(\mu_v)$  are the total, direct and scattered (i.e. diffuse) upwelling transmittances, respectively; downward transmission terms are indicated by their dependence on  $\mu_s$ .  $T_g(\mu_s, \mu_v)$  is the total gaseous transmission and  $s$  denotes the spherical albedo. Most of the variables in eq. 1 and the following equations are wavelength-dependent; we omitted the wavelength subscript  $\lambda$  for the sake of simpler equations.

The adjacency effect is accounted for by the background contribution  $\langle \rho \rangle$ , which is derived from the weighted sum of the measured apparent reflectance around the target and from the target itself (see [30, 31] for details).

A simplified schematic workflow for the radiometric correction module is shown in Fig. 4.

Please place Fig. 4 approximately here.

#### 1) DN-to-TOA-reflectance conversion

The TOA reflectance  $\rho^*$  of a tilted surface is computed by [32]

$$\rho^* = \pi \cdot d^2 \cdot L^* / (ESUN \cdot \mu_s). \quad (2)$$

The at-satellite radiance  $L^*$  is derived from the calibrated DNs by applying the rescaling factors which are included in the metadata [32]. The mean solar exo-atmospheric spectral irradiance  $ESUN$  was derived by applying sensor-specific relative spectral response (RSR) functions to the Thuillier solar spectrum [32, 33].  $\Theta_s$  and the solar azimuth angle  $\Phi_s$ , as well as the Earth-Sun distance  $d$  [34] are computed by using the date and location from the metadata. Though it is generally accepted to only use the sun position at the scene center [9], we compute sun positions for square blocks of 333 Landsat pixels, which results in a roughly 10 km spaced sun position grid; the view zenith  $\Theta_v$  and azimuth angles  $\Phi_v$  are computed likewise.

#### 2) Atmospheric correction

##### a) Gaseous Transmittance

In our current implementation, we focus on the transmittance of water vapor absorption as first approximation. Amongst the radiance-modifying gases, water vapor is the most variable agent. The gaseous transmittance term in eq. 1 then simplifies to:

$$T_g(\mu_s, \mu_v) = T_w(\mu_s) \cdot T_w(\mu_v), \quad (3)$$

where the down-welling water vapor transmittance function is computed by [35]:

$$T_w(\mu_s) = \exp\left(-\frac{0.2385 \cdot a_w \cdot W \cdot M(\mu_s)}{[1 + 20.07 \cdot a_w \cdot W \cdot M(\mu_s)]^{0.45}}\right), \quad (4)$$

which is dependent on the relative air mass  $M$ , on spectral water vapor absorption coefficients  $a_w$  and on the precipitable water vapor  $W$  (measured in centimeters). The relative air mass  $M$  is defined as [35]:

$$M(\mu_s) = 1/[\mu_s + 0.15 \cdot (93.885 - \theta_s)^{-1.253}]. \quad (5)$$

The up-welling water vapor transmittance is calculated by using  $\Theta_v$  and  $\mu_v$  instead of  $\Theta_s$  and  $\mu_s$ . We derived hyperspectral  $a_w$  values from the high-resolution transmission molecular absorption database (HITRAN) [36] with the Landsat RSR functions.

Precipitable water estimates  $W$  are provided by a previously generated MODIS water vapor database. Therefore, we implemented an operational module for automatic data acquisition and processing of MODIS data. For each WRS-2 frame in continental sub-equatorial Africa, day-specific water vapor loadings are derived if possible. We determine spatial averages from all cloud- and sun-glint-free pixels within the WRS-2 footprints, separately for the Terra and Aqua granules. If estimates from both sensors are available, we choose the one with the larger number of valid pixels, which in most cases is Terra for Landsat-7 (99.4%) and Aqua for Landsat-5/8 (99.2%) due to orbital characteristics. If there are less than 10% of valid pixels, the average is expected to be unreliable and we fall back on an alternative parameterization, which applies for missing data of any kind, including the pre-MODIS era, coverage gaps between the swaths, sensor outages and the like. Southern African climate is dominated by a stable and pronounced seasonality into a dry and wet season. Therefore,  $W$  is replaced by an average seasonal proxy derived from a statistical analysis of the complete MODIS water vapor data sequence - if necessary. This fallback climatology was used in 100% of all cases before 2000, 25.3% in 2000, 2.2% in 2001, 2.8% in 2002 and less than 0.4% thereafter, which reflects the phased commissioning of the Terra and Aqua platforms. Exemplary water vapor data are plotted in Fig. 5 for WRS-2 Path/Row 177/072 with daily values in the top panel and the fallback climatology below. Water vapor follows a seasonal cycle with low values in the dry season and high values in the wet season. The standard deviation is higher in the wet season, which could indicate that the error in using the fallback values is smaller during the dry season during which Landsat data availability is higher. In order to document the usability of this approach across the study area, we appended additional examples in the supplemental material section of this article. We chose the Landsat frames with the highest monthly average and standard deviation. The seasonal and annual variability is higher in the latter example (compared to Fig. 5), but a seasonal pattern is evident nonetheless. In addition, we provide maps of the fallback climatology (as shown in Fig. 5 bottom for one frame) for complete sub-equatorial Africa in the supplemental material section.

Please place Fig. 5 approximately here.

## b) *Aerosol optical depth*

### (1) Radiative Transfer Theory

All remaining unknown parameters in eq. 1 can be derived from aerosol optical depth  $\tau_a$  [37],[30], by either using the single scattering approximation [38] or by also considering multiple scattering processes [39]. Equations 6-19 provide the derivation of these atmospheric scattering terms from  $\tau_a$  using the multiple

scattering approach (refer to [37] for the simpler single scattering approximation). The practical estimation of  $\tau_a$  is outlined in the next chapters (2-4).

The path reflectance  $\rho_p$  is defined by an exact treatment of first order scattering and an approximate solution for higher order scatterings:

$$\rho_p = A_{pl} + [3(1 + g)\mu_v\mu_s - 2(\mu_v + \mu_s) + P] \cdot [1 - \exp(\tau/\mu_v - \tau/\mu_s)]/[4(\mu_v + \mu_s)]. \quad (6)$$

The plane albedo  $A_{pl}$  is given by [40]:

$$A_{pl} = 1 - \frac{R(\mu_v) \cdot R(\mu_s)}{4 + 3(1 - g)\tau}, \quad (7)$$

where the reflectance functions  $R(\mu_s)$  (replace  $\mu_s$  with  $\mu_v$  to obtain  $R(\mu_v)$ ) are given by [40]:

$$R(\mu_s) = 1 + 3 \frac{\mu_s}{2} + \left[1 - 3 \frac{\mu_s}{2}\right] \cdot \exp(-\tau/\mu_s) \quad (8)$$

with the asymmetry factor [41]

$$g = [\alpha \cdot (g_1 + g_2) - g_2] \cdot \tau_a / \tau. \quad (9)$$

The total optical depth

$$\tau = \tau_a + \tau_r \quad (10)$$

is the sum of  $\tau_a$  and the molecular optical depth  $\tau_r$  for a standard Rayleigh atmosphere [42]:

$$\tau_r = 0.0088 \cdot \lambda^{-4.15+0.2\lambda}. \quad (11)$$

The total phase function

$$P = P_a \cdot \tau_a / \tau + P_r \cdot \tau_r / \tau \quad (12)$$

for the backscattering angle

$$\psi_- = \cos^{-1}\{-\mu_v\mu_s - [(1 - \mu_v^2)(1 - \mu_s^2)]^{0.5} \cos(\phi_v - \phi_s)\} \quad (13)$$

is obtained from the phase functions for molecular [43]

$$P_r = 0.75 \cdot (1 + \cos^2\psi_-) \quad (14)$$

and aerosol scattering  $P_a$ . The aerosol scattering equation is approximated by a two-term Henyey-Greenstein (TTHG) function with  $g_1 = 0.836$ ,  $g_2 = 0.537$  and  $\alpha = 0.968$ , representing a continental aerosol model [41]:

$$P_a = \frac{(1 - g_1^2) \cdot \alpha}{[1 + g_1^2 - 2g_1 \cos \psi_-]^{1.5}} + \frac{(1 - g_2^2) \cdot (1 - \alpha)}{[1 + g_2^2 + 2g_2 \cos \psi_-]^{1.5}}. \quad (15)$$

The down-welling total scattering transmittance

$$T(\mu_s) = \exp(-(0.52\tau_r + 0.167\tau_a)/\mu_s), \quad (16)$$

the direct portion

$$t_d(\mu_s) = \exp(-\tau/\mu_s), \quad (17)$$

and the scattered portion

$$t_s(\mu_s) = T(\mu_s) - t_d(\mu_s) \quad (18)$$

are readily computed (the up-welling terms are retrieved by replacing  $\mu_s$  with  $\mu_v$ ). The spherical albedo  $s$ , viewed from the ground, is given by:

$$s = \exp(-(\tau_r + \tau_a)) \cdot (0.92\tau_r + 0.333\tau_a). \quad (19)$$

Thus, solving eq. 1 only requires a robust estimation of  $\tau_a$ , which - in the absence of measurements - is commonly derived from the image itself with dark target techniques, assuming that candidate objects exist in every image (e.g. [37]).

## (2) Dark Object Database

For the practical implementation of this concept, we utilize a partially image-based approach by using a pre-compiled Dark Object Database (DODB). The DODB is created from the complete data series of each pixel and guarantees that the dark objects (DOs) are estimated from the same pixels for all bands and that preferably the temporally more persistent ones are used for estimating  $\tau_a$ . This prevents that temporary dark objects (like flood water, burned areas or cloud shadows) are used in some images while completely different DOs are used in other images that do not contain such objects.

The generation of the DODB is a prerequisite for our pre-processing algorithm, thus we based the generation of the DODB on an analysis of the DN's of every available image. DOs are identified in each image by using the red and near infrared band histograms (see next paragraph). After all the individual images are analyzed, the dark object persistency DOP (with  $DOP = [0,100]$ , i.e. the percentage of the time a pixel is dark) is derived for each pixel in the study area. In order to account for potential land cover change during the past 30+ years, we compiled 3 decadal databases, i.e. [1984,1995], [1996,2005] and [2006,2015]. The DODB is designed such that newly acquired data can be simply integrated into the database by updating the DOP score. Fig. 6 displays a subset of the DOP, as well as two examples of Landsat false color images for an area that is characterized by varying water levels (more water during the late wet season, middle) and the presence of burn scars (more likely during the course of the dry season, bottom). The use of the most persistent DOs effectively prevents the usage of the transient dark features, in this case the flood plains and the burned areas, which were also marked as DO in some images (b/w tones in the top panel).

Please place Fig. 6 approximately here.

We identify DOs in the red and near infrared bands for each processed image. The NIR band is utilized because of (i) the decreasing scattering effect at longer wavelengths [44] which increases the darkening effect of shadowed pixels [5] and because (ii) most surface features are bright and thus maximize the contrast to shadowed areas [5]. The red band is used to reject highly turbid water bodies. For both bands, the DOs are retrieved from the lower bounds of the band histograms.

### (3) Image-based estimation of AOD

Based on the persistent DOs, we employ an object-based approach where it is attempted to derive  $\tau_a$  for each non-transient dark object on a physically sound basis, implicitly including the adjacency effect. During the actual radiometric processing, the DOP is available to the algorithm and is used in combination with the  $\rho^*$  image under consideration. We only use pixels with consecutively decreasing  $\rho^*$  for all bands, which represents an important property of an ideal DO as it should follow a  $\lambda^{-x}$  relation [45]. Objects with less than 10 valid pixels or that are in close proximity to clouds ( $<10$  px) are discarded. The object reflectance  $\rho_o^*$ , environment reflection  $\rho_e^*$  and target radius  $r$  are directly inferred from the image, the target altitude is considered by adjusting  $\tau_r$  [42] and the objects are separated into water and topographic shadow targets based on the illumination situation. If  $\rho_e^*$  is smaller than  $\rho_o^*$ , the object is rejected.

Our approach builds on the method of [46], where  $\rho_o^*$  is matched with the spectral reference reflectance  $\rho_s^*$  of a typical dark object. In case of water, the surface reference reflectance is given by:

$$\rho_s = \rho_w + \rho_f t_s(\mu_s)/T(\mu_s), \quad (20)$$

with  $\rho_w$  being the volumic water reflectance that is obtained from a spectral reference library (see below). The Fresnel surface specular reflectance  $\rho_f$  is computed as a function of the incidence angle  $i$  [47], assuming that water surfaces are flat. Aerosol optical depth is determined in a bandwise iterative procedure by gradually adding atmospheric scattering to  $\rho_s$ , thus simulating the reference reflectance  $\rho_s^*$  as it would be sensed at the top of the atmosphere (solve eq. 1 for  $\rho^*$  with  $A = 1$  while increasing  $\tau_a$ , starting with a pure Rayleigh atmosphere, i.e.  $\tau_a = 0$ ). This process terminates once  $\rho_s^*$  and  $\rho_o^*$  match; see [46] for a detailed description. In the case of cast shadow, we use a modified version of this approach where  $\rho_f$  is ignored but the topographic correction factor  $A$  (see topographic correction) is included in order to model the shaded  $\rho_s^*$ . The procedure is outlined in Fig. 4 in simplified form.

In order to (i) obtain a more physically coherent set of the wavelength dependent  $\tau_a$ , (ii) reduce measurement dependent variations [30] and (iii) to have an instrument for gauging the quality of the estimation ( $R^2$ ) [46], we perform a logarithmic regression between  $\tau_a$  and  $\lambda$ . A second-order polynomial fit is employed (eq. 21), which enables the modelling of the inherent  $\ln \tau_a$  curvature of biomass burning and desert dust aerosols as a consequence of accumulation mode dominating aerosol size distributions [48]. If unsuccessful (e.g. if the Ångström exponent  $a_1 > 0$ ), the simpler linear Ångström relation [49] is tried in a second step, where the second order term in eq. 21 is ignored ( $a_2 = 0$ ). If the fit to the Ångström equation is also unsuccessful (e.g. if  $a_1 > 0$ ), the object is rejected.

$$\ln \tau_a = a_0 + a_1 \cdot \ln \lambda + a_2 \cdot (\ln \lambda)^2. \quad (21)$$

Due to the profound variability of naturally occurring water bodies / land surfaces, we pre-compiled spectral reference libraries (for a range of possible conditions) using spectral modelling software. The water color simulator WASI [50] and the leaf optical properties + canopy bi-directional reflectance model PROSAIL [51] are used for the water and cast shadow targets, respectively. For each object, each reference spectrum  $\rho_s$  is tested and the corresponding  $\tau_a$  estimate that yields the best  $R^2$  in the logarithmic regression is retained; if the best  $R^2$  is smaller than 0.1, the object is rejected.

As dark targets are not abundantly available across the images in our study area [52], we compute a  $\tau_a$  scene average from all available targets, weighted by the  $R^2$  of the logarithmic regressions while considering the individual target altitudes [42]. Once  $\tau_a$  is retrieved, the governing equation (eq. 1) can be solved by computing all parameters as described in chapter (1) using eq. 6-19. We introduce some scene fidelity by using coarsely gridded (333 px) sun-target-view angle dependencies that allow the scattering equations to

vary over the scene. In addition,  $\tau_a$  and  $\tau_r$  are adjusted for the pixel's altitude [42] in order to approximate the elevation dependency on the optical depths.

#### (4) Fallback strategy

The African savanna landscape is often bright [53] and a substantial number of Landsat footprints do not contain dark targets (all objects were rejected, see previous section) in which case we cannot retrieve  $\tau_a$  and consequently employ an alternative fallback strategy in a second step.

We implemented a backdoor-interface, where externally generated  $\tau_a$  values can be passed to the algorithm. In this case, the image-based estimation of  $\tau_a$  is skipped and the external values are used instead. We made use of this possibility for the images where we could not identify any valid dark target and re-processed these images with modelled  $\tau_a$  values. This applied to ~14% of the images and especially occurred in the bright landscapes of southern Namibia and Botswana.

We used the  $\tau_a$  values of all the images where the above presented method worked well and modelled the climatic seasonal course of  $\tau_a$  with a dependence on geolocation: we used a multivariate regression model described by the geolocation and the acquisition day-of-years of the available  $\tau_a$  values to estimate a set of spatio-temporal  $\tau_a$  fallback values with a revisited version of [52] (direct modelling of  $\tau_a$  instead of  $\rho_p$ ).

#### 3) Topography correction

Topography correction is done by applying a modified C-correction [12], which is a physically based correction of topography, amended by an empirically derived extra parameter  $C$  [54]. The topographic correction factor in eq. 1 is determined for every image, band and pixel:

$$A = (\cos \theta_s + C \cdot h_0^{-1}) / (\cos i + C \cdot h_0^{-1} h). \quad (22)$$

The illumination angle  $\cos i$  is computed by [55]

$$\cos i = \cos \theta_s \cos \theta_n + \sin \theta_s \sin \theta_n \cos(\phi_s - \phi_n). \quad (23)$$

The topographic slope  $\theta_n$  and aspect  $\phi_n$  are computed with the Horn method [56] from the DEM that was warped to the extent and resolution of the Landsat image under consideration. The  $h$ -factor describes the portion of the sky dome which contributes to the diffuse illumination, where  $h_0$  is the  $h$ -factor at  $\cos i = 0$  [12]. The  $C$ -factor is estimated from a linear regression between  $\cos i$  and the spectral radiance from an inclined surface  $L^*$ . The empirically-derived  $C$ -factor is then the ratio of intercept  $b$  and slope  $m$ :

$$C = b/m. \quad (24)$$

The main focus of attention in computing the  $A$ -factor is to derive this  $C$ -factor. Because of the image-based nature of this correction, we were in need to incorporate several stabilizing constraints in order to estimate the  $C$ -factor in a robust manner for preferably all images:

- The results of the C-correction can be improved if the  $C$ -factor is not derived for the entire Landsat image, but separately for different land cover classes [12]. Thus we split the image pixels into two classes based on an arbitrary Normalized Differenced Vegetation Index (NDVI) threshold of 0.4 [54].
- If there is no equal abundancy of differently illuminated pixels, the regression-based correction often fails to be representative for all topography classes and results in significant under/overcorrection. [54] used a threshold of  $2^\circ$  slope angle for excluding rather flat pixels. We found that this method worked in many cases, but depending on image-content, this constraint is often not sufficient for Landsat frames with unequally mixed and complex terrain. In our operational setting, we found it

very effective to also stratify the image with  $5^\circ$  slope classes. Thus, we estimate the  $C$ -factor for each land cover and slope class separately.

- If the coefficient of determination  $R^2$  is less than 1 % in a given class, we assume that the C-correction did not perform well and we fall back on a simple Minnaert Correction [29] with a fixed Minnaert coefficient of 0.8. This simplifies the  $A$ -factor to

$$A = (\cos \theta_s / \cos i)^{0.8}. \quad (25)$$

The relationship between  $L^*$  and  $\cos i$  is in general not very high; [12] reported  $R^2$  values between 5 % and 30 % for the different bands of their testing image. Such high  $R^2$  values are not always given if the image is not entirely mountainous, thus we set the threshold to 1 % to test if there is at least any evidence of a relation. The class-dependent topographic correction approach implies that one image may be corrected with different methods, giving precedence to the more advanced C-correction as it commonly comes off as winner in comparative topographic studies, e.g. [54].

#### D. Reprojection and Gridding

Once the Landsat data are radiometrically corrected, they are reprojected into the output coordinate system using bilinear resampling and second order polynomial warping. For our study data, we chose a Lambert Azimuthal Equal Area (LAEA) projection with its origin in the centroid of southern Africa. After reprojection, the data are organized in smaller arbitrary tiles. Therefore, a grid in the target projection is created. The grid originates at a custom point  $O_{grid}$  and a new tile is created each  $n_{grid}$  pixels. For our study area, we chose the arbitrary grid to origin at  $O_{grid} = 0^\circ/0^\circ$  in Lat/Lon and  $n_{grid}$  was chosen to be 1000 px in the LAEA projection. In order to avoid confusion about nomenclature, we hereby define the term

- 'grid' as an arbitrary sub-division with square units in the target coordinate system,
- 'tile' as an entity of the grid with a unique tile identifier, e.g. X0003\_Y0002 and
- 'chip' as the individual gridded images that are affiliated with the tile.

The disintegration of the classical Landsat WRS-2 data structure into a new, gridded tile representation has several benefits: (1) easy and rapid pixel based data access, (2) simple co-registration among images – regardless of their initial path and row designation, and thus (3) exploiting the full depth of the Landsat archive, as well as (4) easier ordering of data for study areas that commonly match imperfectly with WRS-2 frames.

## V. Results

#### A. Processing stats

We processed 58,731 Landsat images of 194 WRS-2 frames with the presented method. 41,762 images were fully processed; 16,876 images were terminated in an early stage because the cloud contamination exceeded the maximum allowable cloud cover threshold and 90 images were so far not processed because of a recent failure in the TIRS calibration. The processed images were partitioned into 4,524 tiles and 1,912,733 chips, resulting in a total data volume of 27.86 TB. The core processing was finished in less than 4 days on a moderately sized processing cluster (2 nodes à 56 CPUs). Fig. 7 displays the number of processed chips per tile. The data availability – corrected for redundant data - is highest in the overlap region between two orbits.

Please place Fig. 7 approximately here.

### B. Spectral consistency in overlap regions

Our gridded data structure enables us to analyze the spectral consistency within three types of overlapping image regions:

- LPGS generates redundant data in the overlap region of adjacent WRS-2 frames that were captured in the same orbit. Thus, any difference in this redundant overlap should be caused exclusively by the implemented processing chain and enables the exploration of the inherent systematic errors/uncertainties.
- The revisit overlaps between two adjacent orbits, are captured 7 days apart by the same Landsat system. The data might vary somehow due to different atmospheric situations and oppositional viewing directions, which we assume to have the greatest impact. In addition, rapid phenology processes, as well as process-based change like changing water levels, burned areas or active fires might add to the difference.
- The observation frequency reduces to 1 day in the cross-sensor overlaps if data from several sensors are considered, though the viewing aspect angle is still substantially different. A comparison between the 1 day and 7 day differences might be useful to differentiate between the cross-sensor uncertainties and differences that are caused by surface processes.

The assessment is based on an exhaustive analysis of all overlapping chips. The spectral consistency for every chip pair was assessed by the spectral RMSE of the 6 reflective bands  $b$ , averaged over all overlapping pixels  $p$  (eq. 26). In order to avoid situations with unlike atmospheric conditions, we omitted pixels that are located within a 10 km radius from clouds or cloud shadows.

$$meanRMSE = \frac{1}{n} \sum_{p=1}^n \sqrt{\frac{1}{6} \sum_{b=1}^6 [chip1_{p,b} - chip2_{p,b}]^2}. \quad (26)$$

Fig. 8 displays histograms of this assessment for the three overlap types. Quantiles are drawn in colors. The differences in the redundant overlaps are smallest: the mean of the revisit (cross-sensor) overlaps is 0.70% (0.84%) larger than in the redundant overlaps. The error in the cross-sensor overlaps is usually larger than in the revisit overlaps (0.14% difference in the mean), though the 99% quantile is slightly smaller. The expected accuracy of the full radiometric processing chain is in the order of  $\pm 2.5\%$  reflectance [57]. 98.8% of all redundant chip pairs are within this range. Considering the orbital overlaps, still 92.8% (91.0%) of the revisit (cross-sensor) chip pairs are enclosed.

Please place Fig. 8 approximately here.

### C. Aerosol optical depth

Sunphotometer data from the Aerosol Robotic Network (AERONET) [58] are commonly considered the most accurate terrestrial  $\tau_a$  measurements, and as such, we derived Landsat-equivalent  $\tau_a$  from all available AERONET data in our study area with eq. 21. The relationship ( $R^2$ ) between the satellite-based and the coinciding ground-measured  $\tau_a$  is between 0.72 and 0.79 for all bands (see Fig. 9), intercepts are in the order of 0.02-0.03 and the relationship is very close to the 1-to-1 line. The colors represent different AERONET sites (blue colors: Zambia, red colors: Namibia, green colors: Botswana, no stations for the remaining countries) and the point sizes indicate the mean DOP of the dark targets. The majority of the observations is from the two Mongu sites (23.15°E, 15.25°S) which are drawn as rectangles.

The fallback surface, i.e. the spatio-temporal aerosol climatology was modelled from all available  $\tau_a$  estimates across the whole study area; see [52] for details. Regarding the quality of the aerosol climatology, [52] found that the modelled seasonal aerosol cycle matches well with an average seasonal cycle derived from the

AERONET Mongu station. An updated assessment of this relationship is included in the supplemental material section of this article with  $R^2$  between 0.81 and 0.84 for all bands. Please place Fig. 9 approximately here.

#### *D. Uncertainty in climatic water vapor database*

If daily water vapor values are not available, we rely on a fallback climatology. In order to better understand the potential error and document the uncertainty in using this method, we simulated the potential effect of varying water vapor concentrations. The effect of water vapor absorption is albedo-dependent. Therefore, we picked a Landsat frame which includes both bright and dark surface elements. The Etosha pan is a large salt pan without drainage. The pan is normally very bright but it can be partially flooded after heavy rain and the water can be retained during the course of the dry season. The analysis is based on Landsat-5 data since the TM is the most water vapor impacted Landsat sensor because of its spectral configuration. We chose two cloud-free images that captured a partially flooded pan, one in the dry (08/10/2008) and one in the late wet season (03/22/2009).

We corrected each image with the retrieved daily water vapor values (DV) and with the corresponding climatic average (CA) which would be used in case the daily value would not be available. In addition, we also increased and decreased the climatic average by one and two standard deviations (see table III). We base our analysis on the resulting difference in surface reflectance between the DV and the climatic variants, see Fig. 10. The impact of water vapor is small on the NIR and SWIR2 bands but clearly evident in the SWIR1 band; the VIS bands are hardly affected and thus not displayed here. The effect is more pronounced for very bright surfaces. The daily water vapor estimate of the dry season image is very similar to the climate average, whereas the daily estimate in the wet season is approximately 1.5 standard deviations smaller than the climate average. Thus, the corrected reflectance differs by 0.53% (4.8%) between the DV and CA for the dry (wet) season image for extremely bright features (100% reflectance). The effect is less pronounced for grey objects (50% reflectance): 0.29% (2.6%) difference. For the SWIR1 wet season images, the difference in corrected reflectance between the CA+ (CA++) and the CA- (CA--) variants is approximately 3% (6.2%) for grey objects. In the dry season, the differences are 2.6% and 5.6%, respectively, thus the uncertainty in the wet season is slightly higher.

Please place table III and Fig. 10 approximately here.

#### *E. Topographic correction*

Fig. 11 illustrates the implemented topographic correction for an illustrative sample area. The depicted images were captured over the mountainous Huila province (southern Angola) in the middle of June (winter solstice) under low illumination conditions (sun elevation:  $41^\circ$ ). The elevation ranges between 534 m and 1532 m. Topographic effects were substantially reduced and the corrected image appears to be rather flat.

Fig. 12 demonstrates the effectiveness of the topographic correction by averaging the NIR reflectance (separately for both land cover classes) with dependence on the topographic aspect. The illumination direction (sun azimuth:  $37^\circ$ ) is clearly visible in the uncorrected image (red), whereas the effect is substantially reduced in case of the C-correction (green). In order to show the advantage of our method over the simpler fallback option, we also corrected this image with the Minnaert correction (blue), which has a tendency for overcorrection.

The topographic correction was quantitatively evaluated by comparing the  $R^2$  of the relationship between  $\cos i$  and  $L$ , before and after the radiometric correction. Fig. 13 displays histograms of the difference in  $R^2$ . All images with more than 10% of sloped terrain ( $> 5^\circ$ ) were used to derive the histograms. The  $R^2$  usually decreases after the correction. The strength of the decrease is different for the six spectral bands. The  $R^2$  of

the blue band occasionally increases after the correction and the decrease is more apparent in the infrared bands.

Please place Fig. 11, Fig. 12 and Fig. 13 approximately here.

## VI. Discussion

### A. Dataset consistency

We mainly based the quantitative assessment of the dataset on an exhaustive analysis of all overlapping chip pairs which enables us to gauge the processing consistency. The differences between processed redundant data are attributed to our processing chain. These differences are usually very small and 98.8% of all overlaps were within the expected 2.5% algorithm accuracy [57]; 99.4% differ by less than 3%. The differences between the chips are likely to be caused by the image based methods, where different image content can cause different AOD retrievals. As such, we conclude that the AOD estimation strategy worked reasonably well (see also next chapter) and we expect our processing chain to be of sufficient quality for any following pixel-based compositing or time series analysis application.

The revisit and cross-sensor overlaps are captured on different dates and inhibit different orbital characteristics, different atmospheric states and potentially also land surface change. We presume that opposite view angles have the greatest impact since we do not correct for BRDF. The differences are indeed greater than in the redundant overlaps. Therefore, we assume that the part of the BRDF, attributed solely to the different viewing geometry (which we partially account for by considering the varying sun-target-view geometry in the atmospheric scattering terms), is in the dimension of this difference, i.e.  $< \sim 1\%$  reflectance. As the sun geometry does not change significantly during this time period, we are not able to assess the actual combined effect of BRDF with the performed overlap analysis, though the impact on our dataset is likely to be significantly larger as recently documented by [59]. Accounting for BRDF effects is complex and not yet a standard correction for Landsat data. An optimal BRDF correction would estimate pixel-based bi-directional parameters by using many observations within a short time period [60] – an approach which is not applicable to Landsat data alone [14]. In addition, the parameter estimation with high temporal and coarse spatial resolution data is also problematic in arid areas during the wet season [61]. Thus, the transfer from MODIS-based bidirectional parameters to Landsat data (e.g. [62]) is also restricted in such environments due to the insufficient angular sampling in parts of the year. One notable and promising approach is to infer a globally applicable set of bi-directional parameters by implementing a sampling design, where pixels from a broad range of differently sloped and illuminated pixels in the Landsat orbital overlaps and in different land cover classes are selected [14]. Nevertheless, a-priori knowledge about the land cover, a confident guess about the spectral stability over time and sufficient input data is needed – demands that are still not met everywhere.

Based on a comparison of the two orbit overlap variants, we might be able to differentiate between the cross-sensor introduced error and changes that are attributed to surface change processes (e.g. green flush events, fire or flooding), because the probability of surface change is higher in the revisit case. In general, the cross-sensor overlap differences were larger than the ones in the revisit overlaps. Thus, we conclude that surface change has less impact than cross-sensor calibration errors on average. As an exception to this, the 99% quantile was larger in the revisit overlaps, which could point to surface change. This implies that surface change happens rarely (within 7 days), but if it does, it has a large impact on the differences of overlapping data. The cross-sensor errors might be amplified by the slightly different spectral configuration of the OLI sensor – compared to the TM and ETM+. [63] recently found that the error in surface reflectance is around 2%, which also partially conforms to our results. The mean cross-sensor error is 1.64%, though the errors between overlapping Landsat-7 and Landsat-5 (1.67%) / Landsat-8 (1.60%) chips do not differ significantly

in our case and the OLI-to-ETM+ error is even slightly smaller than the TM-to-ETM+ error. This might be due to the improved radiometric characterization of the OLI sensor, compared to the TM, or presumably another yet unexplored factor.

### *B. Aerosol optical depth*

The estimation of AOD in bright landscapes is a difficult task and the absence of dark objects on a large scale might preclude the usage of image based assessments [64]. [14] mitigated this problem by using a fixed AOD for Eastern Australia. Whilst this is a sound strategy if the AOD is expected to be rather low, errors are inevitably included if the actual aerosol content is high [65]. The AOD in southern Africa is particularly variable [52] due to the accumulation (washout) of aerosols during the dry burning (wet rain) season [66] and regularly assumes large values (see Fig. 9). Therefore, we developed a method where the image-based dark object retrievals were amended by temporal information on the persistency of dark objects. This approach supports the use of the temporally more persistent dark objects, which helped to increase the quality of the AOD retrievals in an environment where few dark objects exist. Transient dark objects like burned areas were thus successfully rejected.

We used AERONET data to gauge the quality of our AOD estimation strategy. Unfortunately, few AERONET sites exist in the study area; the majority of the stations is confined to Zambia, there also exist three (two) stations in Namibia (Botswana) but no station in Angola and Zimbabwe. In addition, most of the stations also have very few high quality Level 2 data. Most of the data is constrained to the two sites in Mongu, Zambia. Fortunately, this location is relatively central in the study area.

It was shown that our physically-based AOD estimation on basis of radiative transfer theory is in very good agreement with available AERONET data (Fig. 9):  $R^2$  range between 0.72 and 0.79, intercepts are low and slopes are near unity. Despite being mainly driven by the Mongu observations (square signatures), the few observations from the other stations (point signatures) also fit reasonably well (Fig. 9). We therefore expect that the calibrated AOD retrievals are reasonably accurate in areas with similar climate and landscape composition as the sites with AERONET coverage, which are especially the wooded and forested regions in Angola and Zambia. It cannot be ruled out that the errors in the more arid ecosystems are larger and there is still a considerable amount of images (14%) where this strategy did not work due to the high albedo and dryness of savannas and deserts. We reasonably accounted for this with the fallback strategy by using modelled values [52]. We consider these values as an acceptable guess for the actual AOD in a spatially and temporally explicit manner (see [52] and the supplemental material of this article).

Due to the above mentioned difficulties and environmental constraints, we recognize that AOD might still be one of the major uncertainties regarding the quality of our dataset. Nevertheless, we assume the AOD estimations and the AOD climatology to be sufficiently precise to give a reasonable characterization of atmospheric scattering in this highly variable ecosystem.

### *C. Water vapor*

It was shown that the effect of water vapor is small in all but the SWIR1 bands. The TM is the most impacted sensor regarding water vapor because the SWIR1 band is substantially influenced by water vapor absorption. The OLI sensor in turn is hardly affected by water vapor absorption because the SWIR1 band was substantially reduced and was moved well away from the absorption band. It was shown that the difference in using a climatically derived water vapor value was approximately 2.6 % reflectance for grey wet season objects if the daily value is approximately 1.5 standard deviations smaller than the climatology. It was also shown that the uncertainty in using the climatology fallback values is slightly higher in the wet season. The corrected reflectance of grey objects in the dry season differed by approximately 2.6% (5.6%) when considering the 1 (2) standard deviation range from the average. The 2 standard deviation range is only slightly higher than the range of the expected overall algorithm accuracy of  $\pm 2.5\%$  [57]. Therefore we consider it important to derive as precise as possible water vapor loadings for the current state of the

atmosphere; but if not available, the use of a climatologic seasonal cycle is a tolerable fallback strategy with a reasonable uncertainty. It might be worth the effort to improve on the water vapor database, e.g. by filling the pre-MODIS era with NOAA AVHRR derived water vapor estimates (e.g. [67]), which would e.g. decrease the risk that inappropriate values would be used in significantly wetter, drier, hotter or colder years. Yet, this strategy would necessitate a thorough examination of the different methods and sensors with regards to the continuity of the estimated values and was not yet implemented. In any case, the presence of a clear seasonality in water vapor should be verified for other study areas before using this fallback strategy.

#### *D. Topographic correction*

It was shown that the topographic correction caused a decrease in the correlation strength between  $\cos i$  and  $L$  in most cases, suggesting that topographic effects are substantially reduced [68]. This effect was more pronounced in the NIR and SWIR bands due to the stronger darkening effect of topographic shadow in the longer wavelength bands [44], where the removal of this effect has a greater impact. Contrary, the blue band was occasionally affected by an increase in the correlation strength. The blue band is not the best estimator of assessing the topographic effect, because thick haze can completely impede a relationship between  $\cos i$  and  $L$ . As such, we tend to not overvalue the  $R^2$  increases in the blue band. Random manual investigation points to a successful topographic normalization, which is also supported by the presented example. In addition, it was shown that the implemented topographic correction performed better than the Minnaert fallback option. Cosine-based corrections tend to overcorrect areas under low illumination conditions [69] and the Minnaert factor's purpose is simply to dampen the correction strength. One way to optimize the Minnaert value would be to estimate it via a linear regression [69] as it is the case for the C-correction. Nevertheless, as the Minnaert correction is the fallback option in the case that the regression-based C-correction did not work, estimating the Minnaert factor would also not work, and as such, a fixed correction factor is required for this purpose.

## VII. Outlook and Conclusion

In a future version of our algorithm, we consider to expand on the radiometric correction module. Further atmospheric gases like ozone or uniformly mixed gases are to be included and we also consider to provide a more spatially explicit treatment of AOD by interpolating between the dark objects as done by [13]. Nevertheless, these strategies will only work if there are plenty of dark objects, which is a severe constraint in southern Africa [53]. Therefore, the benefit of any further correction should be evaluated with regards (i) to the actual improvement in the corrected reflectance, (ii) to the extra cost in computing time, (iii) to the global applicability and (iv) to the global data availability that any new add-on would require.

The geometric correction of Landsat L1T data is commonly precise; [19] report on a global geometric accuracy of  $\sim 50$  m. Nevertheless, southern African L1T data are occasionally poorly co-registered. In our current implementation, we do not account for this potential error source but an additional bulk-image registration module could be included in the processing chain.

We also intend to prepare and adapt our framework for new Landsat-like systems like any upcoming Landsat-9+ spacecraft or the soon-to-be-available Sentinel-2 data. We consider our approach to be sufficiently transferable to similar medium-resolution data as we are already processing data from all available Landsat systems with the same algorithm, which is e.g. not employed in LEDAPS. The gridded data structure, the sensor-specific water vapor absorption coefficients and exo-atmospheric irradiances will facilitate the incorporation of other data, though specific adaptations will surely be necessary. For Sentinel-2 data, it will be mandatory to include a full BRDF treatment because of the larger swath width. Fortunately, the observing geometry and acquisition frequency of Sentinel-2, combined with Landsat, will also facilitate to obtain BRDF parameters from the data itself, which is very difficult with Landsat data alone.

As already highlighted by [9], the development of (i) higher level products, (ii) from multi-sensor data, (iii) provided in arbitrarily – but regularly – divided tiles is a key component in enabling end-users to make the

best use of medium resolution data, both across space and time. Currently, these demands are not fully met. Landsat standard products lack systematic gridding; higher-level Sentinel-2 products are currently not planned; and Landsat-legacy, Landsat-8 and Sentinel-2 data are not processed with the same algorithm. As such, we specifically developed our processing strategy to derive gridded surface reflectance and cloud/shadow products from multi-sensor data – potentially also Sentinel-2 or similar data in the near future. Gridded higher level products will significantly simplify the application of pixel-based algorithms and will also allow a broader range of end-users to perform such analyses. We chose sufficiently elaborate methods with as few as possible input data. As such, we presume that the presented processing framework may even generate consistent and high qualitative data in areas where the general data availability is still low, which we demonstrated for southern Africa. As such, we preferred image-based solutions where applicable, e.g. for aerosol estimation and topographic correction and opted to not account for BRDF effects until the general applicability to medium resolution data without any pre-knowledge of the study area has become more feasible.

## Acknowledgment

Landsat and SRTM data courtesy of the U.S. Geological Survey. The MODIS data used in this study were acquired as part of the NASA's Earth-Sun System Division and archived and distributed by the MODIS Adaptive Processing System (MODAPS). We thank all PIs and their staff for establishing and maintaining the AERONET sites used in this investigation. KAZA TFCA information and boundary courtesy of Peace Parks Foundation. The authors also thank the two anonymous reviewers whose very constructive comments significantly improved the quality of the final manuscript. This work was funded by the Federal Ministry of Education and Research under contract number FKZ-01LG1201C as part of the Southern African Science Service Centre for Climate Change and Adaptive Land Management project.

## References

- [1] W. B. Cohen and S. N. Goward, "Landsat's Role in Ecological Applications of Remote Sensing," *BioScience*, vol. 54, no. 6, pp. 535-545, Jun., 2004; DOI: 10.1641/0006-3568(2004)054[0535:lrieao]2.0.co;2.
- [2] M. A. Wulder, J. C. White, S. N. Goward, *et al.*, "Landsat continuity: Issues and opportunities for land cover monitoring," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 955-969, Mar., 2008; DOI: 10.1016/j.rse.2007.07.004.
- [3] T. Danaher, P. Scarth, J. Armston, *et al.*, "Remote Sensing of Tree-Grass Systems: The Eastern Australian Woodlands," *Ecosystem Function in Savannas: Measurement and Modeling at Landscape to Global Scales*, pp. 175-194, Nov., 2010; DOI: 10.1201/b10275-13.
- [4] C. E. Woodcock, R. Allen, M. Anderson, *et al.*, "Free access to Landsat imagery," *Science*, vol. 320, no. 5879, p. 1011a, May, 2008; DOI: 10.1126/science.320.5879.1011a.
- [5] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, no. 0, pp. 83-94, Mar., 2012; DOI: 10.1016/j.rse.2011.10.028.
- [6] M. A. Wulder, J. G. Masek, W. B. Cohen, *et al.*, "Opening the archive: How free data has enabled the science and monitoring promise of Landsat," *Remote Sens. Environ.*, vol. 122, no. 0, pp. 2-10, Jul., 2012; DOI: 10.1016/j.rse.2012.01.010.
- [7] J. G. Masek, C. Huang, R. Wolfe, *et al.*, "North American forest disturbance mapped from a decadal Landsat record," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 2914-2926, Jun., 2008; DOI: 10.1016/j.rse.2008.02.010.
- [8] R. Sonnenschein, T. Kuemmerle, T. Udelhoven, *et al.*, "Differences in Landsat-based trend analyses in drylands due to the choice of vegetation estimate," *Remote Sens. Environ.*, vol. 115, no. 6, pp. 1408-1420, Jun., 2011; DOI: 10.1016/j.rse.2011.01.021.
- [9] M. C. Hansen and T. R. Loveland, "A review of large area monitoring of land cover change using Landsat data," *Remote Sens. Environ.*, vol. 122, no. 0, pp. 66-74, Jul., 2012; DOI: 10.1016/j.rse.2011.08.024.
- [10] P. Griffiths, S. van der Linden, T. Kuemmerle, *et al.*, "A Pixel-Based Landsat Compositing Algorithm for Large Area Land Cover Mapping," *IEEE J. Sel. Topics Appl. Earth Observ. and Remote Sens.*, vol. 6, no. 5, pp. 2088-2101, Oct., 2013; DOI: 10.1109/jstars.2012.2228167.
- [11] D. P. Roy, J. Ju, K. Kline, *et al.*, "Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 35-49, Jan., 2010; DOI: 10.1016/j.rse.2009.08.011.
- [12] S. Kobayashi and K. Sanga-Ngoie, "The integrated radiometric correction of optical remote sensing imagery," *International Journal of Remote Sensing*, vol. 29, no. 20, pp. 5957-5985, Oct., 2008; DOI: 10.1080/01431160701881889.
- [13] J. G. Masek, E. F. Vermote, N. E. Saleous, *et al.*, "A Landsat surface reflectance dataset for North America, 1990-2000," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 68-72, Jan., 2006; DOI: 10.1109/lgrs.2005.857030.
- [14] N. Flood, T. Danaher, T. K. Gill, *et al.*, "An Operational Scheme for Deriving Standardised Surface Reflectance from Landsat TM/ETM+ and SPOT HRG Imagery for Eastern Australia," *Remote Sensing*, vol. 5, no. 1, pp. 83-109, Jan., 2013; DOI: 10.3390/rs5010083.
- [15] R. DeFries, F. Achard, S. Brown, *et al.*, "Earth observations for estimating greenhouse gas emissions from deforestation in developing countries," *Environmental Science & Policy*, vol. 10, no. 4, pp. 385-394, Jun., 2007; DOI: 10.1016/j.envsci.2007.01.010.
- [16] S. E. Nicholson, "Rainfall and Atmospheric Circulation during Drought Periods and Wetter Years in West Africa," *Monthly Weather Review*, vol. 109, no. 10, pp. 2191-2208, Oct., 1981; DOI: 10.1175/1520-0493(1981)109<2191:raacd>2.0.co;2.

- [17] D. M. Olson, E. Dinerstein, E. D. Wikramanayake, *et al.*, "Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity," *BioScience*, vol. 51, no. 11, pp. 933-938, November, 2001; DOI: 10.1641/0006-3568(2001)051[0933:teotwa]2.0.co;2.
- [18] R. W. S. Fynn, M. Chase, and A. Röder, "Functional Habitat Heterogeneity and Large Herbivore Seasonal Habitat Selection in Northern Botswana," *South African Journal of Wildlife Research*, vol. 44, no. 1, pp. 1-15, 2014/04/01, 2014; DOI: 10.3957/056.044.0103.
- [19] T. R. Loveland and J. L. Dwyer, "Landsat: Building a strong future," *Remote Sens. Environ.*, vol. 122, no. 0, pp. 22-29, Jul., 2012; DOI: 10.1016/j.rse.2011.09.022.
- [20] M. A. Wulder, J. C. White, T. R. Loveland, *et al.*, "The global Landsat archive: Status, consolidation, and direction," *Under review in Remote Sensing of Environment*.
- [21] USGS. (2015, Sept.). *SRTM 1 Arc-Second Global* [Web Page]. Available: <https://lta.cr.usgs.gov/SRTM1Arc>
- [22] B.-C. Gao and Y. J. Kaufman, "Water vapor retrievals using Moderate Resolution Imaging Spectroradiometer (MODIS) near-infrared channels," *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D13, pp. n/a-n/a, 2003; DOI: 10.1029/2002jd003023.
- [23] Z. Zhu, S. Wang, and C. E. Woodcock, "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.*, vol. 159, no. 0, pp. 269-277, Mar., 2015; DOI: 10.1016/j.rse.2014.12.014.
- [24] GDAL, "GDAL - Geospatial Data Abstraction Library, Version 1.10.0," ed: Open Source Geospatial Foundation, 2013.
- [25] O. Tange, "GNU Parallel—the command-line power tool," *The USENIX Magazine*, vol. 36, no. 1, pp. 42-47, 2011.
- [26] D. Frantz, A. Roder, T. Udelhoven, *et al.*, "Enhancing the Detectability of Clouds and Their Shadows in Multitemporal Dryland Landsat Imagery: Extending Fmask," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1242-1246, Jun., 2015; DOI: 10.1109/lgrs.2015.2390673.
- [27] A. Meijster, J. B. T. M. Roerdink, and W. H. Hesselink, "A General Algorithm for Computing Distance Transforms in Linear Time," in *Mathematical Morphology and its Applications to Image and Signal Processing*, vol. 18, J. Goutsias, *et al.*, Eds., ed, Springer US, 2000, ch. 36, pp. 331-340.
- [28] D. Tanré, C. Deroo, P. Duhaut, *et al.*, "Description of a computer code to simulate the satellite signal in the solar spectrum: the SS code," *International Journal of Remote Sensing*, vol. 11, no. 4, pp. 659-668, Apr., 1990; DOI: 10.1080/01431169008955048.
- [29] P. Teillet, B. Guindon, and D. Goodenough, "On the slope-aspect correction of multispectral scanner data," *Canadian Journal of Remote Sensing*, vol. 8, no. 2, pp. 84-106, Aug., 1982; DOI: 10.1080/07038992.1982.10855028.
- [30] J. Hill, "High precision land cover mapping and inventory with multi-temporal earth observation satellite data: the Ardèche experiment," Ph.D. Ph.D. dissertation, Faculty of Geography/Geosciences, Trier University, 1993. [Online].
- [31] D. Tanre, M. Herman, and P. Y. Deschamps, "Influence of the background contribution upon space measurements of ground reflectance," *Appl. Opt.*, vol. 20, no. 20, pp. 3676-3684, Oct., 1981; DOI: 10.1364/ao.20.003676.
- [32] G. Chander, B. L. Markham, and D. L. Helder, "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors," *Remote Sens. Environ.*, vol. 113, no. 5, pp. 893-903, May, 2009; DOI: 10.1016/j.rse.2009.01.007.
- [33] G. Thuillier, M. Hersé, D. Labs, *et al.*, "The Solar Spectral Irradiance from 200 to 2400 nm as Measured by the SOLSPEC Spectrometer from the Atlas and Eureca Missions," *Solar Physics*, vol. 214, no. 1, pp. 1-22, May, 2003; DOI: 10.1023/a:1024048429145.
- [34] J. Spencer, "Fourier series representation of the position of the sun," *Search*, vol. 2, no. 5, pp. 172-172, 1971.
- [35] R. E. Bird and C. Riordan, "Simple Solar Spectral Model for Direct and Diffuse Irradiance on Horizontal and Tilted Planes at the Earth's Surface for Cloudless Atmospheres," *Journal of Climate and Applied Meteorology*, vol. 25, no. 1, pp. 87-97, Jan., 1986; DOI: 10.1175/1520-0450(1986)025<0087:sssmfd>2.0.co;2.
- [36] L. Rothman, I. Gordon, Y. Babikov, *et al.*, "The HITRAN2012 molecular spectroscopic database," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 130, pp. 4-50, Nov., 2013; DOI: 10.1016/j.jqsrt.2013.07.002.
- [37] J. Hill and B. Sturm, "Radiometric correction of multitemporal Thematic Mapper data for use in agricultural land-cover classification and vegetation monitoring," *International Journal of Remote Sensing*, vol. 12, no. 7, pp. 1471-1491, Jul., 1991; DOI: 10.1080/01431169108955184.
- [38] H. Gordon, "A Preliminary Assessment of the Nimbus-7 CZCS Atmospheric Correction Algorithm in a Horizontally Inhomogeneous Atmosphere," in *Oceanography from Space*, vol. 13, J. F. R. Gower, Ed., ed, Springer US, Boston, MA, 1981, ch. 30, pp. 257-265.
- [39] V. V. Sobolev, "Light scattering in planetary atmospheres," (*Translation of Rasseyaniye sveta v atmosferakh planet, Moscow, Izdatel'stvo Nauka, 1972.*) *Oxford and New York, Pergamon Press (International Series of Monographs in Natural Philosophy. Volume 76)*, 1975. 263 p., vol. 1, 1975.
- [40] W. M. Irvine, "Multiple scattering in planetary atmospheres," *Icarus*, vol. 25, no. 2, pp. 175-204, Jun., 1975; DOI: 10.1016/0019-1035(75)90019-6.
- [41] S. Aranuvachapun, "Satellite remote sensing of atmospheric optical depth spectrum," *International Journal of Remote Sensing*, vol. 7, no. 4, pp. 499-514, Apr., 1986; DOI: 10.1080/01431168608954705.
- [42] R. Guzzi, R. Rizzi, and G. Zibordi, "Atmospheric correction of data measured by a flying platform over the sea: elements of a model and its experimental validation," *Appl. Opt.*, vol. 26, no. 15, pp. 3043-3051, Aug., 1987; DOI: 10.1364/ao.26.003043.
- [43] A. Bucholtz, "Rayleigh-scattering calculations for the terrestrial atmosphere," *Appl. Opt.*, vol. 34, no. 15, pp. 2765-2773, May, 1995; DOI: 10.1364/ao.34.002765.
- [44] Y. Luo, A. P. Trishchenko, and K. V. Khlopenkov, "Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America," *Remote Sens. Environ.*, vol. 112, no. 12, pp. 4167-4185, Dec., 2008; DOI: 10.1016/j.rse.2008.06.010.
- [45] P. S. Chavez Jr, "An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data," *Remote Sens. Environ.*, vol. 24, no. 3, pp. 459-479, Apr., 1988; DOI: 10.1016/0034-4257(88)90019-3.
- [46] A. Royer, L. Charbonneau, and P. M. Teillet, "Interannual landsat-MSS reflectance variation in an urbanized temperate zone," *Remote Sens. Environ.*, vol. 24, no. 3, pp. 423-446, Apr., 1988; DOI: 10.1016/0034-4257(88)90017-X.
- [47] S. Kay, J. Hedley, and S. Lavender, "Sun Glint Correction of High and Low Spatial Resolution Images of Aquatic Scenes: a Review of Methods for Visible and Near-Infrared Wavelengths," *Remote Sensing*, vol. 1, no. 4, p. 697, 2009.
- [48] T. F. Eck, B. N. Holben, J. S. Reid, *et al.*, "Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols," *Journal of Geophysical Research: Atmospheres*, vol. 104, no. D24, pp. 31333-31349, 1999; DOI: 10.1029/1999jd900923.
- [49] A. Ångström, "The parameters of atmospheric turbidity," *Tellus*, vol. 16, no. 1, pp. 64-75, 1964; DOI: 10.1111/j.2153-3490.1964.tb00144.x.
- [50] P. Gege, "The water color simulator WASI: an integrating software tool for analysis and simulation of optical in situ spectra," *Computers & Geosciences*, vol. 30, no. 5, pp. 523-532, Jun., 2004; DOI: 10.1016/j.cageo.2004.03.005.
- [51] S. Jacquemoud, W. Verhoef, F. Baret, *et al.*, "PROSPECT + SAIL models: A review of use for vegetation characterization," *Remote Sens. Environ.*, vol. 113, Supplement 1, no. 0, pp. S56-S66, Sept., 2009; DOI: 10.1016/j.rse.2008.01.026.
- [52] D. Frantz, A. Röder, M. Stellmes, *et al.*, "On the derivation of a spatially distributed aerosol climatology for its incorporation in a radiometric Landsat pre-processing framework," *Remote Sensing Letters*, vol. 6, no. 8, pp. 647-656, 2015/08/03, 2015; DOI: 10.1080/2150704x.2015.1070314.
- [53] Y. J. Kaufman and C. Sendra, "Algorithm for automatic atmospheric corrections to visible and near-IR satellite imagery," *International Journal of Remote Sensing*, vol. 9, no. 8, pp. 1357-1381, Aug., 1988; DOI: 10.1080/01431168808954942.

- [54] S. Hantson and E. Chuvieco, "Evaluation of different topographic correction methods for Landsat imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 13, no. 5, pp. 691-700, Oct., 2011; DOI: 10.1016/j.jag.2011.05.001.
- [55] J. A. Smith, "The Lambertian assumption and Landsat data," *Photogrammetric Engineering & Remote Sensing*, vol. 46, pp. 1183-1189, 1980, 1980.
- [56] B. K. P. Horn, "Hill shading and the reflectance map," *Proceedings of the IEEE*, vol. 69, no. 1, pp. 14-47, Jan., 1981; DOI: 10.1109/proc.1981.11918.
- [57] A. Röder, T. Kuemmerle, and J. Hill, "Extension of retrospective datasets using multiple sensors. An approach to radiometric intercalibration of Landsat TM and MSS data," *Remote Sens. Environ.*, vol. 95, no. 2, pp. 195-210, Mar., 2005; DOI: 10.1016/j.rse.2004.12.008.
- [58] B. N. Holben, T. F. Eck, I. Slutsker, *et al.*, "AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization," *Remote Sens. Environ.*, vol. 66, no. 1, pp. 1-16, Oct., 1998; DOI: 10.1016/S0034-4257(98)00031-5.
- [59] J. R. Nagol, J. O. Sexton, D.-H. Kim, *et al.*, "Bidirectional effects in Landsat reflectance estimates: Is there a problem to solve?," *ISPRS Journal of Photogrammetry and Remote Sensing*, no. 0, May; DOI: 10.1016/j.isprsjprs.2014.09.006.
- [60] C. B. Schaaf, F. Gao, A. H. Strahler, *et al.*, "First operational BRDF, albedo nadir reflectance products from MODIS," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 135-148, Nov., 2002; DOI: 10.1016/S0034-4257(02)00091-3.
- [61] A. Huete, K. Didan, T. Miura, *et al.*, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sens. Environ.*, vol. 83, no. 1/2, pp. 195-213, Nov., 2002; DOI: 10.1016/S0034-4257(02)00096-2.
- [62] D. P. Roy, J. Ju, P. Lewis, *et al.*, "Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data," *Remote Sens. Environ.*, vol. 112, no. 6, pp. 3112-3130, Jun., 2008; DOI: 10.1016/j.rse.2008.03.009.
- [63] N. Flood, "Continuity of Reflectance Data between Landsat-7 ETM+ and Landsat-8 OLI, for Both Top-of-Atmosphere and Surface Reflectance: A Study in the Australian Landscape," *Remote Sensing*, vol. 6, no. 9, pp. 7952-7970, 2014.
- [64] S. S. Gillingham, N. Flood, T. K. Gill, *et al.*, "Limitations of the dense dark vegetation method for aerosol retrieval under Australian conditions," *Remote Sensing Letters*, vol. 3, no. 1, pp. 67-76, Jan., 2011; DOI: 10.1080/01431161.2010.533298.
- [65] S. S. Gillingham, N. Flood, and T. K. Gill, "On determining appropriate aerosol optical depth values for atmospheric correction of satellite imagery for biophysical parameter retrieval: requirements and limitations under Australian conditions," *International Journal of Remote Sensing*, vol. 34, no. 6, pp. 2089-2100, 2013/03/20, 2012; DOI: 10.1080/01431161.2012.738945.
- [66] T. F. Eck, B. N. Holben, D. Ward, *et al.*, "Characterization of the optical properties of biomass burning aerosols in Zambia during the 1997 ZIBBEE field campaign," *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 106, no. D4, pp. 3425-3448, Feb., 2001; DOI: 10.1029/2000JD900555.
- [67] J. A. Sobrino, N. Raissouni, J. Simarro, *et al.*, "Atmospheric water vapor content over land surfaces derived from the AVHRR data: application to the Iberian Peninsula," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 37, no. 3, pp. 1425-1434, May, 1999; DOI: 10.1109/36.763306.
- [68] J. Hill, W. Mehl, and V. Radeloff, "Improved forest mapping by combining corrections of atmospheric and topographic effects in Landsat TM imagery," *Sensors and environmental applications of remote sensing*, pp. 143-151, 1995.
- [69] P. Meyer, K. I. Itten, T. Kellenberger, *et al.*, "Radiometric corrections of topographically induced effects on Landsat TM data in an alpine environment," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 48, no. 4, pp. 17-28, Aug., 1993; DOI: 10.1016/0924-2716(93)90028-L.



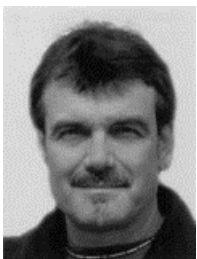
**David Frantz** is a researcher at the Department of Environmental Remote Sensing & Geoinformatics at Trier University, Germany. He received the Diploma degree in Applied Environmental Sciences with a major in remote sensing from Trier University, Germany, 2012 and is currently pursuing the Ph.D. degree. His research interests cover the monitoring of dryland ecosystems using multispectral optical remote sensing data. His main research focus is on the radiometric mass pre-processing of remote sensor data, the development and application of data fusion techniques, as well as the development of pixel-based Landsat compositing algorithms for the monitoring of forest dynamics across large areas.



**Achim Röder** is a senior scientist at the Department of Environmental Remote Sensing & Geoinformatics at Trier University, Germany. Raised as a geographer, he holds a Ph.D. in natural sciences for the development of a remote sensing based monitoring framework for Mediterranean rangelands. His general scientific interest is on the integration of optical remote sensing and geospatial data analysis for environmental studies. In particular, he focuses on the assessment of temporal dynamics of land use/cover using change detection and time series analysis techniques and the development of operational spectral mixture analysis concepts. Furthermore, he is interested in analyzing remote sensing based change dynamics to identify driving factors and assess trade-offs in interdisciplinary contexts.



**Marion Stellmes** is a senior scientist in the Department of Environmental Remote Sensing and Geoinformatics at Trier University, Germany, where she also obtained her Ph.D. in 2011. She is interested in interdisciplinary environmental research, especially in the evaluation of ecosystem goods and services and has a strong background in remote sensing based monitoring of land degradation and desertification in the Mediterranean and Southern Africa. Furthermore, she is specialized in time series analysis of medium and coarse resolution remote sensing time series and has been working for more than ten years in the field of land cover/use change detection. She received her Diploma in Applied Environmental Sciences in 2002, with an emphasis on optical remote sensing and geomathematics.



**Joachim Hill** received a Ph.D. degree from Trier University, Trier, Germany in 1991, and a postdoctoral degree (Habilitation) from the Ludwig-Maximilians-University in Munich in 1993. From 1984 to 1994 he was with the Joint Research Centre of the European Commission in Ispra, Italy. Since 1994, he has been a Full Professor in remote sensing at Trier University, Trier, Germany, where he is Head of the department of Environmental Remote Sensing and Geoinformatics. He is a member of the scientific advisory group of the Environmental Monitoring and Analysis Program (EnMAP), a German hyperspectral satellite mission to be launched in 2018. His research in remote sensing focuses on data processing strategies for monitoring forest ecosystems and land degradation processes.

## Figures

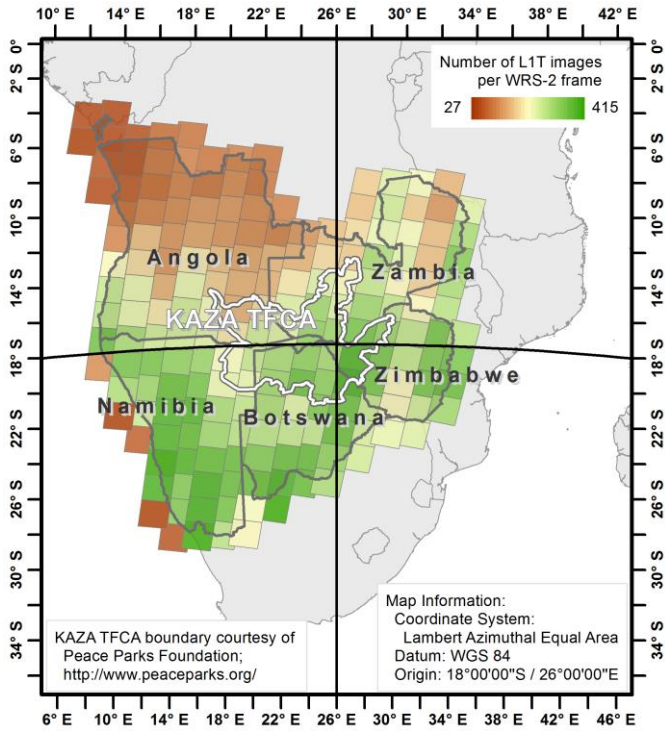


Fig. 1. Location of the study area and number of available L1T images per Landsat frame.

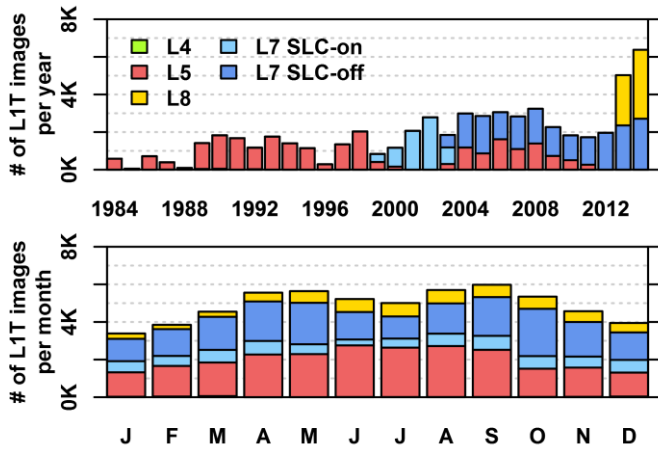


Fig. 2. Number of available L1T images per year (top) and month (bottom).

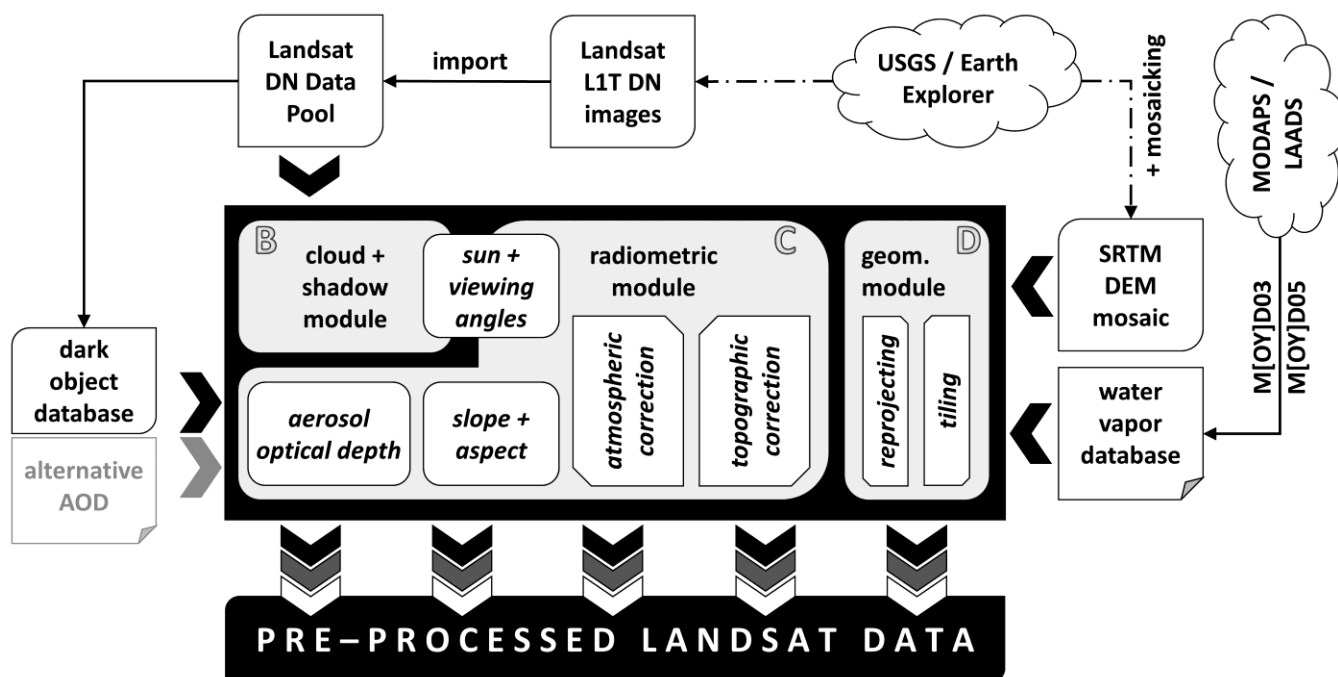


Fig. 3. Schematic workflow of the pre-processing framework. B, C and D refer to the subchapters of the methods section.

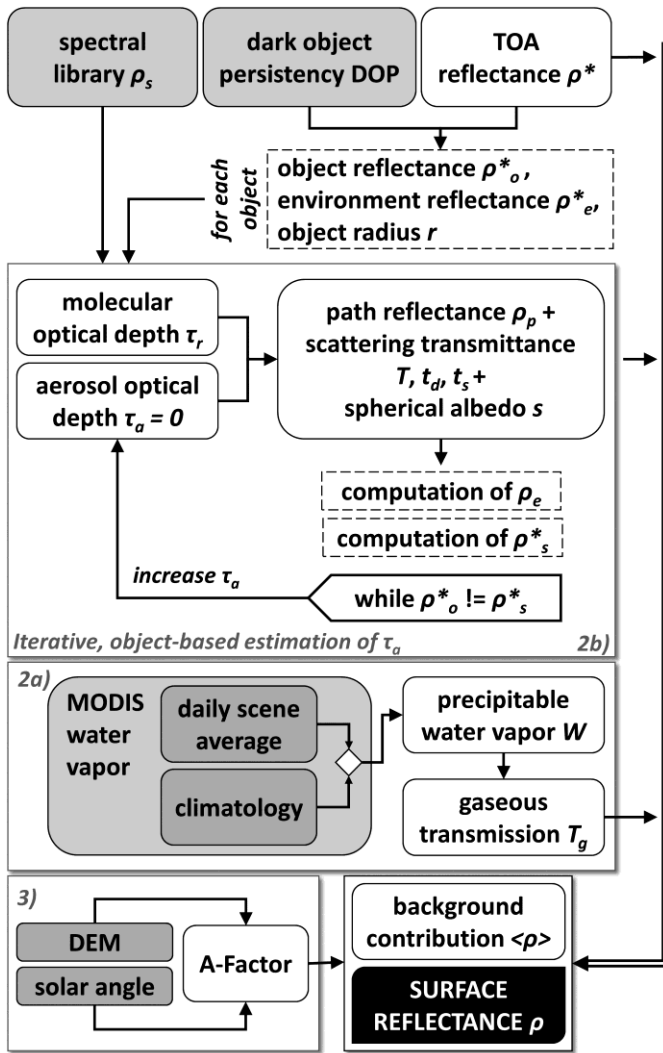


Fig. 4. Workflow for the radiometric processing. The subsections of this chapter are grouped in boxes.

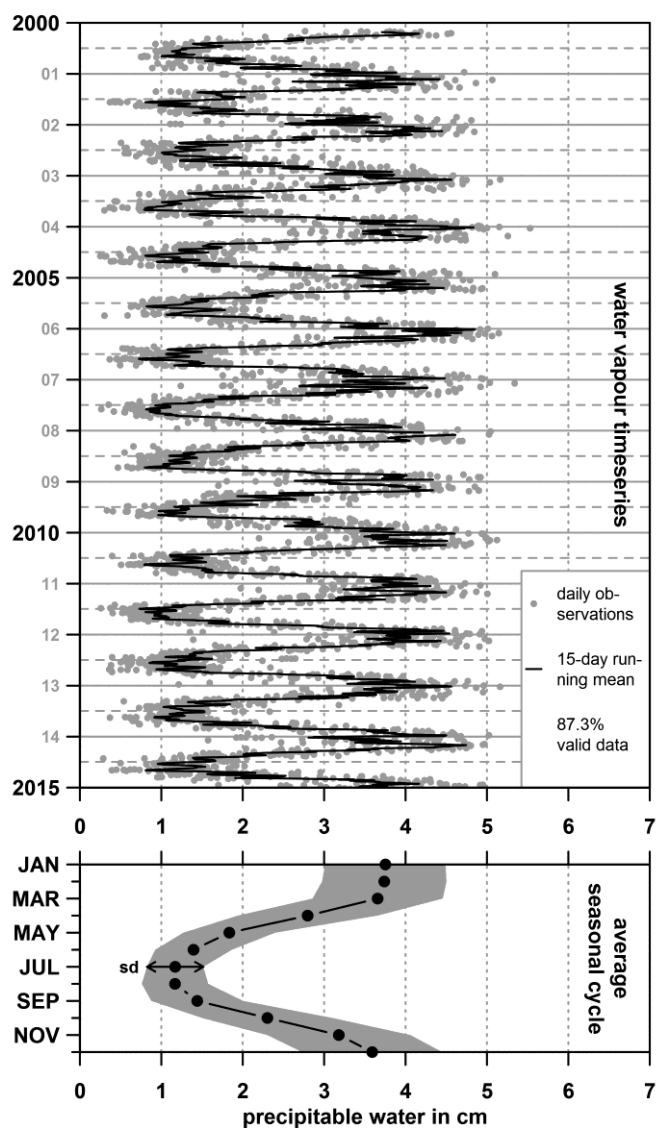


Fig. 5. Visualization of the water vapor database for WRS-2 Path/Row 177/072 (19.71°E, 17.35°S). Top: daily values and 15-day running mean; bottom: average seasonal cycle and standard deviation on monthly basis.

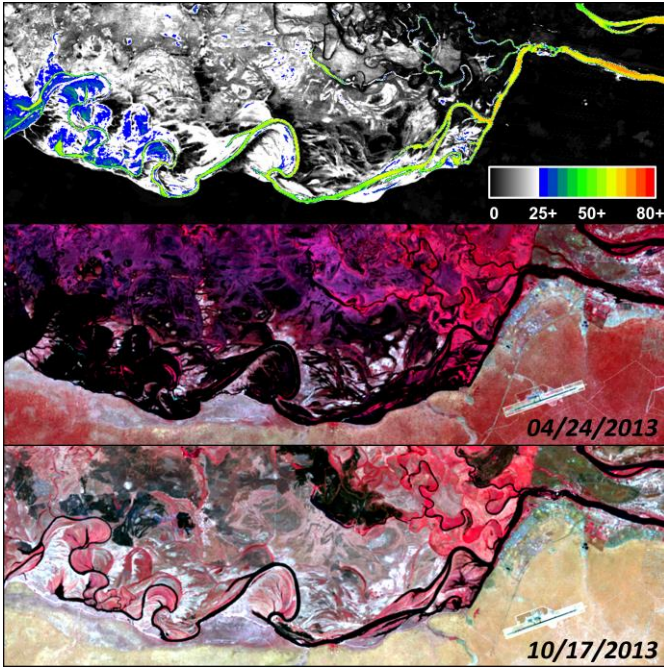


Fig. 6. Visualization of a section of the dark object database and two example images (25.06°E, 17.79°S). Top: Dark object persistency; middle: false-color image of a late wet-season image; bottom: false-color image of a dry-season image. The same stretch was applied to both images with R/G/B = NIR/red/green.

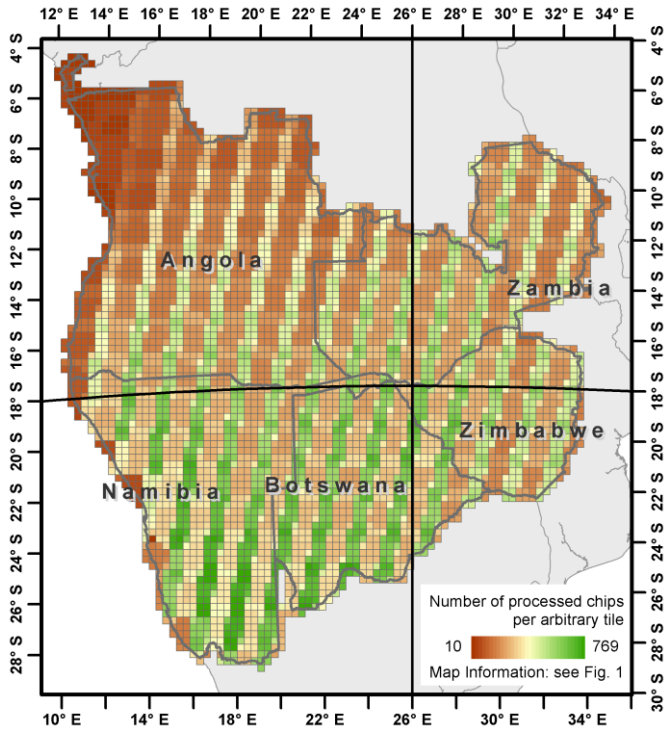


Fig. 7. Number of processed chips per arbitrary tile, corrected for redundant overlaps.

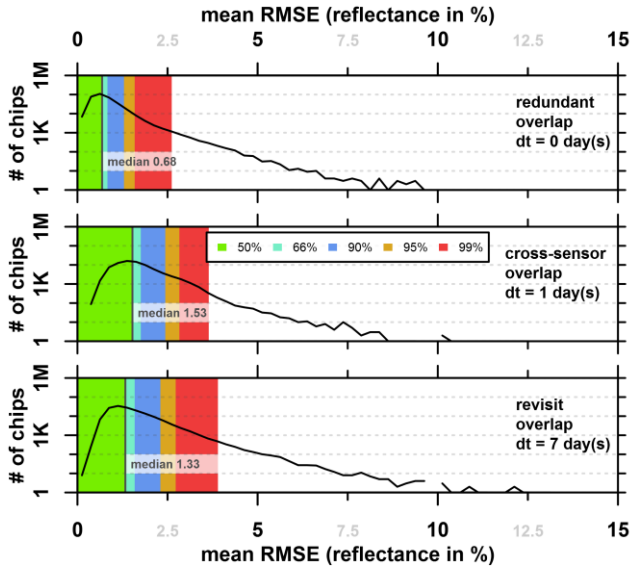


Fig. 8. Histograms of the spectral consistency assessment of overlapping chip pairs. The histogram bin width was set to 0.25% reflectance. The y-axis is drawn logarithmic. Quantiles are drawn in colors.

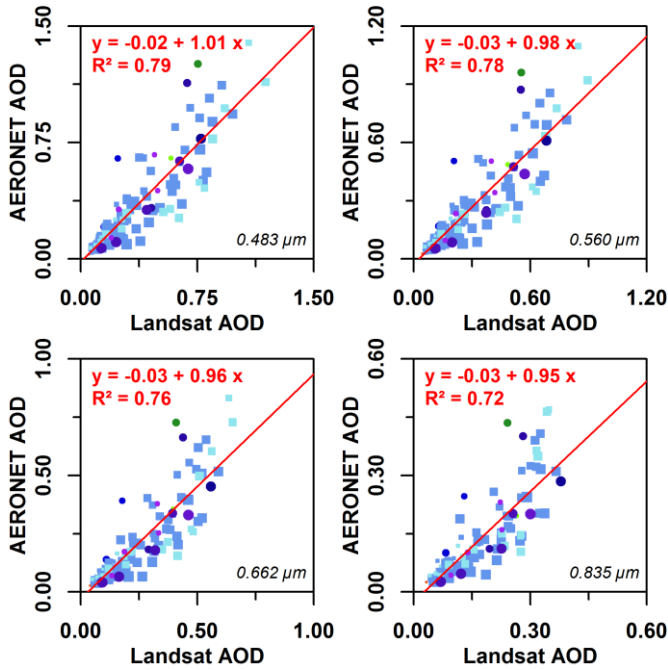


Fig. 9. Linear regression between image-based Landsat and terrestrial AERONET AOD measurements for all stations and all coinciding observations. The colors represent different AERONET sites; blue to purple: Zambia, red/orange: Namibia, green: Botswana. The majority of the data is from the Mongu sites, Zambia.

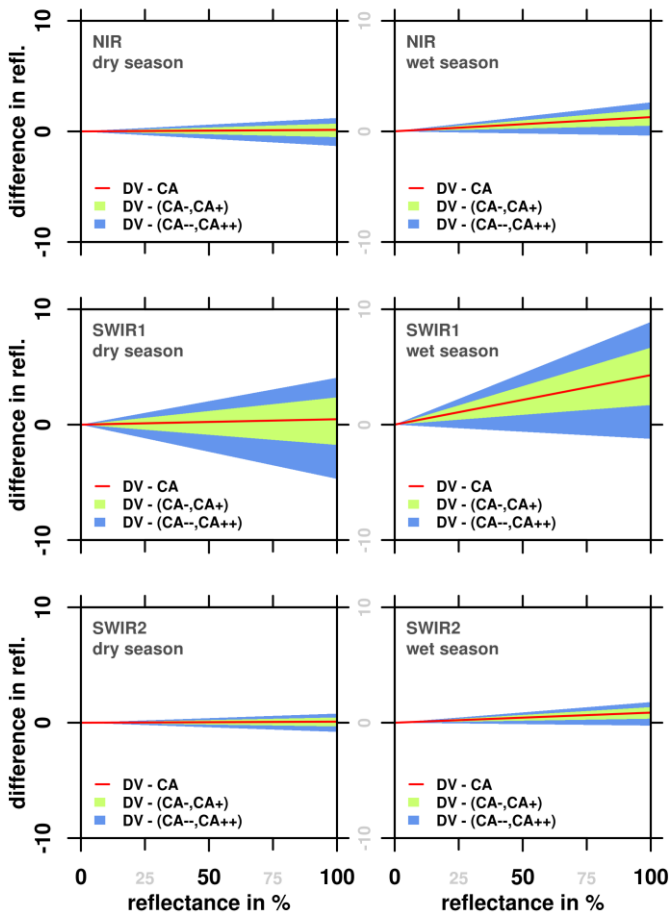


Fig. 10. Potential errors in using climatically derived water vapor estimates instead of daily values. The analysis was performed with two dry/wet season TM images of the Etosha Pan. Both images were corrected with different water vapor values (Table III). The figure displays the difference in surface reflectance (in %) between the correction with day-specific values and the correction with climatically derived variants.

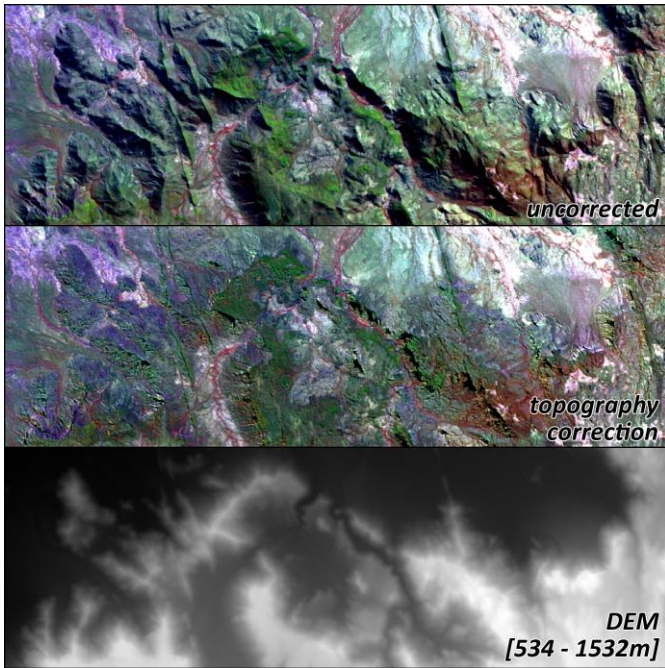


Fig. 11. Illustration of the topography correction for a mountainous Landsat-8 image in Huila, Angola (13.39°E, 15.86°S; 1000 m elevation range) under low illumination conditions (sun elevation: 41°, azimuth: 37°). Top: no topography correction; middle: topographically corrected; bottom: digital elevation model. The same stretch was applied to both images with R/G/B = NIR/SWIR1/red.

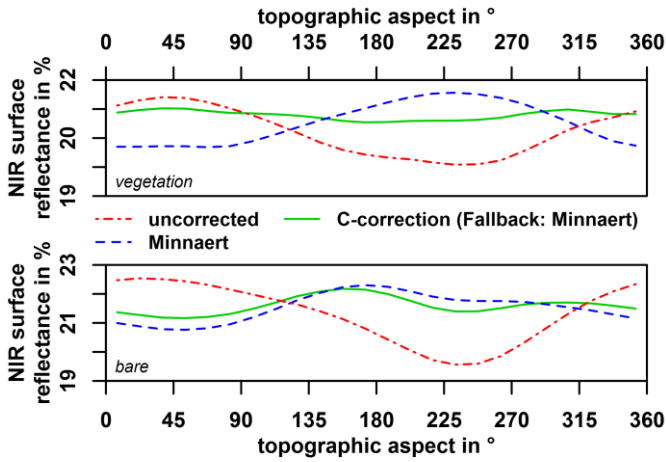


Fig. 12. Average aspect-dependent NIR reflectance for the image (full frame) shown in Fig. 11 for the vegetated (top) and bare (bottom) classes. The image was processed with the implemented C-Correction (green), the Minnaert correction (blue) and with disabled topography correction (red). Flat curves point to a successful topographic normalization.

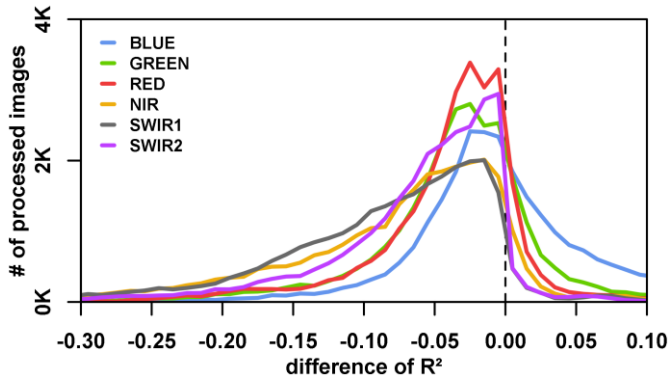


Fig. 13. Topographic correction evaluation. The histogram bin width was set to  $0.01 R^2$ . The correction was quantitatively evaluated by differencing the  $R^2$  of the relationship between  $\cos i$  and  $L$ , before and after the radiometric correction. A decrease in  $R^2$  points to a reduction of topographic effects.

## Tables

TABLE I  
ACQUIRED L1T LANDSAT IMAGES

	Number	Size	Start	End
<i>L4 TM</i>	225	0.03	10/22/1987	03/27/1993
<i>L5 TM</i>	24,204	3.34	04/12/1984	11/17/2011
<i>L7 ETM+ SLC-On</i>	7,166	1.83	06/30/1999	05/28/2003
<i>L7 ETM+ SLC-Off</i>	20,816	4.44	07/22/2003	12/31/2014
<i>L8 OLI</i>	6,332	5.53	04/11/2013	12/31/2014

Number, compressed data volume in TB and time range of the acquired images.

TABLE II  
MATHEMATICAL SYMBOLS

Symbol	Meaning
$\lambda$	Wavelength
$\rho, \rho^*, \rho_p$	Surface reflectance, at-satellite reflectance, path reflectance
$\langle \rho \rangle$	Background contribution to the apparent target reflectance
$\Theta_s, \Theta_v$	Sun and view zenith angles
$\Phi_s, \Phi_v$	Sun and view azimuth angles
$\mu_s, \mu_v$	Cosine of $\Theta_s$ and $\Theta_v$ , indicating downwelling and upwelling terms, respectively
$T, t_d, t_s$	Total transmittance, direct transmittance, scattered (diffuse) transmittance
$T_g, T_w$	Total gaseous transmittance, water vapor transmittance
$a_w, W, M$	Water vapor absorption coefficient, precipitable water vapor, relative air mass
$L^*, d, ESUN$	At-satellite radiance, Earth-Sun distance, mean solar exo-atmospheric irradiance
$A_{pl}, R, g$	Plane albedo, reflectance functions, asymmetry factor
$\tau, \tau_a, \tau_r$	Total, aerosol and molecular optical depth
$P, P_a, P_r$	Total, aerosol and molecular phase functions
$\Psi_-, s$	Backscattering angle, spherical albedo
$g_1, g_2, \alpha$	Asymmetry parameters of the TTHG function
$\rho_o^*, \rho_e^*, r$	Dark object reflectance, environment reflectance and target radius
$\rho_s, \rho_w, \rho_f$	Reference reflectance, volumic water reflectance, Fresnel reflectance
$a_0, a_1, a_2$	Coefficients of the logarithmic $\tau_a$ vs. $\lambda$ regression
$A, C$	Topographic A- and C-Factors
$i, \cos i$	Incidence angle, illumination angle
$b, m$	Intercept and slope of linear regression between $\cos i$ and $L^*$
$h, h_0$	Portion of the sky dome which contributes to the diffuse illumination ( $h_0$ : $h$ at $\cos i = 0$ )
$\Theta_n, \Phi_n$	Topographic slope and aspect

Definition of the mathematical symbols used in this paper.

TABLE III  
PRECIPITABLE WATER VALUES

Water vapor variant		Dry season 08/10/2008	Wet season 03/22/2009
<i>Daily value</i>	<i>DV</i>	1.02	2.32
Climatic average - 2 sd	CA--	0.29	1.96
Climatic average - 1 sd	CA-	0.70	2.85
Climatic average	CA	1.12	3.74
Climatic average + 1sd	CA+	1.53	4.63
Climatic average + 2sd	CA++	1.95	5.52

Precipitable water value variants (in cm) used for simulating the effect of climatic averages on the corrected reflectance of different land surface types.