

Multi-criteria evaluation of topographic correction methods

Ion Sola *, **María González-Audicana** and **Jesús Álvarez-Mozos**

Department of Projects and Rural Engineering. Campus Arrosadía. Public University of Navarre, 31006, Pamplona, SPAIN; E-Mail: maria.audicana@unavarra.es, jesus.alvarez@unavarra.es

* Author to whom correspondence should be addressed; E-Mail: ion.sola@unavarra.es ;

Tel.: +34-948-168-401; Fax: +34-948-169-148.

Abstract:

In the last decades, several topographic correction methods (TOC) have been proposed, but there is not an agreement on the best method. Furthermore, different evaluation criteria have been used in the past, and there is not any simple and objective evaluation procedure to measure the quality of the correction. Consequently, a multi-criteria analysis of widely used topographic correction methods is required that evaluates their performance over different sensors, terrain and temporal configurations. In this work, ten TOC methods were assessed using seven different evaluation strategies. The analysis was carried out for three SPOT5 images acquired over a mountainous area of northern Spain. The images had different acquisition dates and solar angles, so as to evaluate performance under varying illumination conditions. The results obtained showed that Statistic-Empirical method, C-Correction and Sun-Canopy-Sensor+C performed the best, and differences were minor when favorable illumination conditions were considered. For the seven tested evaluation strategies, interquartile range reduction of land covers or the comparison of sunlit and shaded slopes gave very similar results, whereas there were greater contrasts among other criteria.

Keywords: topographic correction; evaluation; comparison; multi-criteria

1. Introduction

The availability of Remote Sensing (RS) data has grown exponentially in the last few years. Accordingly, their applications in different fields, such as land-use/land-cover mapping, vegetation cover monitoring and change detection, retrieval of land cover biophysical parameters, agriculture, or and risk assessment, have increased. To retrieve accurate information from RS scenes, it is necessary to perform pre-processing operations to correct distortions that are inherent to any image acquisition process. The radiance recorded by an optical satellite sensor is affected by several factors, which include sensor and system induced errors, atmosphere, topography and solar illumination angles, that need to be resolved by correction methods (Balthazar et al. 2012; Reese and Olsson 2011; Veraverbeke et al. 2010).

In non-flat areas, illumination correction, also known as topographic correction, is an important step in pre-processing high-resolution RS data (Tan et al., 2013). The radiance detected by sensors can vary significantly depending on the reflectance of land covers and on the slope and aspect of the area (Riaño et al. 2003). The objective of topographic correction is thus to compensate for the differences in solar irradiance between areas with differing slope and aspect and, ultimately, to obtain the radiance values the sensor would have obtained in the case of a perfectly flat surface.

Numerous topographic correction (TOC) methods have been developed to correct topographic effects on the radiance measured by satellites. In those TOC algorithms, the illumination conditions for each pixel are normally estimated using the cosine of the solar incidence angle, $\cos\gamma_i$, based on the solar geometry (i.e., sun position), and the slope and aspect of these pixels (Eq. 1).

$$\cos\gamma_i = \cos\beta\cos\theta_s + \sin\beta\sin\theta_s\cos(\varphi_n - \varphi_s) \quad (1)$$

where β is the slope angle, φ_n the aspect angle, θ_s the solar zenith angle, and φ_s the solar azimuth angle. Both β and φ_n are pixel-based values computed from the digital elevation model (DEM).

According to Balthazar et al. (2012), TOC methods can be grouped into three subcategories: empirical methods, semi-empirical methods, and physically based methods. The first group consists of simple empirical methods, such as band ratioing, that do not require additional ancillary data (Civco 1989;

Ekstrand 1996). These procedures are based on the assumption that radiance values vary, due to topography proportionally in all bands. They are easily implemented, but their output does not have a physical meaning (Blesius and Weirich 2005). The second category groups semi-empirical approaches that need a DEM to represent the solar irradiance differences between slopes of the area to be corrected (Ghasemi et al. 2013; Hantson and Chuvieco 2011; Law and Nichol 2004; Lu et al. 2008; Soenen et al. 2005; Twele and Erasmi 2005). Finally, physically based TOC methods model the full path of radiance through the atmosphere to the target object and backwards (Gu and Gillespie 1998; Sandmeier and Itten 1997; Soenen et al. 2005; Kane et al. 2008; Kobayashi and Sanga-Ngoie 2008; Soenen et al. 2009; Zhang and Gao, 2011).

Due to their relative simplicity and accuracy, semi-empirical TOC algorithms are most often used. Those methods require fitting empirical parameters for the correction of each spectral band, and these parameters can be fit globally to the whole scene or individually for each land cover type. In the latter case, the scene needs to first be stratified or pre-classified on different strata that are corrected individually to achieve a more precise estimation of the correction parameters and, thus a stronger reduction of the topographic effect (Bishop and Colby 2002; Bishop et al. 2003; Szantoi and Simonetti 2013). On the other hand, Reese and Olsson (2011) recommended different sampling strategies for the determination of the correction parameters.

A number of TOC algorithms were proposed in the past. However, there is not an agreement on their performance, as authors used different evaluation procedures that are not easily comparable (Civco 1989; Lu et al. 2008; Richter et al. 2009; Hantson and Chuvieco 2011). Most of these evaluation criteria are summarized in Table 1, including the pros and cons for each.

Table 1. TOC evaluation procedures used in the literature

Evaluation technique	References	Pros	Cons
Visual	Zhang et al. 2015; Singh et al. 2015; Shepherd et al. 2014	Direct indicator. No ancillary data required	Subjective
Correlation $\cos^2\theta_i - L_{sen}$	Gao et al. 2014; Vanonckelen et al. 2014; Moreira and Valeriano 2014	Easy to compute, quantitative, no ancillary data required	Residual correlation if slope orientation determines land cover

Spectral stability of land cover	Ghasemi et al. 2013; Goslee 2012; Moreira and Valeriano 2014	Detects possible biases introduced by the correction	Measure of stability, not a proper TOC evaluation technique
Reduction of land cover variability	Gao et al. 2014; Moreira and Valeriano 2014; Fan et al. 2014	Objective. Analysis on different land covers	Need of reliable information on land covers
Classification accuracy	Hoshikawa and Umezaki 2014; Vanonckelen et al. 2013; Füreder 2010	Assesses the effects of correction on thematic products derived from RS. Quantitative. Analysis on different land covers	Depends on the quality of ground truth data, classification algorithm, etc.
Difference North-facing/South-facing pixels of same land cover	Civco 1989; Fan et al. 2014; Vicente-Serrano et al. 2008; Notarnicola et al. 2014	Good correction, under-correction or over-correction detected	Only selected land cover is assessed
Presence of outliers	Balthazar et al. 2012	Quantifies the percentage of pixels each TOC could correct	Not a proper TOC evaluation technique
Similarity to SH (synthetic images)	Sola et al. 2014	Comparison between corrected image and ideal situation	Need to generate a pair of synthetic images

Visual evaluation of the removal of the topographic effect is, generally, the first indicator of the quality of the correction (Szantoi and Simonetti 2013; Tan et al. 2013; Shepherd et al. 2014). Nevertheless, it is imperative to quantitatively evaluate the results to select the best TOC method (Balthazar et al. 2012).

One of the most widely used quantitative evaluation procedures is the decrease in the dependence between $\cos\gamma_i$ and the radiance of each spectral band after TOC. This can be measured through the decrease in the correlation coefficient (Gao et al. 2014), in the slope of linear regression (Vanonckelen et al. 2014) or both (Gao et al. 2014; Goslee 2012; Szantoi and Simonetti 2013; Zhang and Gao 2009a). A lower dependence between the incidence angle and the corrected radiance indicates a better topographic correction. Obviously, this is not valid in areas where slope and aspect influence land cover distribution. In such areas, residual correlation between radiance and $\cos\gamma_i$ should be expected, even after a successful topographic correction (Hantson and Chuvieco 2011).

Quantification of the reduction of land cover class variability after TOC can be measured through the standard deviation (SD) or the coefficient of variation (CV) of the reflectance within each land cover class (Fan et al. 2014; Gao et al. 2014; Moreira and Valeriano 2014). Land cover homogeneity should theoretically increase after correction, since the intraclass radiometric variations caused by topography are minimized. This criterion is arguably the most objective and quantitatively measurable evaluation

procedure. However, it is only applicable to cases where *a priori* knowledge of land cover distributions is available. When this information is missing, broad land cover classes based on vegetation index thresholds can be created, i.e., vegetation or no vegetation (Szantoi and Simonetti 2013), forest or pastures (Goslee 2012; Lu et al. 2008) or samples of representative land covers (Gao et al. 2014).

Alternatively, the improvement of the accuracy of land cover classification has been considered to be a good measure of the effects of preprocessing (i.e., atmospheric and/or topographic correction) of satellite imagery (Hoshikawa and Umezaki 2014; Moreira and Valeriano 2014; Vanonckelen et al. 2013). A classification based on topographically corrected images should ideally yield a higher accuracy than one using uncorrected data. A similar approach is to evaluate the improvement in biophysical parameter retrievals (Ekstrand 1996; Tokola et al. 2001) or in change detection accuracy (Tan et al. 2013; Vanonckelen et al. 2015). However, these assessments entail their own uncertainties in classification, change detection and retrieval algorithms, and this makes it difficult to directly relate their results to TOC quality (Hoshikawa and Umezaki 2014; Sola et al. 2014).

Some authors previously evaluated TOC methods by extracting samples of certain land covers (i.e., forest or pasture) for north-facing (shaded) and south-facing (sunlit) slopes (Fan et al. 2014; Notarnicola et al. 2014; Goslee 2012; Vicente-Serrano et al. 2008; Riaño et al. 2003). The samples from the north and south slopes were compared before and after TOC through their mean radiance difference or the Root Mean Square Deviation (RMSD) (Schulmann et al. 2015). An ideal TOC would make north and south samples more similar. However, this criterion assumes that land cover characteristics (i.e., forest density) are the same in south-facing and north-facing slopes, although this is not always the case.

An ideal topographic correction should not substantially change the spectral characteristics of land covers (i.e., mean radiance value) (Riaño et al. 2003; Richter et al. 2009). A substantial change would mean that the TOC introduced a bias that violated the spectral stability. This evaluation procedure has been used image-wide (Ghasemi et al. 2013; Goslee 2012; Zhang and Gao 2009a) or stratified by land cover classes (Goslee 2012; Moreira and Valeriano 2014). However, it cannot strictly be considered a

criterion to assess the performance of the correction, but rather as a measure of stability (Baraldi et al. 2010).

Another evaluation criterion is the proportion of statistical outliers generated after TOC. In extreme conditions, some TOC methods can fail and result in abnormally high or low radiance values, i.e., outliers. Unfortunately, this issue is not thoroughly analyzed in the literature. Balthazar et al. (2012) defined an expected range of radiance values and considered outliers to be all the pixels beyond that threshold. Using this description, they then tested the amount of outliers generated by a given TOC.

Finally, Sola et al. (2014) proposed the use of simulated scenarios based on synthetic images to evaluate topographic correction algorithms in a more thorough and objective manner. Synthetic images represent the radiance an optical sensor would receive under specific geometric and temporal acquisition conditions and assuming a certain land cover type. They can be used to select the best performing TOC for each particular situation (e.g., solar angles, spatial resolution, etc.). In particular, using a Synthetic Horizontal image (SH) for a point of reference for the comparative analysis of images obtained after correcting a Synthetic Real image (SR) with different algorithms provided a sound, objective and clear method for the quantitative assessment of those algorithms.

To summarize, many TOC algorithms have been proposed, but most of them have not been fully and consistently evaluated since most studies only considered a limited set of illumination conditions and only one or two evaluation criteria. Therefore, the objective of this paper is to perform a multi-criteria analysis of different topographic correction methods, providing a guideline of use of TOC methods under different conditions, including advantages and shortcomings of each TOC algorithm.

2. Material and methods

2.1. Study area

Three different image acquisition dates were analyzed that corresponded to a single study site in the Pyrenees mountain range, Spain (Fig. 1). For these images, the relief is rough and the valleys are oriented in different directions. The land coverage is primarily coniferous, broadleaf and mixed forest, comprising

60% of the area, whereas 23% of the land coverage is agricultural. Other classes, such as shrubs, rocks or unproductive areas are less frequent. The mean altitude of the area is 380 m, ranging from 67 to 1370 m, and the maximum and mean slopes are 75.6° and 22.2°, respectively. These figures indicate a rough topography for the study area.

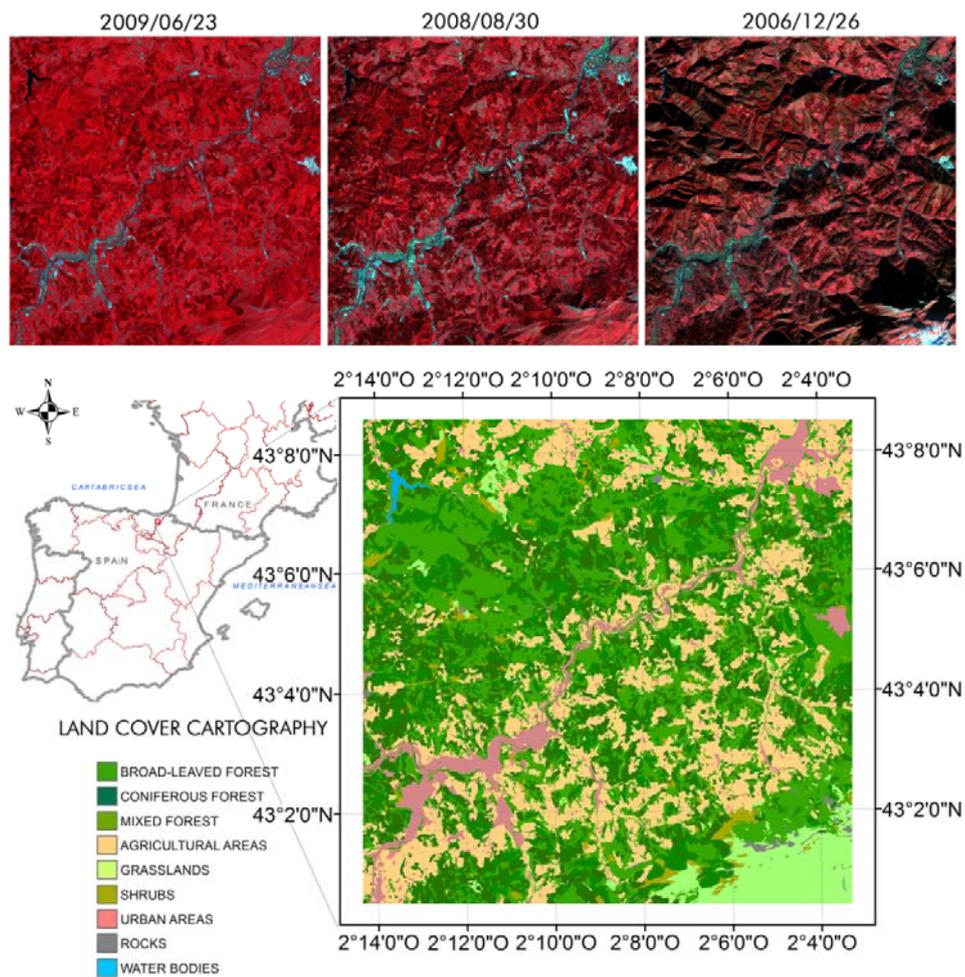


Fig. 1. The study area, located in northern Spain. SPOT5 imagery of June, August and December. Displayed bands: Near Infrared, Red, Green (RGB).

2.2. Data acquisition and Processing

The area corresponded to a 15x15 km subscene of three SPOT5 images acquired under different temporal configurations (Table 2). SPOT5 multispectral scenes are composed by four spectral bands (i.e., green: 0.50 – 0.59 μm , red: 0.61 – 0.68 μm , NIR (near infrared): 0.78 – 0.89 μm and SWIR (short-wave infrared): 1.58 – 1.75 μm) with a spatial resolution of 10 m. The SPOT5 scenes were orthorectified and

converted from digital numbers to top of atmosphere radiance ($\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$) by using the gain and offset values provided in the metadata file for each spectral band.

Table 2. Configuration of SPOT5 scenes for the different images

SPOT5 grid reference	Sun elevation (°)	Sun azimuth (°)	Acquisition date	Acquisition time
35 263	65.80	139.05	2009/06/23	11:03
35 263	53.53	155.02	2008/08/30	11:11
35 263	21.91	165.15	2006/12/26	11:07

The land-use/land-cover (LU/LC) information of the area was obtained from specific cartography of the region of Gipuzkoa (Gobierno Vasco 2007; 2010). Using these maps, the land cover classes of the study area were reclassified into nine broad classes. The percentage of surface covered by each of the nine land cover classes is shown in Table 3.

Table 3. Percentages of area of each land cover for the test site

Land cover class	Code	Area coverage (%)
Broad-leaved forest	BROAD	20.29
Coniferous forest	CONIF	28.50
Mixed forest	MIXED	11.99
Agricultural areas	AGRIC	23.62
Grasslands	GRASS	5.97
Shrubs	SHRUB	3.36
Unproductive	UNPROD	5.25
Rocks	ROCK	0.50
Water bodies	WATER	0.52
TOTAL		100.00

All the topographic parameters necessary to apply each TOC method were retrieved from the original digital elevation model (DEM), at 5 m resolution, and then resampled to 10 m to match the spatial resolution of SPOT5. This DEM, provided by the Spanish National Geographic Institute (IGN), was generated in 2010 from cubic convolution of LIDAR point cloud, with a density of 0.5 points m^{-2} .

2.3. Selected topographic correction algorithms

After a comprehensive literature review, this work selected ten topographic correction algorithms based on their reported popularity and performance that are detailed in Table 4 (Richter 2009; Zhang et al. 2011; Vanonckelen et al. 2013; Moreira and Valeriano 2014). Table 4 excluded algorithms widely tested but with proven poor performance, such as the Lambertian methods of Cosine and Sun-Canopy-Sensor

(SCS). Similarly, algorithms that performed well but had been tested only on a single study were discarded (e.g., Kobayashi and Sanga-Ngoie 2009; Zhang and Gao 2011), since further research was deemed necessary to fully corroborate the results.

Table 4. Expressions of TOC algorithms analyzed

TOC	Expression	Authors
C-Correction (CC)	$L_{corr,\lambda} = L_\lambda \frac{\cos \vartheta_s + C_\lambda}{\cos \gamma_i + C_\lambda}$	Teillet et al. (1982)
Smoothed C-Correction (sCC3)	$L_{corr,\lambda} = L_\lambda \frac{\cos \vartheta_s + C_\lambda}{\cos \gamma'_i + C_\lambda}$	Riaño et al. (2003)
SCS+C (SCSC)	$L_{corr,\lambda} = L_\lambda \frac{\cos \beta \cos \vartheta_s + C_\lambda}{\cos \gamma_i + C_\lambda}$	Soenen et al. (2005)
Statistical-Empirical (SE)	$L_{corr,\lambda} = L_\lambda - (A \cos \gamma_i + B) + \bar{L}_\lambda$	Teillet et al. (1982)
Minnaert (MIN)	$L_{corr,\lambda} = L_\lambda \left(\frac{\cos \vartheta_s}{\cos \gamma_i} \right)^k$	Minnaert (1941)
Enhanced Minnaert (EMIN)	$L_{corr,\lambda} = L_\lambda \cos \beta \left(\frac{\cos \vartheta_s}{\cos \gamma_i \cos \beta} \right)^k$	Smith et al. (1980)
Pixel-based Minnaert (PBM)	$L_{corr,\lambda} = L_\lambda \frac{\cos \beta}{(\cos \beta \cos \gamma_i)^{k^*}}$	Lu et al. (2008)
Modified Minnaert (MM)	$L_{LAMB,\lambda} = L_\lambda \frac{\cos \vartheta_s}{\cos \gamma_i}$ $L_{corr,\lambda} = L_{LAMB,\lambda} \left(\frac{\cos \gamma_i}{\cos \beta_i} \right)^b = L_{LAMB,\lambda} \cdot G$	Richter (1998)
Two stage normalization (2SN)	1F: $L_{corr,\lambda} = L_\lambda + \left[L_\lambda \cdot \left(\frac{\mu_k - X_{ij}}{\mu_k} \right) \right]$ 2F: $L_{corr,\lambda} = L_\lambda + \left[L_\lambda \cdot \left(\frac{\mu_k - X_{ij}}{\mu_k} \right) \cdot C_{2sn} \right]$	Civco (1989)
Slope-Matching (SM)	1F: $L_{corr,\lambda} = L_\lambda + (L_{MAX,\lambda} - L_{MIN,\lambda}) \left(\frac{\mu_w - X_{ij}}{\mu_w} \right)$ 2F: $L_{corr,\lambda} = L_\lambda + (L_{MAX,\lambda} - L_{MIN,\lambda}) \left(\frac{\mu_w - X_{ij}}{\mu_w} \right) \cdot C_{sm}$	Nichol et al. (2006)

Table 5. Nomenclature

Symbol	Explanation
L_λ	Original radiance for band λ
$L_{corr,\lambda}$	Corrected radiance for band λ
ϑ_s	Solar zenith angle
γ_i	Solar incidence angle
C_λ	Empirical parameter used by the CC, sCC3 and SCSC methods
β	Terrain slope
γ'_i	Solar incidence angle obtained from smoothed slope
\bar{L}_λ	Mean radiance of the image for band λ
k	Constant of Minnaert
k^*	Pixel-based k obtained stratifying the image by terrain slope ranges
$L_{LAMB,\lambda}$	Radiance of band λ after a Lambertian correction (i.e., Cosine method)
G	Correction parameter introduced to avoid the overcorrection of poorly illuminated pixels
μ_k	Mean value of the scaled (0–255) $\cos \gamma_i$ for the main cover type
X_{ij}	Scaled (0–255) cosine of γ_i for pixel ij for the main cover type
C_{2sn}	Empirically-derived calibration parameter
$L_{max,\lambda}$ and $L_{min,\lambda}$	Maximum and minimum radiance value for main cover type
μ_w	Mean value of the scaled $\cos \gamma_i$ for the main cover type on sunny slopes
C_{sm}	Modified correction parameter

Some of these TOC methods can be easily implemented, but in others, the authors encourage users to adjust or tune the correction parameters so as to obtain optimum results for their specific dataset. For instance, after preliminary tests with different values, a smooth factor of 3 was selected in sCC3 to calculate the smoothed β , as it provided the best results. Furthermore, following the suggestions of Lu et al. (2008), a second degree polynomic equation was used as the best fit to the regression between Minnaert k and β in PBM, while in MM correction the “strong” correction option was selected and the lower bound g was set to 0.1 after testing. Finally, in 2SN and SM the sunny and shady slopes were automatically masked from the image of aspect generated from the DEM, and only the main land cover (i.e., broad-leaved forest) was considered in the calculation of the correction parameters C_{2SN} and C_{SM} .

In the calculation of TOC parameters, flat pixels were excluded, that is, pixels where $\beta < 5^\circ$. Pixels occluded by surrounding topography, as in pixels cast in shadow, or self-shadowed pixels (i.e., pixels where $\cos\gamma_i < 0$) were also masked out for that purpose. However, TOC algorithms were later applied image-wide with no mask used. Additionally, bidirectional reflectance distribution function (BRDF) was not considered, as this information is rarely, if ever, available, and is extremely difficult to obtain for regional studies (Goslee 2012). Moreover, for high spatial resolution sensors with a small field of view, the solar viewing geometry is approximately constant for flat surfaces. Therefore, BRDF variations for a certain cover type due to geometry changes are small (Richter, 1998). Finally, no corrections were made in the view direction as little impact of sensor viewing angle on reflectance in temperate latitudes has been reported in the literature (Nagol et al. 2015).

2.4. Evaluation strategies

2.4.1. TOC ranking based on multi-criteria

To have a general overview of the performance of the TOC algorithms tested over the three images, the results obtained through the seven different evaluation procedures were standardized and assembled in a multi-criteria ranking of TOC methods. To do this, the results of each criterion were standardized by

subtracting the mean from each feature and dividing that value by its standard deviation (SD) so that standardized values above zero refer to TOC methods performing higher than the average, and values below zero refer to poor performers. An ensemble criterion was then obtained as the mean of the seven standardized criteria.

2.4.2. Visual analysis

Visual evaluation of the images was independently performed by 10 Remote Sensing (RS) scientists and engineers who were asked to compare pairs of corrected images with no information about the TOC method used. Overall, there were 1350 pairwise comparisons derived from 45 comparisons of three images by each of the ten participants, and each method was compared a total of 90 times (9 comparisons by each of the 10 participants) for each image. Based on those pairwise comparisons, a ranking of TOC methods was built, based on the number of times each TOC method was considered visually better than the other method in the pairwise comparison.

2.4.3. Correlation analysis

Quantification of the reduction of the dependence between $\cos\gamma_i$ and the radiance of each spectral band after image correction was computed by fitting a linear regression, and analyzing its slope and correlation coefficient (R).

2.4.4. Stability of land cover radiometry

In this work, the study area was classified into nine broad land cover classes based on local cartography (Table 3). The relative difference of median radiance (RDMR) was calculated for each land cover by comparing the median radiance of land covers before and after the correction (Eq. 2). This method is more reliable than considering the mean, which has more outliers. Subsequently, the area-weighted RDMR was computed for each spectral band. An ideal TOC method should be stable and result in RDMR values close to zero.

$$RDMR = \frac{(\tilde{L}_{corr,\lambda} - \tilde{L}_\lambda) \cdot 100}{\tilde{L}_\lambda} \quad (2)$$

where $\tilde{L}_{corr,\lambda}$ and \tilde{L}_λ are the topographically corrected and uncorrected median radiances of each land cover class, respectively.

2.4.5. Intraclass interquartile range reduction

The reduction of intraclass variability after correction can be measured using different statistical measures (e.g., variance or SD). The major drawback of this procedure is that the presence of outliers, generated by TOC algorithms on unfavorable conditions, could produce distorted SD results. This effect could instead be minimized by using the interquartile range (IQR), the difference between the upper (Q3) and lower (Q1) quartiles, that is much less sensitive to outliers. Consequently, this modification was implemented in this study. To calculate IQR, images were stratified in the nine land cover classes explained in section 2.2 (Table 3). The main statistics for minimum, Q1, median, Q3 and maximum radiance of each land cover class were calculated before and after correction. The area-weighted IQR reduction (Eq. 3) was computed after.

$$IQR_{reduction} = 100 - \frac{IQR_{corr,\lambda} \cdot 100}{IQR_\lambda} \quad (3)$$

where $IQR_{corr,\lambda}$ and IQR_λ are the topographically corrected and uncorrected IQR of each land cover class, respectively.

Ideally, IQR should decrease after TOC. An advantage of this criterion is that it allows a per class evaluation of the influence of TOC on different land covers that might be more or less affected by topography (Table 1).

2.4.6. Comparison of coniferous forest radiometry between sunlit and shaded slopes

In this work, 2000 pixels of coniferous forest were randomly selected for each image. Half of them were located in sunlit slopes (i.e., slope aspect = solar azimuth $\pm 10^\circ$) and the other half in shaded slopes. The radiance difference between sunlit and shaded slopes was then computed.

2.4.7. Percentage of outliers

Previously, most evaluation of TOC algorithms were carried out on favorable conditions (Zhang and Gao 2009b; Richter et al. 2009; Vanonckelen et al. 2014), and the few problematic pixels located on weakly illuminated slopes ($\cos\gamma_i \leq 0$) were normally masked and excluded from evaluation. Consequently, it has been deemed better to leave those pixels uncorrected than to correct them (Baraldi et al. 2010). However, it seems interesting to assess the performance of TOC algorithms on the whole image, without including any masks, and thus it is necessary to correct even those extreme pixels located on weakly illuminated slopes. When a TOC algorithm fails at correcting those pixels, however, it may create outliers, and a method that generates too many outliers cannot be recommended. In this work, pixels of topographically corrected scenes with radiance values higher than the maximum original radiance or lower than the minimum were considered to be statistical outliers, and their percentage in the image was calculated.

2.4.8. Synthetic images

Sola et al. (2014) proposed a novel evaluation strategy of the performance of TOC based on the comparison between Synthetic Horizontal images (SH) and topographically corrected Synthetic Real images (SR_{TOC}) (Fig. 2). The latter were generated considering the real DEM of an area, while the former were obtained by running the simulation model over an ideal flat DEM. SR can be corrected with the TOC method being evaluated, whereas SH corresponds to the ideal situation of the at-sensor radiance with no influence of topography. The SH image was used as a reference to compare how close the corrected SR image was to the ideal correction by using image quality indexes such as the Structural SIMilarity index (SSIM) proposed by Wang et al. (2004). The mean SSIM index (MSSIM) can be used as an overall image quality metric to quantitatively rank the performance of TOC methods. It is also useful to apply the SSIM index locally rather than globally (Wang et al. 2005) to generate a SSIM map from the local statistics within an 11 x 11 circular-symmetric Gaussian weighting function, which moves pixel-by-pixel over the

image (Brunet et al. 2012). The SSIM maps generated for each topographically corrected image show the performance of the correction pixel-by-pixel, and the combination of SSIM maps and MSSIM index provides a useful tool to decide the best TOC method for each scene.

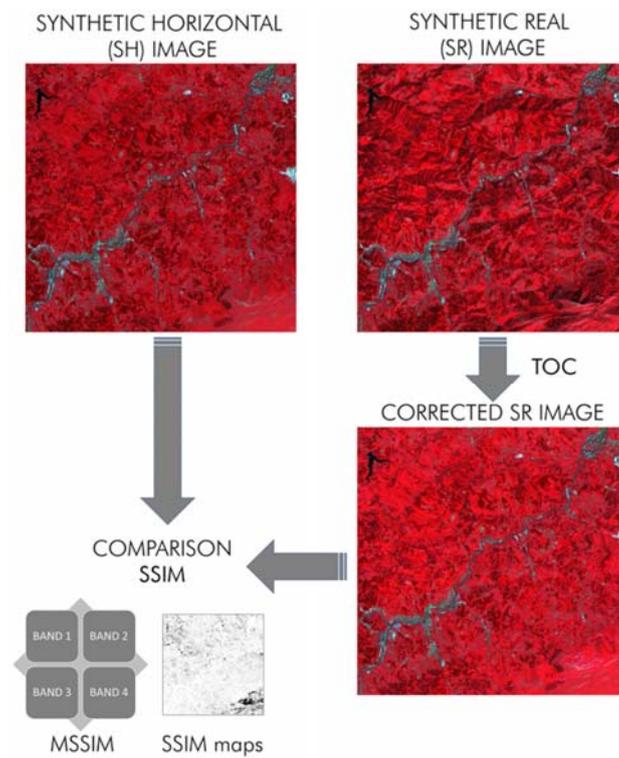


Fig. 2. Scheme of the novel evaluation strategy of the performance of TOC methods based on synthetic images

3. Results

Multi-criteria analysis of different topographic correction (TOC) methods was performed to evaluate their advantages and disadvantages, with the aim of providing a guideline for using TOC methods under various conditions.

3.1. Visual analysis

Visual analysis of three distinct images was carried out to allow direct pair-wise comparison of the different TOC methods. In Fig. 3, false color composite (Near Infrared, Red and Green) of corrected

images from August are shown. The topographic effect was strong (Fig. 3a), and the differences between TOC methods can clearly be seen (Fig. 3b-k).

In this particular case, CC and SE visually performed very well, while 2SN, SM and SCSC corrected most of the shadowed areas. In turn, sCC3 showed a corrected image where shadows were only partially removed. On the other hand, methods based on Minnaert (i.e., MIN, EMIN and PBM) successfully corrected the differences on radiance introduced by the topography in most of the scenes, although visual evaluation was negatively affected by the presence of outliers in some areas. Lastly, MM failed at correcting the topography, and yielded unreliable results.

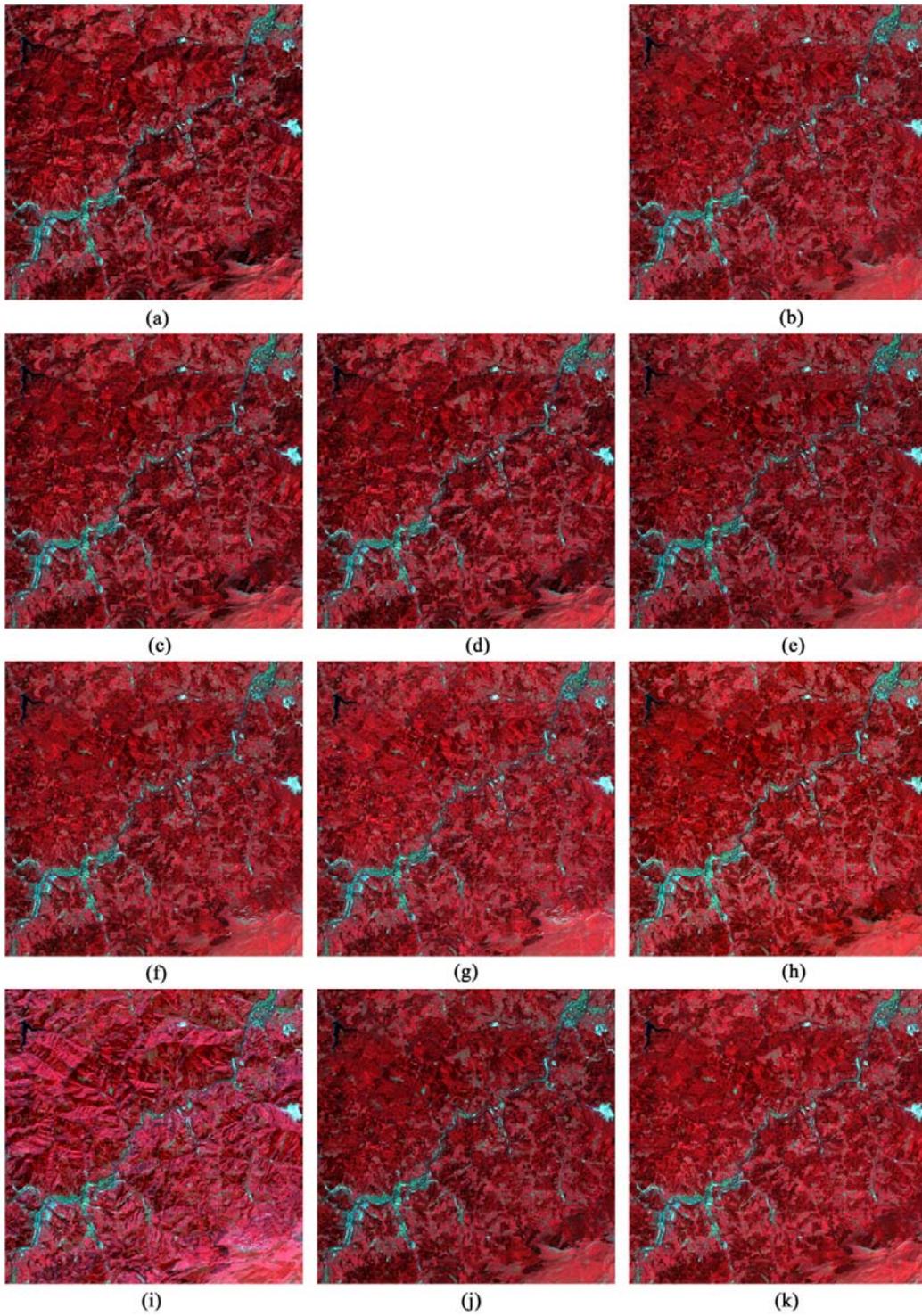


Fig. 3. Original (a) and corrected images of August with TOC methods (b) CC (c) sCC3 (d) SCSC (e) SE (f) MIN (g) EMIN (h) PBM (i) MM (j) 2SN (k) SM

In contrast to the images from August, there were minor topographical effects in the images from June (Fig. S1a). Therefore, the differences between TOC methods were slight (Fig. S1b-k). The image of December, however, corresponded to an extreme scenario with a strong topographic effect and shadows in most zones of the study area (Fig. S2). In this case, none of the tested TOC algorithms (Figs. S2b-k) succeeded in completely correcting the tremendous topographic effect on the original image (Fig. S2a). The shadowed areas corresponding to pixels where no direct irradiance is impinging on the surface were only partially corrected with SE and SM, whereas outliers in this area were clearly noticeable for the other methods. Minnaert based methods (MIN, EMIN, PBM and MM) also did not extract reliable spectral information from those areas, while MM had a very poor performance.

Ranking of methods according to visual analysis provided preliminary results on the quality of each TOC method, although in some cases the evaluation was limited by the slight differences among corrected images to compare.

Table 6. Number of times each TOC method was superior (out of 90) and average ranking of methods by visual analysis for the 3 images (1= best, 10 = worst)

TOC	Number of times each TOC is superior			Visual ranking			
	June	August	December	June	August	December	Overall
CC	61	72	49	4	2	5	5
sCC3	10	12	62	9	9	3	7
SCSC	72	56	49	3	4	5	3
SE	88	90	86	1	1	1	1
MIN	32	22	26	6	7	7	6
EMIN	30	32	26	7	8	7	8
PBM	24	44	9	8	6	9	9
MM	0	0	0	10	10	10	10
2SN	75	68	59	2	3	4	2
SM	58	54	84	5	5	2	4

As seen in Table 6, the ranking of methods varied among the different dates, and the methods that ranked in the first position were different for each date. SE, 2SN and SCSC, in descending order, performed the best in June, with only minor differences among them. In August, SE again ranked first, with results significantly better than CC, 2SN and SCSC. Finally, in December, there were greater differences among methods, with SE and SM showing superior performance. Those methods were capable of extracting spectral information even from shaded slopes. A moderate performance was obtained by CC,

SCSC and sCC3, with similar results among the three methods. MIN, EMIN and PBM gave intermediate results and ranked very similar for the three dates, with some problems of over-correction in weakly illuminated slopes (Fig. 3f, 3g and 3h). Conversely, MM ranked last for the three dates analyzed, showing a poor correction of the topographic effect.

3.2. Correlation analysis

Further investigation of the different TOC methods was done via correlation analysis. The slope and correlation coefficient (R) of the regression between $\cos\gamma_i$ and the radiance of each spectral band are shown in Fig. 4 for the original image (left) and for the ten topographically corrected images for each of the three dates. The ideal correction method should lead to R and slope values of zero.

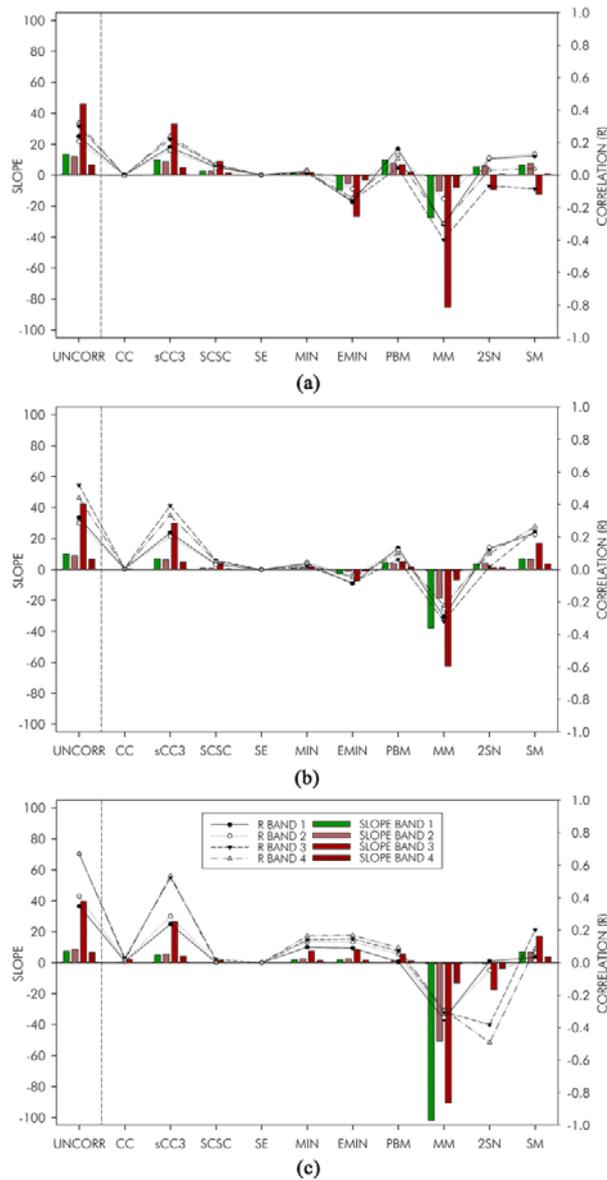


Fig. 4. Slope and correlation coefficient of regression between $\cos\gamma_i$ and the radiance of each spectral band after TOC. **(a)** June **(b)** August **(c)** December. The uncorrected image (UNCORR.) is represented as a reference. The closer the slope and R are from 0, the better.

As seen in Fig. 4, positive slopes and high R-values were observed for the original images, due to the topographic effect, especially for the NIR band. According to this criterion, SE, CC, SCSC and MIN methods were the best methods for reducing the dependence of spectral radiance on illumination in the three images studied, with minimum slopes and R-values. EMIN, PBM, 2SN and SM methods also performed well, but showed a negative correlation for some bands and a positive correlation for others.

The former indicates over-correction, whereas the latter is a sign of incomplete removal of the topographic effect. Finally, sCC3 only partially reduced the correlation between radiance and illumination, while MM overcorrected the original image, showing a negative slope.

3.3. Radiometric stability of land covers

The stability of land covers, measured through the RDMR, is shown in Fig. 5. The original radiometry of land covers was strongly modified by some TOC methods, especially PBM, with a RDMR of land covers up to 80% in the hardest scenario for the infrared bands.

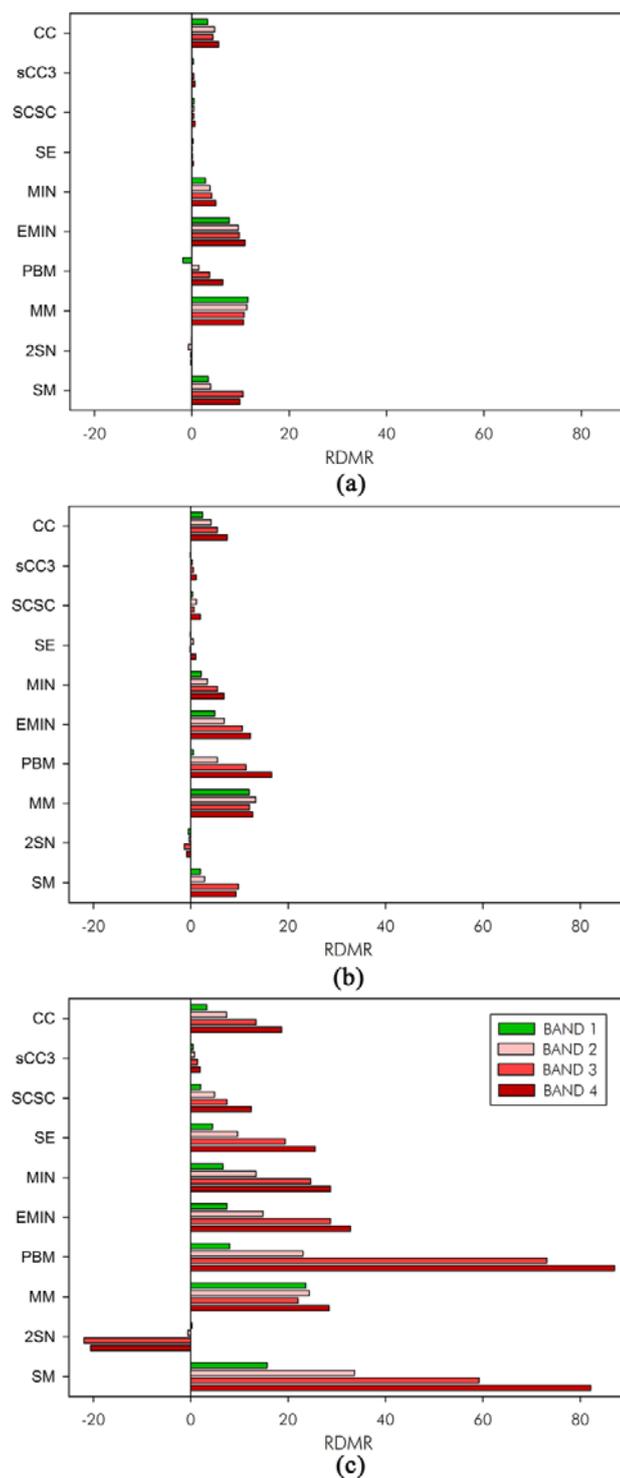


Fig. 5. Radiometric stability of land covers represented by the RDMR after TOC. (a) June (b) August (c) December. The smaller RDMR the better.

It is noticeable that radiance variations were positive for most methods, with 2SN being the only TOC method with negative RDMR values, especially on IR bands in December. Variations were minor in June and August, with RDMR values below 10% in most cases. As expected, results were generally worse in December. Overall, sCC3 was the TOC method that altered the radiometry the least, with RDMR below 5%, even in December, followed by SCSC, CC and SE.

3.4. Intraclass IQR reduction

Intraclass interquartile range (IQR) reduction criterion depends on the amount of topographic effects to correct each case. Therefore, IQR reductions are expected to be lower in June than in August or December because the topographic effect was less pronounced.

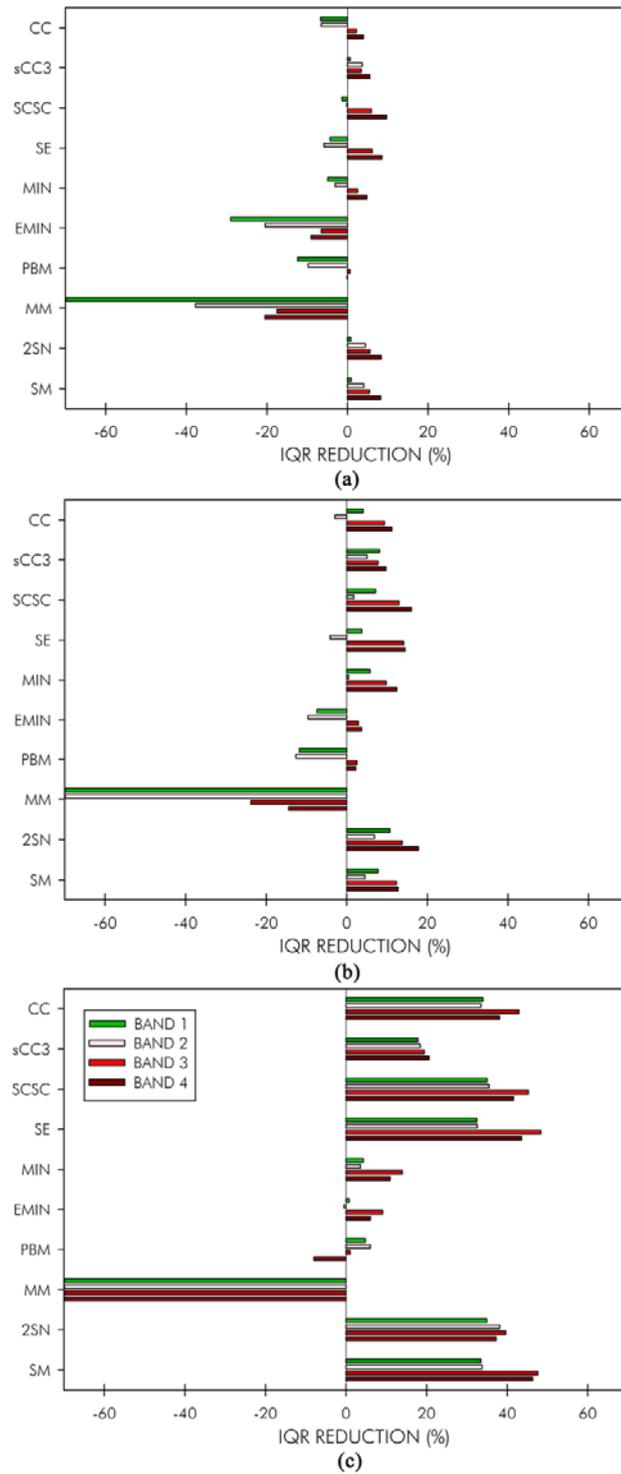


Fig. 6. Intra-class IQR reduction. Weighted average of 8 land covers. **(a)** June **(b)** August **(c)** December. The biggest positive IQR reduction (%) the better. The values of MM are out of range in some cases.

This is clearly visible in Fig. 6, where the area-weighted IQR reduction (Eq. 3) in June was lower than 10% for sCC3, 2SN and SM, whereas CC, SCSC, SE and MIN showed slight negative IQR reduction in the visible bands. The other three methods increased the original IQR to a greater extent, again especially in the visible bands. Negative IQR reduction rates indicate over-correction, resulting in more heterogeneous land covers after TOC. On the contrary, in August the degree of IQR reduction increased in the best performing TOC algorithms, SCSC, 2SN and SE, whereas EMIN, PBM and especially MM showed a poor performance with negative IQR reduction in the visible bands. In December, CC, sCC3, SCSC, SE, 2SN and SM reduced the original IQR of land covers in all of the spectral bands in a range of 20% to 50%. In contrast, methods such as MM, MIN, EMIN and PBM performed significantly worse.

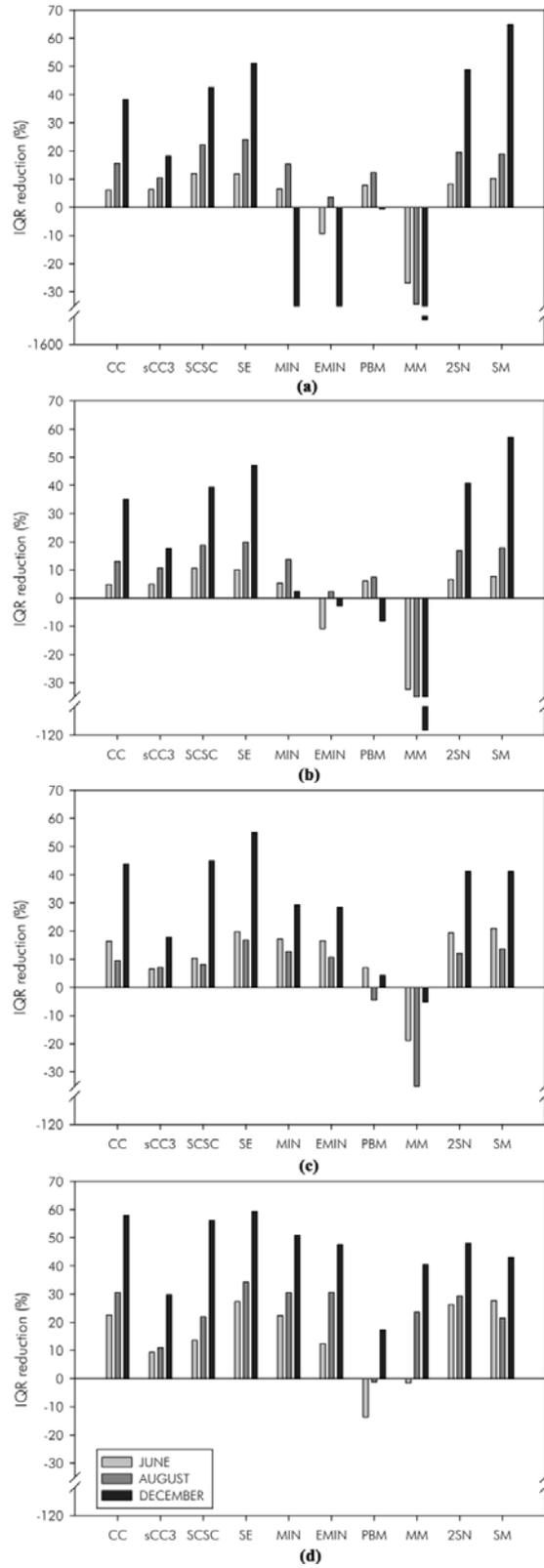


Fig. 7. Intra-class IQR reduction of band 3 for different land covers. (a) Broad-leaved forest (b) Mixed forest (c) Grasslands (d) Rocks. The biggest positive IQR reduction (%) the better.

The IQR reduction was also analyzed per land cover to independently evaluate the performance of TOC methods on each land cover. Fig. 7 shows the results of band 3, which was particularly affected by the topographic effect, for four predominant land covers of the study site on the three dates considered. The four land covers showed a similar trend of increase in the IQR reduction in the image from December for the best performing method. The greatest IQR reduction was achieved for rocks in most TOC methods for all the spectral bands. Also, broad-leaved forest, mixed forest and grasslands showed a strong IQR reduction in most TOC methods, especially for December. As for the algorithms, some methods, such as SE, SM or SCSC, strongly reduced the original IQR of the four displayed land covers on the three dates. Other methods, like CC, sCC3 and 2SN, did not reduce IQR much, but never came into negative values. Finally, MIN, EMIN, PBM and especially MM gave poor results.

3.5. Comparison of conifer forests radiometry between sunlit and shaded slopes

In the uncorrected scenes, the radiance difference between conifer forests located on sunlit slopes and shaded slopes increased when the topographic effect was more severe (UNCORR. in Fig. 8). Ideally TOC methods should reduce this difference and bring it close to zero, meaning that sunlit and shaded slopes had been homogenized, but this was not always the case. In fact, most TOC methods reduced this difference too much and showed negative difference values for some spectral bands (i.e., band 1 and band 2) (Fig. 8), which is a sign of over-correction. This was especially true for December. Again, MM had a poor performance with a strong over-correction in most cases that resulted in pixels of shaded slopes having a much higher radiance than on sunlit slopes. SM and SE were the most successful in homogenizing sunlit and shaded slopes for the three dates, with only slight differences among them. Some other methods, such as SCSC, 2SN or PBM, performed well for June and August, but not in December.

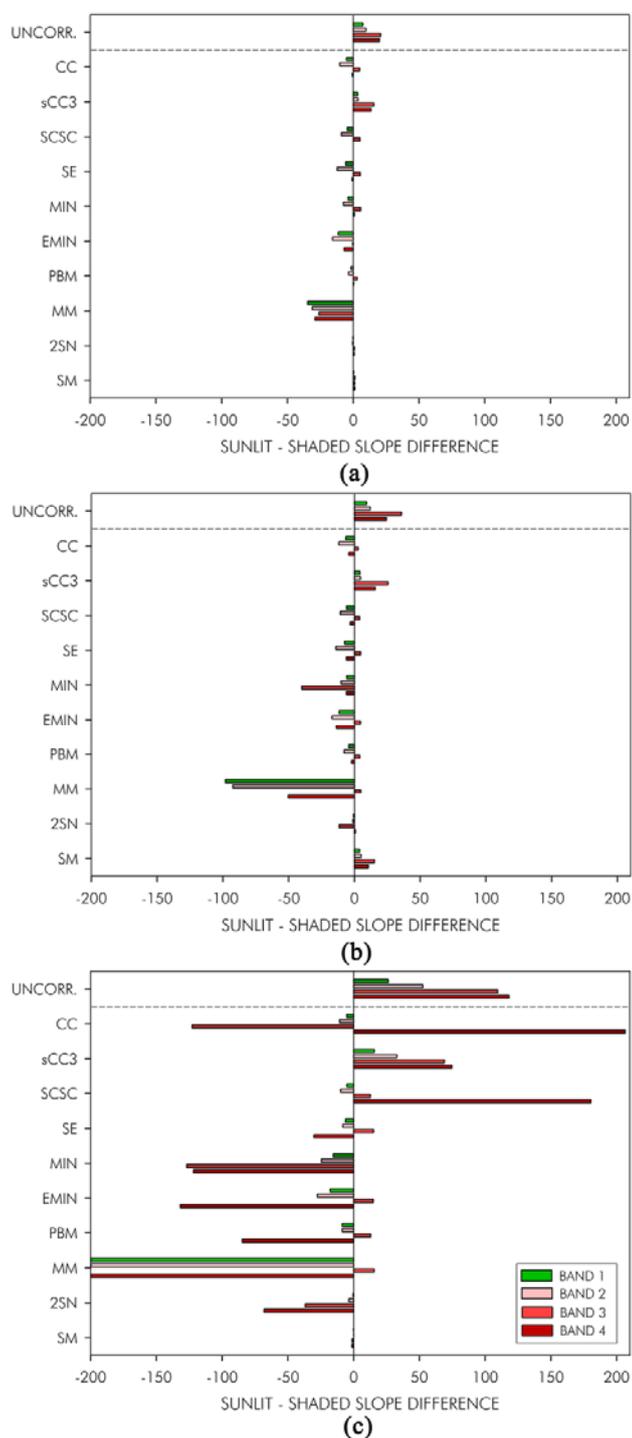


Fig. 8. Radiance difference of conifer forest on sunlit-shaded slopes ($\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$) after TOC. The uncorrected image (UNCORR.) is represented as a reference. (a) June (b) August (c) December. The closer to zero the difference is, the better.

3.6. Percentage of outliers

Having few outliers is an important characteristic for successful TOC methods. The percentage of outliers was low, always below 1%, in June and August for all of the TOC methods, as seen in Fig. 9. This percentage rose dramatically in December for some of the methods, including MM, PBM, EMIN and MIN, especially in band 4. According to this criterion, the best method was sCC3.

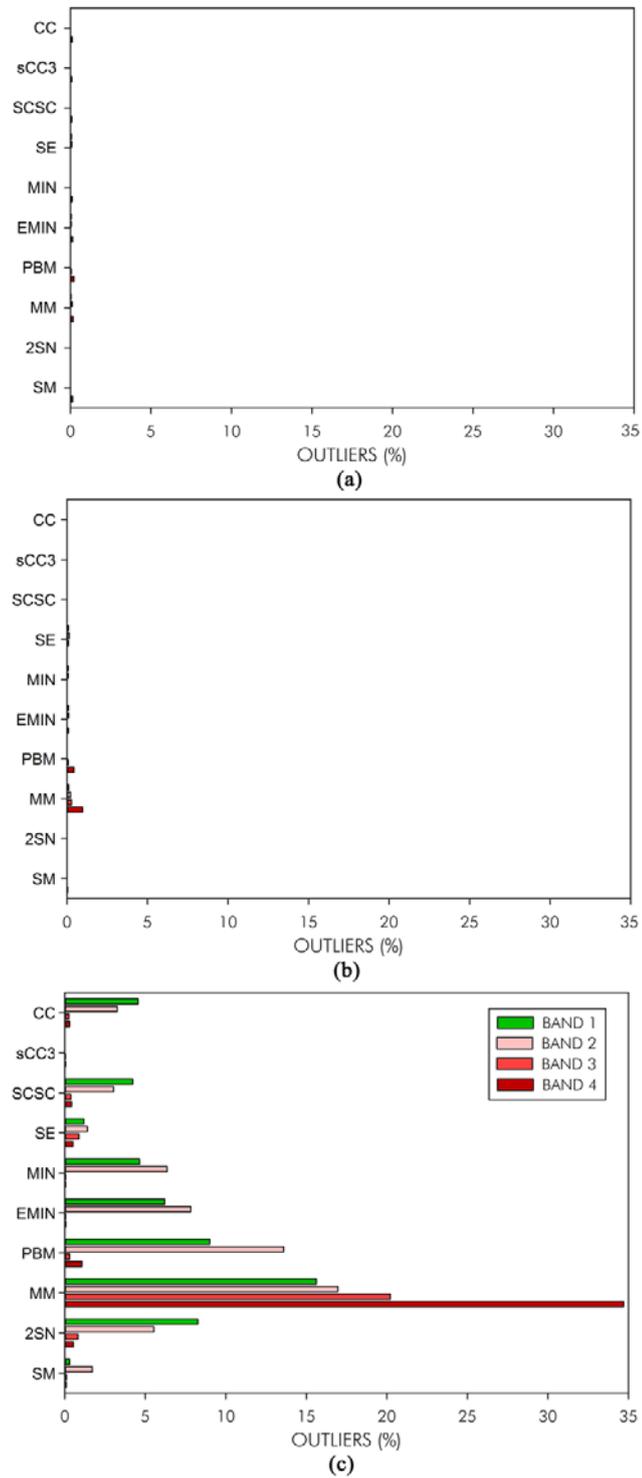


Fig. 9. Percentage of outliers generated after TOC for the four spectral bands and the three dates tested. (a) June (b) August (c) December. The smaller the percentage, the better.

3.7. Evaluation using synthetic images

The pixel-based agreement between a topographically corrected SR image (SR_{TOC}) and a SH image of the same area was measured by Structural SIMilarity index (SSIM). To obtain a global measure, the MSSIM index was computed, where a higher MSSIM is better (Table 7).

Table 7. MSSIM index between SH and SR_{TOC} . The MSSIM of the uncorrected image (UNCORR.) is given as a reference. Higher MSSIM values mean a better TOC. Figures highlighted in green and red represent the best and worst TOC method for each band and date, respectively.

TOC	JUNE				AUGUST				DECEMBER			
	Band 1	Band 2	Band 3	Band 4	Band 1	Band 2	Band 3	Band 4	Band 1	Band 2	Band 3	Band 4
UNCORR	0.911	0.912	0.881	0.759	0.839	0.836	0.775	0.628	0.538	0.517	0.310	0.255
CC	0.980	0.964	0.993	0.992	0.956	0.943	0.981	0.985	0.770	0.796	0.578	0.588
sCC3	0.935	0.928	0.924	0.821	0.881	0.872	0.849	0.717	0.643	0.638	0.411	0.384
SCSC	0.955	0.945	0.958	0.893	0.930	0.919	0.954	0.889	0.498	0.508	0.310	0.254
SE	0.968	0.954	0.984	0.970	0.950	0.938	0.961	0.940	0.753	0.771	0.628	0.611
MIN	0.969	0.941	0.991	0.970	0.955	0.938	0.980	0.970	0.635	0.673	0.405	0.499
EMIN	0.950	0.987	0.917	0.916	0.964	0.971	0.915	0.909	0.633	0.672	0.397	0.472
PBM	0.647	0.676	0.968	0.944	0.651	0.687	0.895	0.886	0.411	0.441	0.054	0.197
MM	0.750	0.899	0.946	0.958	0.625	0.807	0.974	0.967	0.203	0.322	0.587	0.549
2SN	0.948	0.924	0.943	0.826	0.864	0.846	0.813	0.662	0.619	0.598	0.330	0.297
SM	0.936	0.741	0.924	0.846	0.903	0.676	0.821	0.674	0.698	0.686	0.394	0.410

Most TOC methods achieved MSSIM values higher than 0.8 in June and August. It is easily observed that most TOC methods improved the original situation (UNCORR. in Table 7). Among the ten TOC methods evaluated, CC ranked first for this criterion. This algorithm achieved a successful correction for all of the bands and dates. Furthermore, SE, MIN and SCSC performed well in most cases, although the correction was substandard in December. Methods such as EMIN or 2SN performed better in visible bands in June and August, but failed in IR bands, bands 3 and band 4, in December. In general, the results of the comparison between SH and SR_{TOC} were significantly inadequate for December, congruous with other criteria. For this criterion, PBM obtained the lowest values, with an even poorer performance than MM, which is the worst performing TOC method for all of the other criteria.

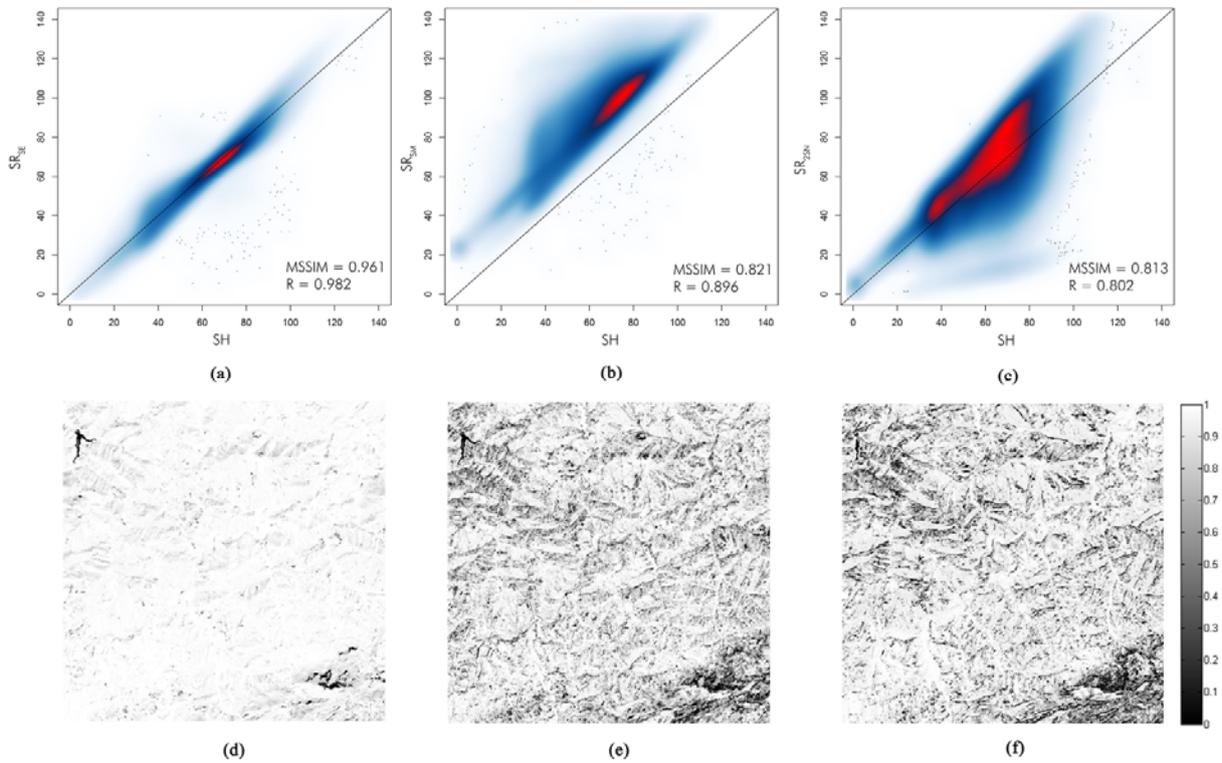


Fig. 10. Scatterplots and SSIM maps between SR_{TOC} (Y axis) and SH (X axis) in $W m^{-2} sr^{-1} \mu m^{-1}$. Only band 3 of the August scene is represented, and only results of three TOC methods are represented. The closer the scatterplot to the 1:1 line and the higher the SSIM (displayed in white), the better. **(a)** Scatterplot of SE **(b)** Scatterplot of SM **(c)** Scatterplot of 2SN **(d)** SSIM map of SE **(e)** SSIM map of SM **(f)** SSIM map of 2SN.

To further compare SR_{TOC} and SH images, scatterplots were represented along with the pixel based SSIM maps (Fig 10). To be more concise, only scatterplots for the third band of the August image were displayed for three TOC methods, SE, SM and 2SN. SE showed a good performance, with most pixels located close to the 1:1 line on the scatterplot and high SSIM values. SM was an intermediate case, with more differences between SH and SR_{TOC} depicted in both the scatterplot and the SSIM map. The scatterplot showed a clear increase in the original radiometry of land covers in the SR_{TOC} image compared to the SH image, in line with the results shown in Fig. 5. Finally, 2SN failed to successfully remove the topographic effect, showing a poor performance in the scatterplot and more dark areas clearly observed in the lower left part of the SSIM map, which corresponds to a low similarity between the SR_{TOC} image and the SH image.

3.8. TOC ranking based on standardized multi-criteria

To provide an ensemble evaluation criterion, the results obtained through the seven different evaluation procedures were standardized and assembled in a multi-criteria ranking of TOC methods (Table 8).

Table 8. Standardized multi-criteria ranking of TOC methods (the higher score, the better). Figures highlighted in green and red represent the best and worst TOC according to each criterion, respectively.

IMAGE	CRITERION	CC	sCC3	SCSC	SE	MIN	EMIN	PBM	MM	2SN	SM
JUNE	VISUAL	0.54	-1.17	0.91	1.44	-0.44	-0.50	-0.70	-1.51	1.01	0.44
	CORR	1.05	-1.11	0.47	1.08	0.84	-0.44	-0.12	-2.07	0.24	0.06
	STABILITY	-0.10	0.93	0.90	0.96	0.04	-1.37	0.18	-1.76	0.94	-0.72
	IQR RED	0.21	0.56	0.57	0.41	0.32	-0.78	-0.04	-2.56	0.66	0.65
	SUNLIT/SHADED	0.23	-0.22	0.30	0.13	0.32	-0.16	0.59	-2.68	0.77	0.72
	OUTLIERS	0.47	0.70	0.64	0.01	0.38	-0.99	-1.15	-1.77	1.50	0.22
	SR/SH	1.20	-0.28	0.38	0.96	0.93	0.47	-2.00	-0.53	-0.12	-1.02
	MEAN	0.51	-0.09	0.60	0.71	0.34	-0.54	-0.46	-1.84	0.71	0.05
AUGUST	VISUAL	0.95	-1.16	0.39	1.58	-0.81	-0.46	-0.04	-1.58	0.81	0.32
	CORR	0.98	-1.60	0.64	1.04	0.72	0.47	0.10	-1.41	0.17	-1.10
	STABILITY	-0.03	1.02	0.89	1.03	0.07	-0.93	-0.89	-1.85	0.96	-0.29
	IQR RED	0.30	0.39	0.46	0.36	0.37	-0.01	-0.10	-2.78	0.57	0.45
	SUNLIT/SHADED	0.44	0.07	0.46	0.34	-0.10	0.12	0.54	-2.78	0.60	0.30
	OUTLIERS	0.60	0.60	0.60	-0.06	0.26	0.03	-0.45	-2.66	0.59	0.51
	SR/SH	1.14	-0.57	0.60	0.90	1.07	0.81	-1.20	-0.40	-0.99	-1.34
	MEAN	0.63	-0.18	0.58	0.74	0.23	0.00	-0.29	-1.92	0.39	-0.17
DECEMBER	VISUAL	0.14	0.58	0.14	1.40	-0.65	-0.65	-1.23	-1.53	0.48	1.33
	CORR	0.88	-1.88	0.96	1.00	0.04	0.04	0.60	-1.33	-0.69	0.39
	STABILITY	0.61	1.20	0.85	0.35	0.13	-0.04	-1.72	-0.27	0.60	-1.71
	IQR RED	0.38	0.28	0.39	0.39	0.22	0.20	0.18	-2.84	0.38	0.40
	SUNLIT/SHADED	0.21	0.30	0.29	0.38	0.24	0.30	0.35	-2.84	0.35	0.42
	OUTLIERS	0.36	0.68	0.37	0.53	0.25	0.13	-0.25	-2.74	0.09	0.60
	SR/SH	1.32	-0.03	0.10	1.39	0.25	0.18	-2.03	-0.89	-0.50	0.20
	MEAN	0.56	0.16	0.44	0.78	0.07	0.02	-0.59	-1.78	0.10	0.23
AVERAGE		0.57	-0.03	0.54	0.74	0.21	-0.17	-0.45	-1.85	0.40	0.04

Among the methods tested, SE could globally be considered the best, as it performed well according to most of the criteria for the three dates. Similarly, CC, SCSC and 2SN also obtained good results, with only minor differences between all these best-performing TOC methods. In contrast, MM performed significantly worse than other TOC methods.

Other methods, such as MIN, EMIN, PBM or SM, had inconsistent results, with good performances according to some criteria and poor results according to others. For example, SM ranked in the first positions for some criteria in December, like IQR reduction or sunlit shaded slopes, but had poor results in terms of stability, as it modified the radiometry of land covers. This was further confirmed by the synthetic image evaluation procedure for June and August. In general, Minnaert-based methods MIN, EMIN and PBM performed slightly worse than the top ranked TOC methods. In particular, MIN and

EMIN showed a poor performance in the visual assessment, and EMIN also performed poorly in terms of stability, similar to PBM. On the contrary, sCC3 performed the best in this criterion, but was unsuccessful in reducing the IQR of the radiance of land covers.

4. Discussion

Out of the ten TOC methods compared in this study, SE, CC and SCSC, in this order, seemed to outperform the others. They resulted in a better reduction of the topographic effect according to the different criteria used. The results obtained by the better TOC methods, SCSC, SE and CC, are in line with results previously shown by Soenen et al. (2008), who suggested that SE, CC, MIN and SCSC corrections all reduced the topographic effect to a similar extent, but were still subject to overcorrection effects in steep, shaded slopes. In line with the results from Riaño et al. (2003), sCC3 was the best method for retaining the spectral characteristics of each band, although this method did not successfully remove the dependence of radiance on illumination as the authors suggested. It could be concluded that this method performed well only under good illumination conditions, such as those considered by Riaño et al.

In general, Minnaert-based methods performed significantly worse than the best performing TOC methods for all three dates. This was in line with previous investigations, where some authors (e.g., Hantson and Chuvieco 2011) claimed better performances under larger solar elevation angles. Conversely, it contrasted with other authors who claimed a poor performance under low solar elevation angles, about 25° (Zhang and Gao 2009b), similar to our image from December. Finally, the worst corrections were obtained with MM and PBM. These results contrast with what was previously shown by Richter et al. (2009), who claimed that MM achieved the best visual ranking compared with other frequently used TOC methods. This could partly be due to the inconsistent nature of these methods, where it is necessary to tune and optimally select a number of empirical parameters for each particular case or dataset.

For the different evaluation criteria, some are hard to compare in a multitemporal manner, as their values depend on the illumination conditions of each particular date. For instance, the difference between conifer forests located on sunlit or shaded slopes increased when the topographic effect was more severe.

Moreover, the comparison between SH and SR_{TOC} measured through the SSIM index also depended on the date. Similarly, the same rate of IQR reduction could be a sign of successful correction in June, but not in December. On the other hand, some other criteria did not follow this trend. For example, in the visual assessment (Table 6), SM ranked higher in December than in June or August. This appears to be caused by the weaker topographic effect to be corrected in the images of June and August, which resulted in fewer differences between TOC methods. Therefore, the visual assessment as a criterion to compare TOC methods in favorable conditions may not be adequate. According to the correlation criterion, SE, CC and SCSC were the best methods for reducing the dependence of spectral radiance on illumination in the three studied images, with minimum slopes and R-values. MIN also performed well in June and August for reducing a dependency on spectral radiance. In contrast, residual correlation remained after the correction in December. These results are in line with previous studies (Gao et al. 2014; Soenen et al. 2008; Wu et al. 2008).

Table 8 also showed that some TOC evaluation criteria gave very similar results, while others displayed significant contrast. For example, SE ranked first in December according to the visual analysis and synthetic image criteria, but ranked only fifth in stability. Related, the criteria that appeared to be the most representative of the global multi-criteria analysis were the reduction of intraclass IQR and the comparison of sunlit and shaded slopes, a clear signal for their usefulness in assessing the performance of TOC methods.

Land covers behave differently when TOC methods are applied based to their degree of non-Lambertian behavior (Mariotto and Gutschick, 2010). Thus, it is interesting to analyze the results per land cover. When the reduction of IQR for land-covers was analyzed for the NIR band (Fig. 7), the best results were obtained for rocks. This could be due to the structure of this land cover being more homogeneous, and, consequently, their reflective behavior more controlled by the topography. Further, this land cover is frequently located on very steep slopes. On the other hand, mixed forests and broad-leaved forests showed slightly lower degrees of IQR reduction for the NIR band. The reason for this could be due to the differences in forest structure and canopy self-shadowing, which is not considered in this work, but would

be a major issue to take into account in future developments in this field (Gu and Gillespie, 1998; Soenen et al., 2005; Kane et al. 2008).

Among the spectral bands, the NIR and SWIR (bands 3 and 4) had the highest difference between shaded and sunlit slopes on the uncorrected images, probably due to a particularly strong topographic effect on these spectral bands, which is in line with previous studies (Nichol et al. 2006). Nevertheless, most TOC methods achieved a great reduction of these differences for the NIR and SWIR bands. Similar to what was found by Balthazar et al. (2012), statistical outliers were found mostly in very low-illuminated areas, generally shaded slopes, those with low values of $\cos\gamma_i$. Consequently, the amount of outliers significantly increased in December (Fig. 9). Finally, the use of synthetic images suggested a general good performance of SE and CC, which is consistent with the results obtained with other criteria.

In conclusion, the performance of 10 widely used TOC methods were assessed through a multi-criteria analysis based on images acquired on three different dates, covering a range of illumination conditions from moderate to severe. The results obtained have been used to provide basic guidelines on the use of TOC methods for different statistics, as summarized in a table of pros and cons (Table 9):

Table 9. Pros and cons of TOC methods for different characteristics

TOC	Pros	Cons
CC	Good performance in all conditions according to most criteria	Not the best in stability and comparison between sunlit and shaded slopes
sCC3	Stable land cover radiometry in all conditions	Topographic effect not completely removed
SCSC	Good performance in all conditions, very stable and without outliers	Originally designed only for forested areas
SE	Good performance in all conditions. The best according to the visual and correlation analysis, especially in severe conditions	Land cover radiometry not completely stable in severe conditions
MIN	Good or moderate performance in favorable conditions	Problems of overcorrection in poorly illuminated areas. Presence of outliers in poorly illuminated areas
EMIN	Good or moderate performance in favorable conditions	No improvement observed when compared to MIN
PBM	Good performance in comparison between sunlit and shaded slopes and correlation analysis	Land cover radiometry not stable. Presence of outliers in poorly illuminated areas. Open to users' arbitrary decisions
MM	There are no pros for this method	Poor performance in all conditions. Open to users' arbitrary decisions
2SN	Best method in favorable conditions. Good performance in terms of visual assessment and IQR reduction of land covers	Poor performance with synthetic images. Open to users' arbitrary decisions
SM	Reasonable performance, especially in severe conditions	Land cover radiometry not stable and poor performance with synthetic images. Open to users' arbitrary decisions

5. Conclusions

This paper aimed to perform a multi-criteria analysis of different topographic correction (TOC) methods applied under different conditions, including advantages and shortcomings of each TOC algorithm. The multi-criteria analysis showed that the use of a single evaluation procedure to assess the quality of topographic correction algorithms might be inappropriate, as the quality of the correction depends on several factors. Some evaluation criteria had similar rankings to the multi-criteria ranking, for example, intraclass IQR reduction or the comparison of sunlit and shaded slopes. However, other criteria had more of a contrast in their rankings when compared to the multi-criteria ranking. Consequently, the use of a multi-criteria approach to assess the global performance of topographic correction algorithms is strongly recommended. If a single evaluation criterion was to be recommended, the intraclass IQR reduction should be used, since it was the most representative of the multi-criteria ranking. However, land cover distribution is a prerequisite for this criterion to be utilized.

TOC performance depended strongly on the magnitude of the topographic effect to correct, and this depended on the illumination conditions of the acquisition. In favorable conditions, such as June, most TOC algorithms performed adequately and the differences between the best TOC methods (SE, 2SN, SCSC and CC) were minor. In these favorable conditions, the selection of one algorithm or another seemed to have little impact on the outcome of the correction. However, as the topographic effect became stronger, as for August and especially December, differences between TOC algorithms became more apparent. According to our results, methods including slope smoothing (i.e., sCC3) or based on the Minnaert approach (i.e., MIN, EMIN or PBM) should be avoided when poor illumination conditions were considered. Slope smoothing methods were unsuccessful at reducing the intraclass IQR, while the Minnaert approach introduced too many artifacts and failed in terms of radiometric stability.

As a practical recommendation, the TOC algorithms that achieved a best performance were SE, CC and SCSC, and these could be recommended for most situations. Other methods, such as 2SN and SM, performed inconsistently, with results varying from good to moderate depending on the image,

illumination conditions or method implementation. Finally, some methods were found to be too complex to apply, as they required many parameters. Others, such as PBM, MM, 2SN and SM, were open to arbitrary decisions of the user to adapt the correction to each specific dataset. These issues are not minor, since every processing algorithm should be as simple as possible to facilitate its use and implementation in automated image processing chains.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Public University of Navarre (UPNA). The authors would also like to thank the Spanish National Geographic Institute (IGN) for providing the test data. Part of the research presented in this paper is funded by the Spanish Ministry of Economy and Competitiveness in the frame of the ESP2013-48458-C4-2-P project.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

- Balthazar, V., Vanacker, V., & Lambin, E.F. (2012). Evaluation and parameterization of ATCOR3 topographic correction method for forest cover mapping in mountain areas. *International Journal of Applied Observation and Geoinformation*, 18, 436-450.
- Baraldi, A., Girona, M., & Simonetti, D. (2010). Operational Two-Stage Stratified Topographic Correction of Spaceborne Multispectral Imagery Employing an Automatic Spectral-Rule-Based Decision-Tree Preliminary Classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 112-146.
- Bishop, M.P., & Colby, J.D. (2002). Anisotropic reflectance correction of SPOT-3 HRV imagery. *International Journal of Remote Sensing*, 23, 2125-2131.
- Bishop, M.P., Shroder Jr., J.F., & Colby, J.D. (2003). Remote sensing and geomorphometry for studying relief production in high mountains. *Geomorphology*, 55, 345-361.
- Blesius, L., & Weirich, F. (2005). The use of the Minnaert correction for land-cover classification in mountainous terrain. *International Journal of Remote Sensing*, 26, 3831-3851.
- Brunet, D., Vrscay, E. R., & Wang, Z. (2012). On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 21, 4, 1488-1499.
- Civco, D.L. (1989). Topographic Normalization of Landsat Thematic Mapper Digital Imagery. *Photogrammetric Engineering & Remote Sensing*, 55, 1303-1309.

- Ekstrand, S. (1996). Landsat TM-based forest damage assessment: Correction for topographic effects. *Photogrammetric Engineering & Remote Sensing*, 62, 151-161.
- Fan, Y., Koukal, T., & Weisberg, P.J. (2014). A sun-crown-sensor model and adapted C-correction logic for topographic correction of high resolution forest imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96, 94-105.
- Gao, M.L., Zhao, W.J., Gong, Z.N., Gong, H.L., Chen, Z., & Tang, X.M. (2014). Topographic Correction of ZY-3 Satellite Images and Its Effects on Estimation of Shrub Leaf Biomass in Mountainous Areas. *Remote Sensing*, 6, 2745-2764.
- Ghasemi, N., Mohammadzadeh, A., & Reza Sahebi, M. (2013). Assessment of different topographic correction methods in ALOS AVNIR-2 data over a forest area. *International Journal of Digital Earth*, 6, 504-520.
- Gobierno Vasco (2007). Cartografía de hábitats, vegetación actual y usos del suelo de la Comunidad Autónoma del País Vasco. (URL: <http://www.geo.euskadi.eus/geograficos/habitats-vegetacion-actual-y-usos-del-suelo/s69-geodir/es/> (accessed 04/2015)).
- Gobierno Vasco (2010). Mapa forestal del País Vasco. Año 2010 (URL: http://www.geo.euskadi.eus/s69-geodir/es/contenidos/ds_geograficos/md_ideeu_invent_forest_2010/es_def/index.shtml (accessed 01/2016)).
- Goslee, S.C. (2012). Topographic Corrections of Satellite Data for Regional Monitoring. *Photogrammetric Engineering & Remote Sensing*, 78, 973-981.
- Gu, D., & Gillespie, A. (1998). Topographic normalization of Landsat TM images of forest based on subpixel Sun-canopy-sensor geometry. *Remote Sensing of Environment*, 64, 166-175.
- Hantson, S., & Chuvieco, E. (2011). Evaluation of different topographic correction methods for Landsat imagery. *International Journal of Applied Observation and Geoinformation*, 13, 691-700.
- Hoshikawa, K., & Umezaki, M. (2014). Effects of terrain-induced shade removal using global DEM data sets on land-cover classification. *International Journal of Remote Sensing*, 35, 1331-1355.
- Kane, V. R., Gillespie, A. R., McGaughey, R., Lutz, J. A., Ceder, K., Franklin, J. F. (2008). Interpretation and topographic compensation of conifer canopy self-shadowing. *Remote Sensing of Environment*, 112, 3820-3832.
- Kobayashi, S., & Sanga-Ngoie, K. (2008). The integrated radiometric correction of optical remote sensing imageries. *International Journal of Remote Sensing*, 29, 5957-5985.
- Kobayashi, S. & Sanga-Ngoie, K. (2009). A comparative study of radiometric correction methods for optical remote sensing imagery: the IRC vs. other image - based C - correction methods. *International Journal of Remote Sensing*, 30:2, 285-314.
- Law, K.H., & Nichol, J. (2004). Topographic correction for differential illumination effects on IKONOS satellite imagery. In, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (pp. 641-646).
- Lu, D., Ge, H., He, S., Xu, A., Zhou, G., & Du, H. (2008). Pixel-based Minnaert Correction Method for Reducing Topographic Effects on a Landsat 7 ETM+ Image. *Photogrammetric Engineering & Remote Sensing*, 74, 1343 – 1350.
- Mariotto, I., & Gutschick, V.P. (2010). Non-Lambertian Corrected Albedo and Vegetation Index for Estimating Land Evapotranspiration in a Heterogeneous Semi-Arid Landscape. *Remote Sensing*, 2, 926-938.
- Minnaert, M. (1941). The reciprocity principle in lunar photometry. *Astrophysical Journal*, 93, 403-410.

- Moreira, E.P., & Valeriano, M.M. (2014). Application and evaluation of topographic correction methods to improve land cover mapping using object-based classification. *International Journal of Applied Earth Observation and Geoinformation*, 32, 208-217.
- Nagol, J.R., Sexton, J.O., Kim, D.-H., Anand, A., Morton, D., Vermote, E., & Townshend, J.R. (2015). Bidirectional effects in Landsat reflectance estimates: Is there a problem to solve? *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 129–135.
- Nichol, J., Hang, L.K., & Sing, W.M. (2006). Empirical correction of low Sun angle images in steeply sloping terrain: a slope-matching technique. *International Journal of Remote Sensing*, 27, 629-635.
- Notarnicola, C., Callegari, M., De Gregorio, L., Sonnenschein, R., Remelgado, R., & Ventura, B. (2014). A novel topographic correction for high and medium resolution images by using combined solar radiation. Proceedings, IGARSS 2014 symposium, Québec, Canada, (pp. 1053-1056).
- Reese, H., & Olsson, H. (2011). C-correction of optical satellite data over alpine vegetation areas: A comparison of sampling strategies for determining the empirical c-parameter. *Remote Sensing of Environment*, 115, 1387-1400.
- Riaño, D., Chuvieco, E., Salas, J., & Aguado, I. (2003). Assessment of different topographic corrections in Landsat-TM data for mapping vegetation types. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 1056-1061.
- Richter, R. (1998). Correction of satellite imagery over mountainous terrain. *Applied Optics*, 37, 4004-4015.
- Richter, R., Kellenberger, T., & Kaufmann, H. (2009). Comparison of topographic correction methods. *Remote Sensing*, 1, 184-196.
- Sandmeier, S., & Itten, K.I. (1997). A Physically-Based Model to Correct Atmospheric and Illumination Effects in Optical Satellite Data of Rugged Terrain. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 708-717.
- Schulmann, T., Katurji, M., & Zavar-Reza, P. (2015). Seeing through shadow: Modelling surface irradiance for topographic correction of Landsat ETM+ data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 99, 14-24.
- Shepherd, J. D., Dymond, J. R., Gillingham S. & Bunting P. (2014). Accurate registration of optical satellite imagery with elevation models for topographic correction. *Remote Sensing Letters*, 5, 637-641.
- Smith, J.A., Lin, T.L., & Ranson, K.J. (1980). The Lambertian Assumption and Landsat Data. *Photogrammetric Engineering & Remote Sensing*, 46, 1183-1189.
- Soenen, S.A., Peddle, D.R., & Coburn, C.A. (2005). SCS+C: A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 2148-2159.
- Soenen, S.A., Peddle, D.R., Coburn, C.A., Hall, R.J., & Hall, F.G. (2008). Improved topographic correction of forest image data using a 3-D canopy reflectance model in multiple forward mode. *International Journal of Remote Sensing*, 29, 1007-1027.
- Soenen, S.A., Peddle, D.R., Coburn, C.A., Hall, R.J., & Hall, F.G. (2009). Canopy Reflectance Model Inversion in Multiple Forward Mode: Forest Structural Information Retrieval from Solution Set Distributions. *Photogrammetric Engineering & Remote Sensing*, 75, 361-374.
- Sola, I., González de Audicana, M., Álvarez-Mozos, J., & Torres, J.L. (2014). Synthetic images for evaluating Topographic Correction Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 52, 1799 – 1810.

- Szantoi, Z., & Simonetti, D. (2013). Fast and Robust Topographic Correction Method for Medium Resolution Satellite Imagery Using a Stratified Approach. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 6, 1921-1933.
- Tan, B., Masek, J.G., Wolfe, R., Gao, F., Huang, C., Vermote, E.F., Sexton, J.O., & Ederer, G. (2013). Improved forest change detection with terrain illumination corrected Landsat images. *Remote Sensing of Environment*, 136, 469-483.
- Teillet, P.M., Guindon, B., & Goodenough, D.G. (1982). On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, 8, 84-106.
- Tokola, T., Sarkeala, J., & Van der Linden, M. (2001). Use of topographic correction in Landsat TM-based forest interpretation in Nepal. *International Journal of Remote Sensing*, 22, 551-563
- Twele, A., & Erasmi, S. (2005). Evaluating topographic correction algorithms for improved land cover discrimination in mountainous areas of central Sulawesi. In S. Erasmi, B. Cyffka, & Kappas; (Eds.), *Remote Sensing & GIS for environmental studies*. Göttingen.
- Vanonckelen, S., Lhermitte, S., Balthazar, V., & Van Rompaey, A. (2014). Performance of atmospheric and topographic correction methods on Landsat imagery in mountain areas. *International Journal of Remote Sensing*, 35, 4952-4972.
- Vanonckelen, S., Lhermitte, S., & Van Rompaey, A. (2013). The effect of atmospheric and topographic correction methods on land cover classification accuracy. *International Journal of Applied Earth Observation and Geoinformation*, 24, 9-21.
- Vanonckelen, S., Lhermitte, S., & Van Rompaey, A. (2015). The effect of atmospheric and topographic correction on pixel-based image composites: Improved forest cover detection in mountain environments. *International Journal of Applied Earth Observation and Geoinformation*, 35, Part B, 320-328.
- Veraverbeke, S., Verstraeten, W.W., Lhermitte, S., & Goossens, R. (2010). Illumination effects on the differenced Normalized Burn Ratio's optimality for assessing fire severity. *International Journal of Applied Observation and Geoinformation*, 12, 60-70.
- Vicente-Serrano, S.M., Pérez-Cabello, F., & Lasanta, T. (2008). Assessment of radiometric correction techniques in analyzing vegetation variability and change using time series of Landsat images. *Remote Sensing of Environment*, 112, 3916-3934.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; Simoncelli, E. P. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13, 600-612.
- Wang, Z.; Bovik, A.C.; Simoncelli, E.P. (2005). Handbook of Image and Video Processing, 2nd edition. Academic Press, Inc., Orlando, FL, USA.
- Wu, J., Bauer, M.E., Wang, D., & Manson, S.M. (2008). A comparison of illumination geometry-based methods for topographic correction of QuickBird images of an undulant area. *ISPRS Journal of Photogrammetric Remote Sensing*, 63, 223-236.
- Zhang, W.C., & Gao, Y.N. (2009a). LULC Classification and Topographic Correction of Landsat-7 ETM+ Imagery in the Yangjia River Watershed: the Influence of DEM Resolution. *Sensors*, 9, 1980-1995.
- Zhang, W.C., & Gao, Y.N. (2009b). A simple empirical topographic correction method for ETM plus imagery. *International Journal of Remote Sensing*, 30, 2259-2275.
- Zhang, W.C., & Gao, Y.N. (2011). Topographic correction algorithm for remotely sensed data accounting for indirect irradiance. *International Journal of Remote Sensing*, 32, 1807-1824.

Zhang, Z. M.; De Wulf, R. R.; Van Coillie, F. M. B.; Verbeke, L. P. C.; De Clercq, E. M.; Ou, X. K. (2011). Influence of different topographic correction strategies on mountain vegetation classification accuracy in the Lancang Watershed, China. *Journal of Applied Remote Sensing*, 5, 1-22.

Zhang, Y.; Yan, G.; Bai, Y. (2015). Sensitivity of Topographic Correction to the DEM Spatial Scale. *IEEE Geoscience and Remote Sensing Letter*, 12, 53-57.

Zhang, Z. M.; De Wulf, R. R.; Van Coillie, F. M. B.; Verbeke, L. P. C.; De Clercq, E. M.; Ou, X. K. (2011). Influence of different topographic correction strategies on mountain vegetation classification accuracy in the Lancang Watershed, China. *Journal of Applied Remote Sensing*, 5, 1-22.

Zhang, Y.; Yan, G.; Bai, Y. (2015). Sensitivity of Topographic Correction to the DEM Spatial Scale. *IEEE Geoscience and Remote Sensing Letter*, 12, 53-57.