RELATIVE RADIOMETRIC NORMALIZATION USING SEVERAL AUTOMATICALLY CHOSEN REFERENCE IMAGES FOR MULTI-SENSOR, MULTI-TEMPORAL SERIES

Charles Hessel[1,*], Rafael Grompone von Gioi[1], Jean-Michel Morel[1], Gabriele Facciolo[1], Pablo Arias[1], Carlo de Franchis[1,2]

[1] Université Paris-Saclay, CNRS, ENS Paris-Saclay, Centre Borelli, 94235, Cachan, France – [2] Kayrros SAS

KEY WORDS: Relative radiometric normalization, Pseudo-invariant features, Historical satellite time series, Multi-sensor

ABSTRACT:

We propose a method for the relative radiometric normalization of long, multi-sensor image time series. This allows to increase the revisit time under comparable conditions. Although the relative radiometric normalization is a well-studied problem in the remote sensing community, the availability of an increasing number of images gives rise to new problems. For example, given long series spanning several years, finding features that are maintained through the whole period of time becomes arduous. Instead, we propose in this paper to use automatically detected reference images chosen by maximization of a quality metric. For each image, two affine correction models are robustly estimated using random sample consensus, using the two closest reference images; the final correction is obtained by linear interpolation. For each pair of source and reference images, pseudo-invariant features are obtained using a similarity measure invariant to radiometric changes. A final tone-mapping step outputs the images in the standard 8-bits range. This method is illustrated by the fusion of time series of Sentinel-2 at correction levels 1C, 2A, and Landsat-8 images. By using only the atmospherically corrected Sentinel-2 L2A images as anchors, the full output series inherits this atmospheric correction.

1. INTRODUCTION

Given the numerous satellites observing the earth, any point on the ground is present in multiple and long image time series. It becomes indispensable to build more complete time series obtained by fusing the output of different imaging devices (with roughly comparable resolution). However, due to the differences between sensors and acquisition modes, such a task is more intricate than expected. In particular, resolutions are different, dynamic ranges are different, color bands are different. Still more challenging, the pixels where the ground is visible and stable can vary in each image due to atmospheric perturbations, clouds, and human actions. Thus equalization requires the application of a different sieve to each image.

Besides facilitation of human visual scene interpretation, the need for relative radiometric normalization in satellite time series is acknowledged since the seventies, with the start of the Landsat missions. Coppin and Bauer noted in 1996 that there were two outstanding requirements for satellite time series analysis: multi-temporal image registration and radiometric calibration [Coppin, Bauer, 1996]. Even now, radiometric normalization is a fundamental step in many applications [Canty, Nielsen, 2008]: to prepare data for change detection [Paolini et al., 2006, Jianya et al., 2008]; tracking vegetation indices over time, for example with the Normalized Difference Vegetation Index (NDVI) [Du et al., 2002]; for supervised and unsupervised land cover classification; for multi-temporal Satellite image mosaicing [Rahman et al., 2015]. The method presented in this paper was successfully used in [Drouyer, 2020] to improve the accuracy of parking occupancy estimation with time series of PlanetScope images.

According to Du *et al.* [Du et al., 2002], there are four main reasons for the variations of sensor measurements of the same scene at different dates: (1) changes in satellite sensor calibration over time, (2) differences in illumination and observation angles, (3) variation in atmospheric effects, and (4) changes in target

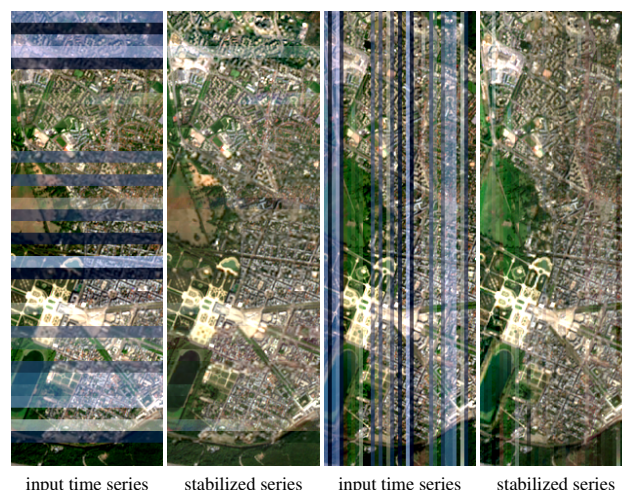input time series   stabilized series   input time series   stabilized series

Figure 1. Composite image, formed by concatenating horizontal (left side) or vertical (right side) bands from each image of the time series. The time series contains a total of 39 images, captured by Sentinel-2 (correction levels 1C and 2A) and Landsat-8. Our algorithm automatically normalizes the spectral values in the series using a few well chosen references.

reflectance. In the case of multi-satellite fusion, a fifth reason is the difference of sensors.

Figure 1 shows the aspect of a time series without radiometric normalization. The image is recomposed by concatenating horizontal or vertical bands from the inputs. After the relative radiometric normalization, seams between the bands are much less visible, even though bands come from images taken at different dates and with different sensors. The first and last images were captured on Feb. 18, 2018 and Sep. 21, 2019, respectively, over Versailles (France). Their size is 2×5 km. Figure 2 shows the temporal coherence of a region within this series after the application of our algorithm.

Radiometric correction methods are generally classified in two

---

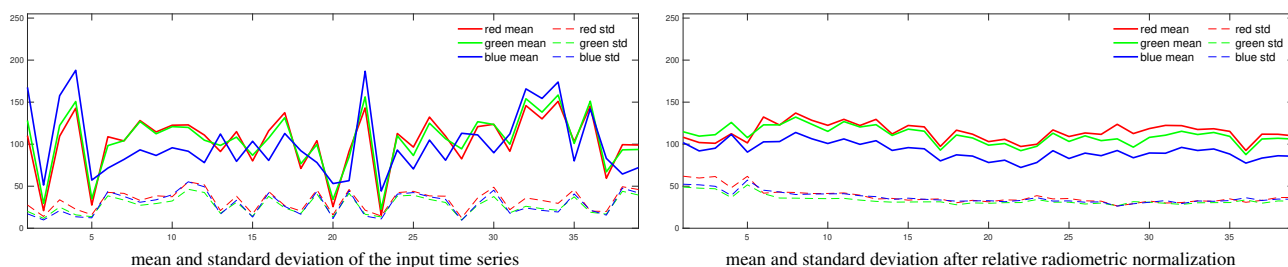* Corresponding author (charles.hessel@ens-paris-saclay.fr)

Figure 2. Temporal evolution of the mean and variance of a small sample in the initial (left) and radiometrically normalized (right) time series. The sample is displayed in Figure 3; this is a urban area. Our algorithm outputs a series with consistent mean and variance through time. Small variations are unavoidable due to changes in observation and illumination angles, in the resolution, calibration and spectral filters of the different sensors, and in the reflectance on the ground.

categories: the first encompasses the ones that aim at mapping the sensors' values to physical measurements like ground radiance or reflectance [Yuan, Elvidge, 1996]. These methods require external measurements. The second category includes methods whose objective is simply to normalize the sensors' values, so that the relative changes remaining are only due to changes in the target reflectance. These methods use only information from the time series itself. In many cases, the retrieval of surface reflectance to correct the sensor calibration, differences in illumination, view angles and atmospheric effects, is difficult or even impossible [Du et al., 2002]. This second category is therefore a desirable alternative. Moreover, absolute radiometric correction is in many cases not required.

In the 1996 review of seven relative radiometric normalization techniques [Yuan, Elvidge, 1996] the authors noticed that all used affine models whose parameters were found by regression. The assumption is that among the different sources of change between the same channel of two images at two different dates, linear effects dominate nonlinear effects [Du et al., 2002]. Hall *et al.* [Hall et al., 1991] noted that the affine nature of the transformation linking two images can be destroyed by the heterogeneity in the atmospheric properties across the scene and the non-linearity in calibration differences between sensor. From our own experiments, the former is a far more serious issue. We indeed generally keep in our time series images with transparent clouds, but the non homogeneity of these clouds hinders a perfect correction for these images.

The most important and difficult parts of relative radiometric correction are the selection of pseudo-invariant features (PIFs) [Du et al., 2002, Canty et al., 2004], and the regression method used to find the affine model parameters [Syariz et al., 2019].

Hall *et al.* [Hall et al., 1991] proposed a method to rectify the colors of Landsat-5 images, at different dates and with different sensors. Their algorithm identifies "radiometric control sets, *i.e.*, sets of scene landscape elements with a mean reflectance which is expected to change little with time." They construct those sets by selecting the extremes of the Kauth-Thomas greenness-brightness scattergram [Kauth, Thomas, 1976], which are supposed to have the same average surface reflectance between images. These pixels may not be the same in different images.

To select the PIFs, [Du et al., 2002] used principal component analysis (PCA) and quality control. They tested their procedure using three Landsat-5 Thematic Mapper images of the same area in years 1986, 1987, and 1991. Their algorithm detects and removes cloud and water pixels, found using simple thresholds on the pixels' values. This method assumes that the affine transformation of the PIFs in the couple of images is close to the

identity. Their quality control consists in computing a linear correlation coefficient for the candidate PIFs; then verifying that the obtained major axis is close to unity. While it is not close enough, the parameters are iteratively updated and a new axis is computed. The method [Xu et al., 2012] also requires this linear relationship to accurately find the pseudo-invariant features. Our method, however, makes no assumption on the PIFs.

Multivariate alteration detection (MAD) transformation was proposed by [Canty et al., 2004] to obtain invariant pixels. This technique was proposed in [Nielsen et al., 2002, Nielsen et al., 1998]. They first form linear combinations of the intensities for all channels in the two images (at different dates). The combination's coefficients are determined so that the positive correlation between the two obtained images is minimized. This is meant to enhance the actual changes between the two images as much as possible; then pixels with smallest normalized MAD components are selected for the regression. Only one image is used as a reference.

In 2015, Lin *et al.* [Lin et al., 2015] did simultaneous cloud detection and radiometric normalization. They argue that both objectives overlap, as both require to find invariant pixels. Based on the method of [Du et al., 2002], they proposed a weighted principal component analysis to help removing outliers and finding invariant pixels. High weights are assigned to points with strong spectral similarities, which are measured using Euclidean distance, spectral angle and spectral correlation. These three measures were proposed for change detection [Carvalho Júnior et al., 2011] and adapted to the pseudo-invariant feature selection [De Carvalho et al., 2013].

In this paper, we propose an algorithm addressing these problems, able to produce time series from different statellite sources with uniform contrast and color on relevant stable pixels. We shall test our method on time series with mixed Sentinel-2 at correction level L1C and L2A, and Landsat-8 data. Since Sentinel-2 L2A images are atmospherically corrected, the final time series with normalized radiometry inherits this atmospheric correction without the need of other external data.

Unlike other methods, our algorithm first automatically chooses in the series regularly distributed reference images that maximize three criteria: ground visibility, local contrast and sensor quality (resolution and correction level). Since the time series may span several years, the reference samples must be dense enough to cope with seasonal color changes. All remaining images are then equalized with the two closest references. To this aim, relevant stable ground pixels are found in these images by comparing their gradients' angles pairwise. Then an affine radiometric correction matching the colors of these pixels to those of both

nearest reference images is robustly estimated using random sample consensus. The final correction is a temporal linear interpolation of both affine corrections.

The method we propose makes no assumption on the input images: they can be of any range, color, number of channels, resolution or from any sensor. This means that we do not rely on thresholds that have to be set for each different satellite. Furthermore, our method handles well small images (*e.g.* 500×500 pixels), which usually cause problems for algorithms that require to find, for example, "black pixels" from areas with zero reflectance, to estimate the thickness of the atmosphere.

Our aim is therefore threefold: (1) to create *visualizable* time series from more than one satellite, in order to ease humans' interpretation, inspection and annotation; (2) to prepare the data for detection algorithms that expect inputs with consistent color, contrast, resolution and dynamic ranges; (3) to obtain atmospherically corrected images without any knowledge of the atmospheric conditions, but rather by using as example the available and already-corrected Sentinel-2 L2A images.

Our contributions are: (1) A general method for relative radiometric normalization of long and heterogeneous time series; (2) A complete pipeline for the preparation of multi-satellite data; (3) A simple example-based tool for atmospheric correction: given some atmospherically corrected images inserted in the time series, our method propagates the correction to neighboring frames, without requesting external data.

2. METHOD DESCRIPTION

Sentinel-2 visible bands have a resolution of 10m, while Landsat-8 panchromatic band is at 15m and its visible blue, green and red bands are at 30m resolution. To create a time series with images from both satellites, the pixels' size must be made uniform. We thus use the panchromatic band to pansharpen Landsat-8 visible bands to 15m/pixel and then upsample them to 10m/pixel by interpolation. Gdal[1] pansharpening tool is used for the first step. This algorithm first upsamples the low resolution bands, then computes the ratio between the panchromatic band and the weighted average of the upsampled bands, and finally applies this ratio to the different spectral bands to pansharpen. The time series is then registered with sub-pixel accuracy using the phase correlation method [Foroosh et al., 2002]. We sample the image after translation by spline interpolation [Briand, Davy, 2019].

Sentinel-2 and Landsat-8's sensors output very different values. Landsat-8 typically has values around 10 times higher than Sentinel-2 for the same ground reflectance. Although our method is absolutely independent from this, we applied a first rough normalization of the images in function of their sensors, so as to place most of their values in the range $[0, 1]$. This normalization is needed, for example in Figure 1, to display images from the two satellites side by side with comparable values.

The overall image series processing chain then works as follows: (1) Apply the ground visibility detector, remove exceedingly cloudy images from the time series; (2) Apply quality metrics and select reference images as local maxima of the global quality score; (3) Find stable pixels by comparing the gradient angles in the pairs of source and reference images; (4) Correct the spectral values based on relevant pixels and the two closest references images; (5) Apply a final tone-mapping step. These steps are summarized in Algorithm 1. Next sections detail each step.

[1] https://gdal.org/index.html

Algorithm 1: Relative Radiometric Normalization

input : registered time series (possibly mutli-sensor)
output : time series with normalized radiometric values
Apply the ground visibility detector
foreach *image* **do** Compute image quality `// Equation (6)`
Select as key images the ones with highest quality
foreach *image* **do** `// estimate and apply corrections`
    foreach *channel* **do**
        for *previous key, next key* **do**
            compute mask for the pair (current, key)
            estimate the affine model with RANSAC
        interpolate the parameters
        correct the image
foreach *image* **do** Apply tone-mapping `// Equation (11)`

2.1 Ground visibility detector

Rather than a single image spectral cloud detector such as [Zhu et al., 2015], we use a *ground visibility detector* [Grompone von Gioi et al., 2020]. This is a simple and fast detector perfectly adapted to our problem. The difference is that in addition to the clouds, all elements of the scene that appear only once in the time series are detected as *invisible*, which means that they are unstable parts that should not be selected as pseudo-invariant features. This will be typically the case for cloudy pixels or for the waters of maritime shores or lakes, which may vary strongly in color and texture [Guindon, 1997, Du et al., 2002].

The ground visibility detector [Grompone von Gioi et al., 2020] works as follows: given a set of registered images, the ground visibility masks for the images are all initialized as *not visible*; then all pairs of images are compared locally. The comparison is performed on the angles of the image gradients, which means that it is insensible to affine modifications of the contrast. When a match is found, the corresponding parts are marked as *visible* in both image ground visibility masks. More specifically: for each pair of images, an image of angles difference is computed. These differences are small when the region is the same in the two images; on the other hand, they are large when the region has changed. The presence of new objects, clouds and water are the main reasons for large gradient angle differences. Match decisions are taken in regions where differences are small, which are obtained with a region growing algorithm. The *a contrario* statistical framework [Desolneux et al., 2000, Desolneux et al., 2008] is used to decide if the match is meaningful or not.

So as to reduce noise, we apply the detector on the average of the images' spectral bands. We call u_n the image at index n in the time series, and $u_{c,n}$ the channel c of this image. The gray-level image is obtained as:

$$\bar{u}_n = \frac{1}{C} \sum_{c=1}^{C} u_{c,n}, \qquad (1)$$

with C the number of channels. The ground visibility masks are then obtained as

$$\{m_1, ..., m_N\} = \text{ground visibility}\left(\{\bar{u}_1, ..., \bar{u}_N\}\right), \qquad (2)$$

with N the length of the time series. Images whose ground visibility is below 75% are removed from the series.

Figure 4 displays some results obtained with this ground visibility detector.

Temporal evolution of a patch taken in the input time series

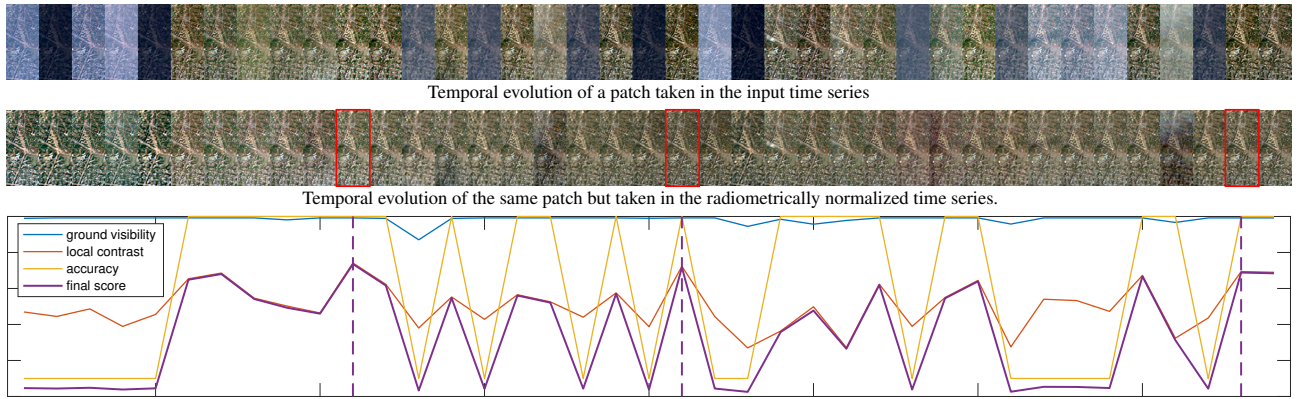Temporal evolution of the same patch but taken in the radiometrically normalized time series.

Figure 3. Value of the three metrics: proportion of visible ground S, local contrast C, accuracy A, and final score Q. The values are all in $[0, 1]$. A sample of the initial and stabilized time series (same series as in Figure 1) is displayed on the top and center rows, respectively. The key images found by the method are at index 11, 21 and 38 (over 39 images); they are highlighted in red in the normalized sequence and indicated with vertical dashed lines in the graph.
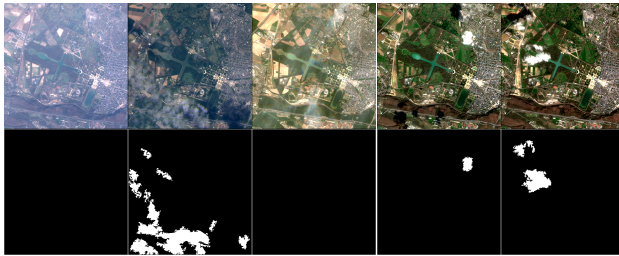

Figure 4. Results of the ground visibility detector [Grompone von Gioi et al., 2020]. Some of the input images from the time series are shown on the top row; corresponding ground visibility maps are shown below. Black means visible, white means invisible.

2.2 Detection of key images

Robust relative radiometric normalization of long time series requires several reference images. Our method relies on some well-chosen *key images*, selected as local maxima of the product of three metrics.

First, the **proportion of visible ground**, which is the proportion of pixels in the image that are not clouds, water or other factors rendering the pixel unreliable. This score is directly derived from the ground visibility detector:

$$S(u_n) = \frac{1}{|\Omega|} \sum_{x \in \Omega} m_n(x), \tag{3}$$

where Ω is the set of pixels in the image u_n. The mask m_n has value 1 when the pixel is visible ground, and 0 otherwise.

Second, the **local contrast**, measured as the local standard deviation:

$$C(u_n) = \frac{1}{|\Omega'|} \sum_{x \in \Omega'} \frac{\sqrt{(k * \bar{u}_n^2)(x) - (k * \bar{u}_n)^2(x)}}{\mathrm{std}(u_{\Omega'})}, \tag{4}$$

where k is a uniform kernel of size 15×15 pixels, and Ω' is the set of visible pixels in the image. That is, we measure the local contrast of the visible values only. This value is divided by the standard deviation of the image restricted to the visible ground pixel, namely, $\mathrm{std}(u_{\Omega'})$, so as to remain independent from the dynamic range of the input image. This measure is high when the contrast is located mostly in the high frequencies of the image. As a consequence, blurred and hazy images will have a low score.

Third, the **accuracy**. High resolution images are to be preferred, because they contain richer spectral information. The weights are arbitrary and favor Sentinel-2 L2A images.

$$A(u_n) = \begin{cases} 1 & \text{if Sentinel-2, Level-2A} \\ 0.1 & \text{if Sentinel-2, Level-1C and Landsat-8} \end{cases} \tag{5}$$

This is notably useful when Sentinel-2 and Landsat-8 images are mixed together.

These metrics are then multiplied to obtain a score Q for each image:

$$Q(u_n) = S(u_n) \cdot C(u_n) \cdot A(u_n). \tag{6}$$

Finally, we look for local maxima in this series of scores. An image at index n is declared a *key* if its score is superior to all other scores in a local time window of size $[n-w, n+w]$, where n is the index of the image in N images long time series and w is a parameter, by default set to 9.

The scores $S(u_n)$, $C(u_n)$, $A(u_n)$, and $Q(u_n)$ are displayed in Figure 3. The selected key images are framed in the time series and indicated with dashed vertical lines in the graph.

2.3 Pseudo-invariant features

Once the key images have been found, we use them as targets for the rest of the time series. We use the two closest keys to correct the spectral values of each image. We then obtain two set of parameters, which are linearly interpolated in function of the time difference in the series. However, we cannot rely on all pixels to estimate the correction. Most of the time, a large proportion of the observed scene has changed: apparition of new objects, changes in the aspect of the vegetation, humidity, clouds, are frequent sources of changes in the actual ground reflectance. Furthermore, changes in the angle of view or time of the day also modify the measured values. We need first to remove as many outliers as possible.

We compute the difference of angles between the gradients of the input and the key image. These angle differences are normalized in $[0, 1]$. They are averaged locally (by a uniform square kernel k of size 3×3) and then clipped to obtain a binary mask.

$$D(u, v) = [k * (|\mathrm{angles}(\nabla u, \nabla v)|/\pi)] < \tau, \tag{7}$$

where u is the current image and v the previous or next key. The second term in the convolution is an image of normalized gradient angles between u and v, *i.e.* with values in $[0, 1]$. The threshold τ is obtained so that D always contains 10% of the image pixels. In other words, the selected PIFs are the 10% pixels with the smallest angle difference. This ensures that we always have enough PIFs for the following regression step. We show in Figure 5 the aspect of these masks.

2.4 Robust regression strategy

As discussed earlier, we assume that affine effects dominate between two images from different dates and possibly from different sensors. This can be modeled as:

$$v_c = a_c \cdot u_c + b_c, \tag{8}$$

where c is the channel, u the input image, and (a_c, b_c) are the parameters of our affine model for this specific channel.

Denoting by u_D the input image restricted to its persistent pixels as computed in Section 2.3, we want

$$\arg\min_{a_c, b_c} \|v_{c,D} - a_c \cdot u_{c,D} - b_c\|, \tag{9}$$

with $\|\cdot\|$ a certain norm. An affine model is computed for each channel independently.

Knowing that we generally still have outliers at these points, we need a robust way of estimating the affine model's parameters. We thus opted for the RANSAC strategy [Fischler, Bolles, 1981], because it can handle a large proportion of outliers.

However, RANSAC would fail in estimating this model because images tend to have large areas with uniform colors. Indeed, if a large number of pixels are located on the same place, there are great chances that the other values will not impact the regression much. Thus, before applying RANSAC, we remove points in very dense regions, so as to reduce their weight when estimating the parameters of the affine model. The objective is to use points as uniformly disposed as possible in the dynamic range.

De-densification of the bi-temporal scattergram   We compute the histogram of the bi-temporal scattergram; this is a joint histogram of the source and key images. That is, we quantize the scattergram and count the number of points in each cell of the grid. Then, we randomly remove points in cells that are above a threshold. In practice we proceed in a faster and approximately equivalent way: We successively compute the histogram (using 100 bins) of the source and key images and each time remove at random points in bins whose count is above 3% of the total number of points.

Algorithm 2 summarizes the RANSAC procedure. We use the orthogonal regression (total least squares) [Markovsky, Van Huffel, 2007] in RANSAC. Such a technique is used when there are uncertainties in both variables. This amounts to use PCA on the inliers and keep the major axis.

We also tried to use the L^1 norm minimization and robust statistics, *e.g.*: $a_c = \mathrm{mad}\,\{v_c\}\,/\,\mathrm{mad}\,\{u_c\}$ and $b_c = \mathrm{median}\,\{v_c\} - a_c \cdot \mathrm{median}\,\{u_c\}$, but found that RANSAC was preferable, because it performs well even in the very challenging cases where the affine assumption is not respected, for example in presence of semi-transparent clouds that do not cover the whole image.

Algorithm 2: RANSAC: random sample consensus

input : source and target images, restricted to the persistent pixels, and de-densified
output : affine model parameters: $a_{\text{best}}, b_{\text{best}}$
$t \leftarrow 1000$ // initialize
while $t > 0$ **do**
  Select two points at random
  $(a, b) \leftarrow$ compute parameters of line passing through
  $d \leftarrow$ orthogonal distances from the points to the line
  $i \leftarrow d < 20\sigma$ // inlier set
  if $|i| > |i_{best}|$ **then**
    $(i_{\text{best}}, a_{\text{best}}, b_{\text{best}}) \leftarrow (i, a, b)$
    $i' \leftarrow i$
    do
      $(a', b') \leftarrow$ orthogonal regression(i')
      $d' \leftarrow$ orthogonal distances to the line
      $i' \leftarrow d' < 20\sigma$ // refined inlier set
      if $|i'| > |i_{best}|$ **then**
        $(i_{\text{best}}, a_{\text{best}}, b_{\text{best}}) \leftarrow (i', a', b')$
      $t \leftarrow t - 1$
    while $|i'| > |i_{best}|$ **and** $t > 0$
  $t \leftarrow t - 1$

Noise estimation   The RANSAC method requires to establish a threshold for the definition of its inlier set, which is usually related to the uncertainty or the noise of the data points. The sources of uncertainty in our model are: the image noise and the model deviations: misalignment, non-lambertian surfaces, inhomogeneity of the atmosphere, *etc*. We can determine the noise present on each image σ_n using the Ponomarenko noise estimation algorithm [Ponomarenko et al., 2007, Colom, Buades, 2013]. Then, in order to account for the model deviations we set the inlier threshold to 20σ, with $\sigma = \mathrm{median}(\{\sigma_n \,|\, n = 1...N\})$, and σ_n the noise standard deviation of image u_n.

We show in Figure 5 scatterplots of the bitemporal pairs of source and target images. The affine model found by RANSAC is traced for each channel. The bright colored points are the inliers found by RANSAC, while the rest of the colored points are the rest of the PIFs. The points in gray represent values removed either by the difference of angles measure, or the de-densification step. Remark that most points that do not correspond to an affine relation between the input and reference images have been removed by these masks. This indicates that the angle measure succeeded in removing a large part of the outliers.

2.5 Dynamic range compression

So far, we have stabilized the contrast of the time series, but its dynamic range has not been altered. Since most applications assume images in the typical 8-bits dynamic range, a strong tone-mapping step is required. This is also useful for human interpreters who visualize the images on standard screens with limited dynamic range.

Tone-mapping can be either global or local. The latter is generally able to perform a stronger compression of the dynamic, yet the former is generally simpler and generates fewer artifacts. We use a simple linear compression of the dynamic with a slight gamma-correction. The minimum and maximum values of the tone-mapped time series, that will be mapped to 0 and 255 respectively, are computed so as to clip the minimal number of pixels but also to avoid strong compression of the contrast. We use the median 1st and 99th percentiles, respectively. That is:
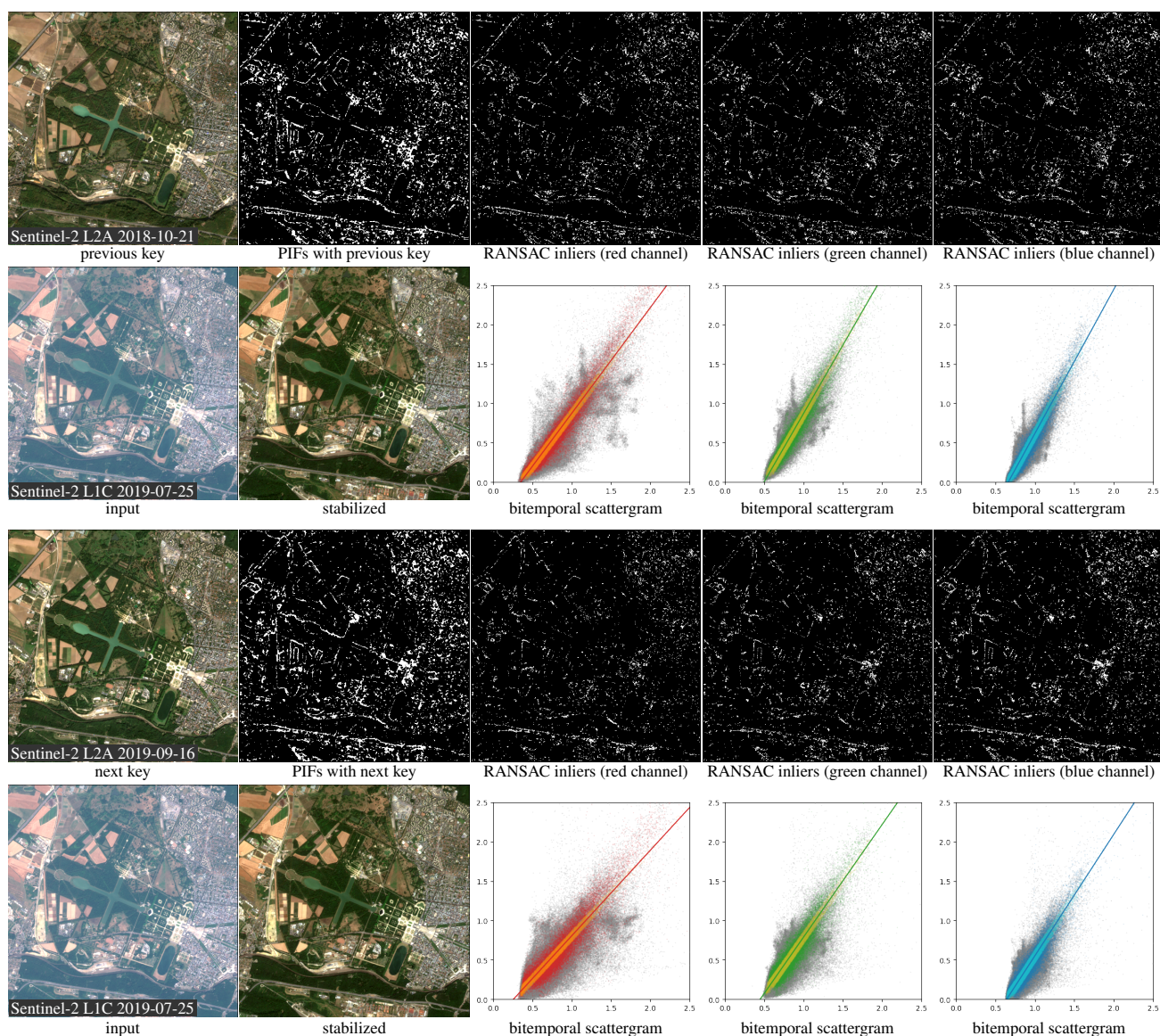
Figure 5. The top two rows show the pair (input, previous key), and the bottom two rows the pair (input, next key). Next to the keys are displayed the pseudo-invariant features obtained with the angle difference. The three columns on the right show the bitemporal scattergrams for each channel and the RANSAC inliers. The gray values of the scattergram represent points excluded by the mask. The colored values are the PIFs, and among them, the bright ones are the RANSAC inliers, also displayed in the images above.
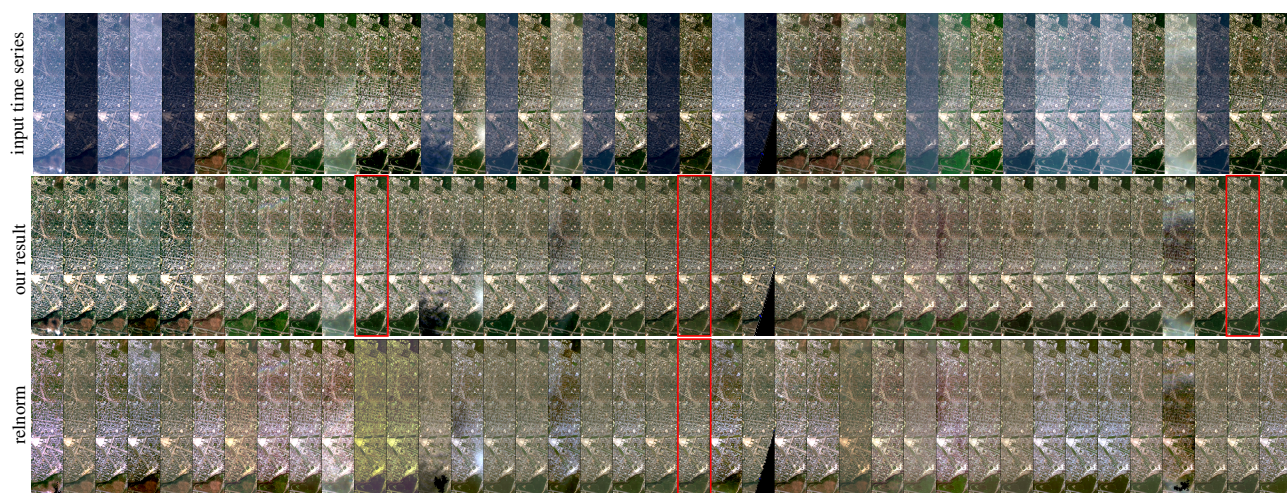

Figure 6. Top row: inputs, with Sentinel-2 (L1C, L2A) and Landsat-8 images. Center row: our relative radiometric normalization result. Bottom row: results with the method "relnorm" [Goslee et al., 2011].

Table 1. Pixels' temporal standard deviations, summarized by quantiles; "Versailles" time series. Series are normalized by their standard deviation before the measure. The lower the better.

Quantile	Input	Naive	Relnorm	Ours
0.25	0.4895	0.2934	0.3070	**0.1833**
0.50	0.5375	0.3387	0.3756	**0.2547**
0.75	0.6610	0.4290	0.4644	**0.3897**

$$\begin{cases} \beta^{\min} = \text{median}(\{\text{percentiles}(\bar{u}_n; 1) \,|\, n = 1 \ldots N\}) \\ \beta^{\max} = \text{median}(\{\text{percentiles}(\bar{u}_n; 99) \,|\, n = 1 \ldots N\}) \,. \end{cases} \tag{10}$$

This is computed using the average of the channels, denoted by $\bar{u}$. Finally, the tone-mapped image z is obtained after clipping and application of a slight gamma-correction:

$$z_c = \text{clip}_{[0,1]}\big((u_c - \beta^{\min})/(\beta^{\max} - \beta^{\min})\big)^{\frac{3}{4}} \,. \tag{11}$$

All images shown in this paper are obtained this way.

3. EVALUATION AND COMPARISONS

Figure 1 and 2 display the result of our method for a 1.5 year long time series made of 29 Sentinel-2 images (7 with correction level L1C, 22 with L2A) and 10 Landsat-8 images. Figure 5 shows the pseudo-invariant features (PIFs) obtained for two couples of images. The accompanying scatterplots show the estimated affine translations for each channel. The input image, which contains haze, is corrected.

Figure 6 shows a rectangular detail of the "Versailles" series and the normalized result. This is an urban area whose color should remain roughly constant over time. This series covers one year and a half, from January 2018 to September 2019. In the normalized version, changes in vegetation color are visible on the bottom of the patch, but the town region is well stabilized. Images selected as key are framed in red; they were captured at dates 2018/06/25, 2018/10/21, and 2019/09/16.

We compared our method with the "relnorm" function from the "landsat" R package[2]. This method was proposed by Goslee *et al.* in [Goslee et al., 2011]. Results are presented in Figure 6, last row. It needs a manually selected reference image; we used the one at index 21. It is framed in red. We passed as cloud masks the result of our ground visibility detector. The method used for the regression was set to the recommended "Major Axis". Our method is significantly more stable in this example.

We measured an execution time of 7min for our algorithm for this series of 39 images. This is more than $50\times$ faster than "relnorm", that required around 398min for the same sequence. Tests were carried out on an 3.1 GHz Dual-Core Intel Core i7.

Table 1 quantitatively evaluates the stability of our results. We proceeded as follows: first, we normalized each evaluated time series by its global standard deviation. This makes the measure independent of the output dynamic range. Then we removed a local temporal average of the pixel's color, so as to discard variations due to seasonal changes. We averaged seven dates, *i.e.*, approximately three months in the tested series. Finally, the pixels' standard deviation across time is measured. The final result is the average of the three channels. We summarized the results for the "Versailles" time series (used in Figure 1 to 6) with three quantiles. The "naive" method is the independent

---

[2] `https://rdrr.io/github/phiala/landsat/`
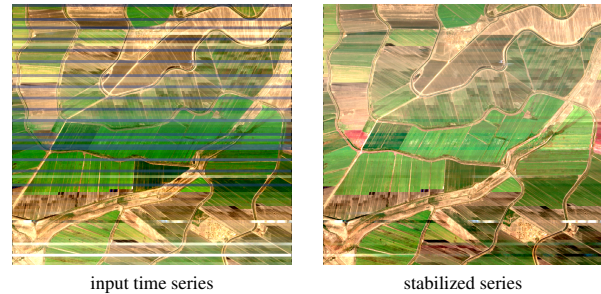

input time series   stabilized series

Figure 7. Composite images made in the same way as Figure 1. Time series with 93 images, 18 Landsat-8 and 75 Sentinel-2 L2A, from 2019/01 to 2020/05 in the region of Rabak (Sudan).

normalization of each image by removing its mean and dividing by its standard deviation. Our method gives a remarkably more stable series.

Figure 7 shows the result of our relative radiometric normalization on another sequence. This scene is more difficult because it is mainly composed of agricultural crops with many seasonal changes. Our method nonetheless produces a stable result in which radiometric changes are due to changes in the ground reflectance. The crops change but the tracks are stable.

Limitations Images with non homogeneous transparent clouds (haze and cirrus) are difficult to handle, because they do not respect the assumption of a global affine modification of the radiometric values. Even if the RANSAC method finds good parameters in most cases, some images are not perfectly normalized, for example the fourth image from the right in Figure 6. Another limitation lies in the selection of PIFs based on the angle of the image gradients. Textured regions, although they could also be relevant in the regression, tend to be rejected because their gradients are not stable temporally.

4. CONCLUSION

We presented a pipeline for the relative normalization of multi-satellite image time series. We tested our method with freely available Sentinel-2 and Landsat-8 images. The results demonstrate good temporal stabilization. The key and most challenging parts of our method are first, to automatically select good reference images, second, to sort out the persistent pixels between each image and its reference, third, to robustly estimate the affine correction parameters. Using quality metrics, similarity masks based on the angles of images gradients, and a random sample consensus method, we succeeded in obtaining an algorithm that gives plausible results in any situation.

As a most basic utilization, this algorithm can fill missing L2A images in Sentinel-2 time series using the L1C, which are always available. A more sophisticated usage, demonstrated in this paper, is the fusion of time series from different satellites. Our algorithm does not need external information, and makes no assumption on the sensors. This means that our method can directly be used with images from other satellites.

We expect to extend this algorithm to make an online version. Indeed, for the moment we require the full time series to find the reference images, compute the affine correction models. However, it is desirable to have a method that allows to add new images to an already stabilized time series.

The code of our method is available online at `https://github.com/chlsl/rrn-multisensor-multidate`.

ACKNOWLEDGMENTS

REFERENCES

Briand, T., Davy, A., 2019. Optimization of Image B-spline Interpolation for GPU Architectures. *Image Processing On Line*, 9, 183–204.

Canty, M. J., Nielsen, A. A., 2008. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. *Remote Sensing of Environment*, 112(3), 1025–1036.

Canty, M. J., Nielsen, A. A., Schmidt, M., 2004. Automatic radiometric normalization of multitemporal satellite imagery. *Remote sensing of Environment*, 91(3-4), 441–451.

Carvalho Júnior, O. A. et al., 2011. A new approach to change vector analysis using distance and similarity measures. *Remote Sensing*, 3(11), 2473–2493.

Colom, M., Buades, A., 2013. Analysis and Extension of the Ponomarenko et al. Method, Estimating a Noise Curve from a Single Image. *Image Processing On Line*, 3, 173–197.

Coppin, P. R., Bauer, M. E., 1996. Digital change detection in forest ecosystems with remote sensing imagery. *Remote sensing reviews*, 13(3-4), 207–234.

De Carvalho, O. A. et al., 2013. Radiometric normalization of temporal images combining automatic detection of pseudo-invariant features from the distance and similarity spectral measures, density scatterplot analysis, and robust regression. *Remote Sensing*, 5(6), 2763–2794.

Desolneux, A., Moisan, L., Morel, J.-M., 2000. Meaningful Alignments. *Int. J. Comput. Vis.*, 40(1), 7–23.

Desolneux, A., Moisan, L., Morel, J.-M., 2008. *From gestalt theory to image analysis: a probabilistic approach*. 34, Springer.

Drouyer, S., 2020. Parking occupancy estimation on planetscope satellite images. *IGARSS*.

Du, Y., Teillet, P. M., Cihlar, J., 2002. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. *Remote Sens. Environ.*, 82(1), 123–134.

Fischler, M. A., Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.

Foroosh, H., Zerubia, J. B., Berthod, M., 2002. Extension of phase correlation to subpixel registration. *IEEE Trans. Image Process.*, 11(3), 188–200.

Goslee, S. C. et al., 2011. Analyzing remote sensing data in R: the landsat package. *J. Stat. Softw.*, 43(4), 1–25.

Grompone von Gioi, R., Hessel, C., Dagobert, T., Morel, J. M., de Franchis, C., 2020. Temporal repetition detection for ground visibility assessment. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*.

Guindon, B., 1997. Assessing the radiometric fidelity of high resolution satellite image mosaics. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(5), 229–243.

Hall, F. G., Strebel, D. E., Nickeson, J. E., Goetz, S. J., 1991. Radiometric rectification: toward a common radiometric response among multidate, multisensor images. *Remote sensing of environment*, 35(1), 11–27.

Jianya, G., Haigang, S., Guorui, M., Qiming, Z., 2008. A review of multi-temporal remote sensing data change detection algorithms. *The Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B7), 757–762.

Kauth, R. J., Thomas, G., 1976. The tasselled cap–a graphic description of the spectral-temporal development of agricultural crops as seen by landsat. *LARS symposia*, 159.

Lin, C.-H. et al., 2015. Radiometric normalization and cloud detection of optical satellite images using invariant pixels. *ISPRS journal of photogrammetry and remote sensing*, 106, 107–117.

Markovsky, I., Van Huffel, S., 2007. Overview of total least-squares methods. *Signal processing*, 87(10), 2283–2302.

Nielsen, A. A., Conradsen, K., Andersen, O. B., 2002. A change oriented extension of EOF analysis applied to the 1996–1997 AVHRR sea surface temperature data. *Physics and Chemistry of the Earth, Parts A/B/C*, 27(32-34), 1379–1386.

Nielsen, A. A., Conradsen, K., Simpson, J. J., 1998. Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: New approaches to change detection studies. *Remote Sens. Environ.*, 64(1), 1–19.

Paolini, L., Grings, F., Sobrino, J. A., Jiménez Muñoz, J. C., Karszenbaum, H., 2006. Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies. *International Journal of Remote Sensing*, 27(4), 685–704.

Ponomarenko, N. N., Lukin, V. V., Zriakhov, M. S., Kaarna, A., Astola, J., 2007. An automatic approach to lossy compression of AVIRIS images. *Geoscience and Remote Sensing Symposium*, IEEE, 472–475.

Rahman, M., Hay, G., Couloigner, I., Hemachandran, B., Bailin, J., 2015. A comparison of four relative radiometric normalization (RRN) techniques for mosaicing H-res multi-temporal thermal infrared (TIR) flight-lines of a complex urban scene. *ISPRS J. Photogramm.*, 106, 82–94.

Syariz, M. A. et al., 2019. Spectral-consistent relative radiometric normalization for multitemporal Landsat 8 imagery. *ISPRS J. Photogramm.*, 147, 56–64.

Xu, Q., Hou, Z., Tokola, T., 2012. Relative radiometric correction of multi-temporal ALOS AVNIR-2 data for the estimation of forest attributes. *ISPRS J. Photogramm.*, 68, 69–78.

Yuan, D., Elvidge, C. D., 1996. Comparison of relative radiometric normalization techniques. *ISPRS Journal of Photogrammetry and Remote Sensing*, 51(3), 117–126.

Zhu, Z., Wang, S., Woodcock, C. E., 2015. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, 159, 269–277.