

Analogue-based colorization of remote sensing images using textural information



Mathieu Gravey*, Luiz Gustavo Rasera, Gregoire Mariethoz

University of Lausanne, Faculty of Geosciences and Environment, Institute of Earth Surface Dynamics, Switzerland

ARTICLE INFO

Keywords:
Geostatistics
Satellite image
Multi-point
Spectral enhancement

ABSTRACT

Satellite images are richer than ever before. For example, new Landsat-8 images with their 11 bands carry much more information than older generations of satellites. These differences in spectral representation imply a major difficulty for assessing long-term land surface changes. The easiest solution is to reduce the information of the most recent product, for example by only keeping a subset of the Landsat-8 bands that matches old imagery. To avoid such loss of information, we propose a new method based on multiband spatial pattern matching. We are focusing on increasing the spectral resolution of archive satellite images to the same level of spectral resolution and coverage as modern imagery. Our method uses analogous scenes taken from modern satellites, which have conceptually the same role as the training images used in multiple-point geostatistics simulation. The spectral characteristics of the training image are then transferred to a target archive image, where new synthetic spectral bands are generated. A spatial pattern matching procedure is used to control this transfer, resulting in preservation of spatial and spectral coherence in the results. We illustrate the methodology on Landsat 8 and Corona imagery. The proposed method was benchmarked against other state-of-the-art colorization techniques, and it shows globally better results.

1. Introduction

Multispectral images play a crucial role in a variety of scientific fields, ranging from geoscience to urban planning and ecology. Their spectral information enables using classification (Bruzzone and Demir, 2014) or object detection (Cheng and Han, 2016) to differentiate Earth surface properties such as land cover types (Vaduva et al., 2013) and, crucially, allows monitoring changes in those properties.

In the last decade, archive images have become available through major change in data policies, such as the release of formerly classified imagery (USGS, 2008) and the waiving of restrictions for the civil use of very fine-resolution imagery. In addition, access to archive images has dramatically improved thanks to free online portals like those of the USGS, the ESA, and platforms such as Google and Amazon. As a result, multi-temporal information is now widely used for remote sensing change detection. For example, such data make it possible to measure urban growth (Weng, 2001), to determine the human impact on rainforests (Hansen et al., 2008), and the evolution of water bodies (Pekel et al., 2016). For such applications, historic remote sensing data are invaluable because they allow quantifying long-term changes induced by climate change or by direct human activity (Lu et al., 2004).

Despite the above advances in the possibilities offered by Earth

observation, a significant remaining difficulty for long-term change detection is the inconsistency between different satellite sensors, which are due to the rapid evolution of sensors and the limited lifetime of each platform. The oldest continuously available satellite data set dates back to only about 50 years, with Landsat 1 launched in 1972 that had only 4 spectral bands (Zanter, 2005). Today, Landsat 8 has 11 bands (Zanter, 2005), and other platforms propose up to 220 bands such as the Hyperion sensor on EO-1 (Beck, 2010). Pre-Landsat imagery is available, albeit with limited and infrequent coverage and often restricted to panchromatic information, starting from 1960 with the Corona spy imaging program (USGS, 2008). This great variability in the information content of multi-temporal images makes their analysis difficult (Yang et al., 2013). The changes in characteristics between the different generations of products particularly affect studies related to change detection, which cannot use the full potential of satellite archives. As a result, in most cases satellite-based long-term change detection studies use post-1984 data, because 1984 is the launch year of Landsat-4 and its modern Thematic Mapper. Among many examples, a few specific papers that face this problem are for the study of long-term changes in Arctic vegetation (Fraser et al., 2011), for the detection of changes in Mediterranean urban and vegetation cover (Alphan and Yilmaz, 2005), for wetland monitoring in Zambia (Munyati, 2010) or for assessment of

* Corresponding author.

E-mail address: mathieu.gravey@unil.ch (M. Gravey).

the tourism boom in Egypt (Dewidar, 2002). The temporal range of such studies would be greatly increased if they could include images dating from as early as 1972.

This paper focuses on the enrichment of old remote sensing images such that they become comparable with present-day acquisitions. While old and recent images may differ in terms of spatial, temporal and spectral resolutions, we only focus here on the spectral aspect. In order to compare time-series images that have different spectral bands, or different bands definitions, one needs to find a common reference frame where both old and new images can be represented and compared. An usual and straightforward approach is to decimate the data to a common denominator by removing the specific spectral bands that are missing in the oldest part of the collection, or to compute indices that are easier to compare (Cohen et al., 2002). Nevertheless, this approach is suboptimal because the information from the most recent sensor is not used. Here, we investigate the possibility of enhancing the spectral characteristics of the old images such that they become comparable with more recent ones for studies of land surface changes.

The premise of the approach proposed in this paper is that spectral enhancement of remote sensing data can be treated as an image colorization problem. Given an older image with a limited number of spectral bands (the target image), we assume that we have another, more recent, and spectrally richer image of similar surface characteristics (the training image). Patterns are defined that characterize the spectral signature of pixels, as well as in their spatial neighborhood (we call those spatio-spectral patterns). Then, based on the training image, we identify correspondences by matching the spatio-spectral patterns, and these correspondences are used to import spectral signatures from the training image to the target image (Fig. 1).

The training image can be an analogue image with similar surface features to the target domain, or a recent image of the area of interest that contains all bands information.

When spectrally enhancing satellite sensor images, two possible situations can be encountered, as illustrated in Fig. 2:

- Spectral disaggregation: when the information present in the old image integrates all bands of interest, but in a convoluted representation. This is, for example, the case when the panchromatic

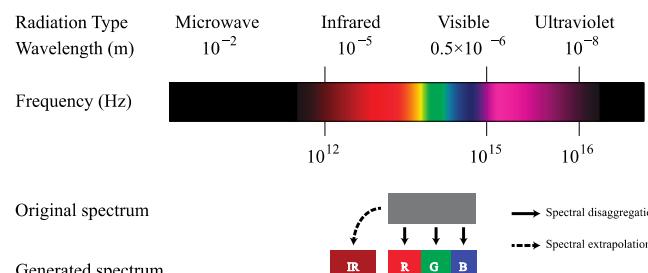


Fig. 2. The difference between spectral disaggregation and spectral extrapolation.

band is available in an old image and has to be disaggregated to the three RGB color bands. In this example, the panchromatic band is defined as the integral of all three-color wavebands.

- Spectral extrapolation: when the target band is not contained at all in the old image. A good example of this situation is when the panchromatic information is available, and the near-infrared (NIR) is the target band to recover. The NIR band corresponds to wavelengths that are not contained in the panchromatic band. However, extrapolating the NIR information could potentially have applications, such as computing the normalized difference vegetation index (NDVI). Logically, spectral extrapolation tends to be more challenging than spectral disaggregation.

In both cases, a naive approach would consist of estimating the spectral response in an unknown band based on the values of the known bands, using an assumption of continuity in the spectral signature of the observed surface. However, this is often not appropriate because the spectral response of a given object can be completely different in each spectral band. Therefore, one cannot simply interpolate band values across the electromagnetic spectrum. The approach we propose in this paper for spectral disaggregation does not make assumptions on the continuity of the spectral signatures, and instead relies on analogies to extend the spectral signatures of historical remote sensing images. This paper focuses on both issues of spectral disaggregation and

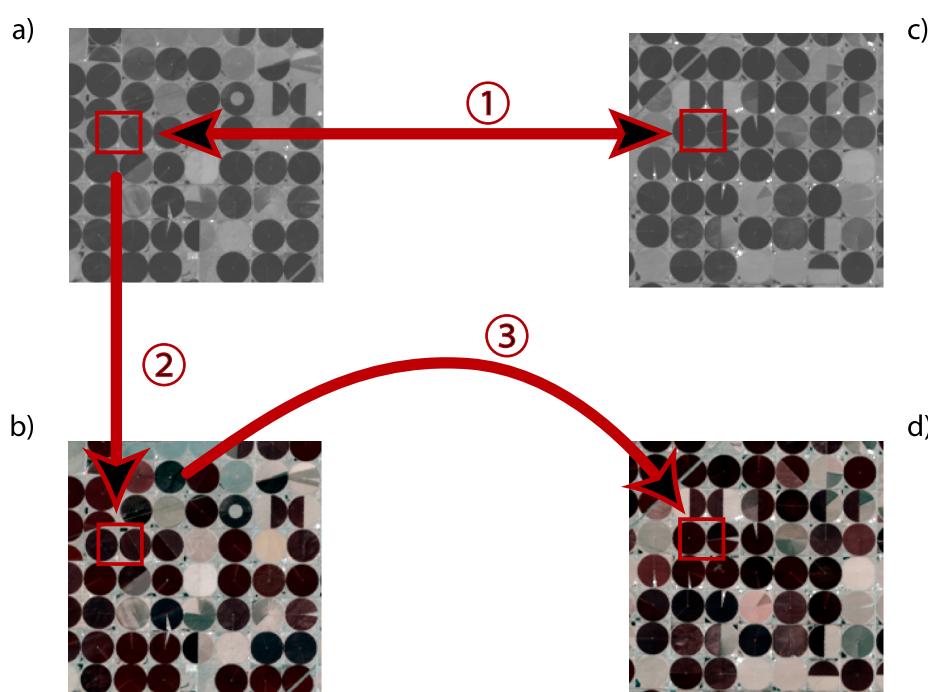


Fig. 1. The principle of image colorization based on a training image. (a) grayscale training image, (b) color training image associated to (a), (c) target grayscale image to colorize, (d) colorized output of (c) using (a) and (b). ① spatial pattern matching, ② color identification and ③ color import.

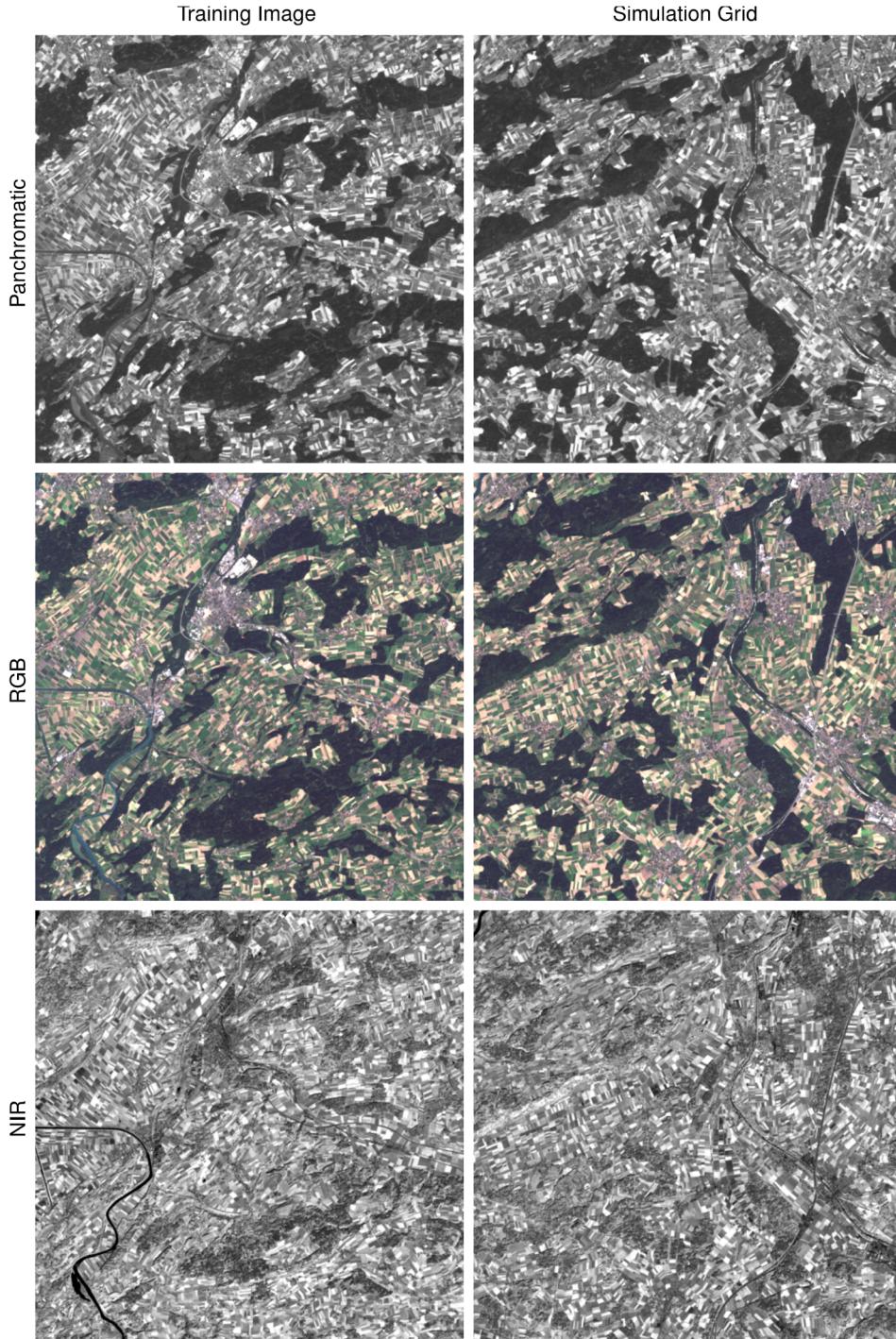


Fig. 3. Training Image and reference scenes in Panchromatic (PAN), RGB and NIR. Landsat id: LC81950272015242LGN00.

extrapolation. It studies, for the first time, the possibility of spectral enhancement of remote sensing images based on training examples. Several algorithms coming from both computer graphics and multiple-point statistics are tested in the context of spectral enhancement, and a new approach that outperforms these classical algorithms is proposed.

This remaining of the paper is structured as follows. We start reviewing related problems in other fields and the different solutions that have been proposed. We then propose our algorithm and compare the results with other approaches. Finally, after discussing the results, we present guidelines to obtain optimal results with minimal computational cost.

2. Training-based colorization algorithms

In this section, we first explore the application of the work done by [Welsh \(2002\)](#), which proposes a training-based colorization approach of grayscale images by using pattern-matching applied independently to each pixel. As will be discussed later, this method has limitations due to the absence of spatial dependence between the simulated data. We then investigate the Direct Sampling multiple-point statistics method ([Mariethoz et al., 2010](#)), which has the particularity of taking into account the spatial dependence between simulated values. Finally, we propose an improvement on the aforementioned methods with the use of a guided simulation path that allows reducing the prediction

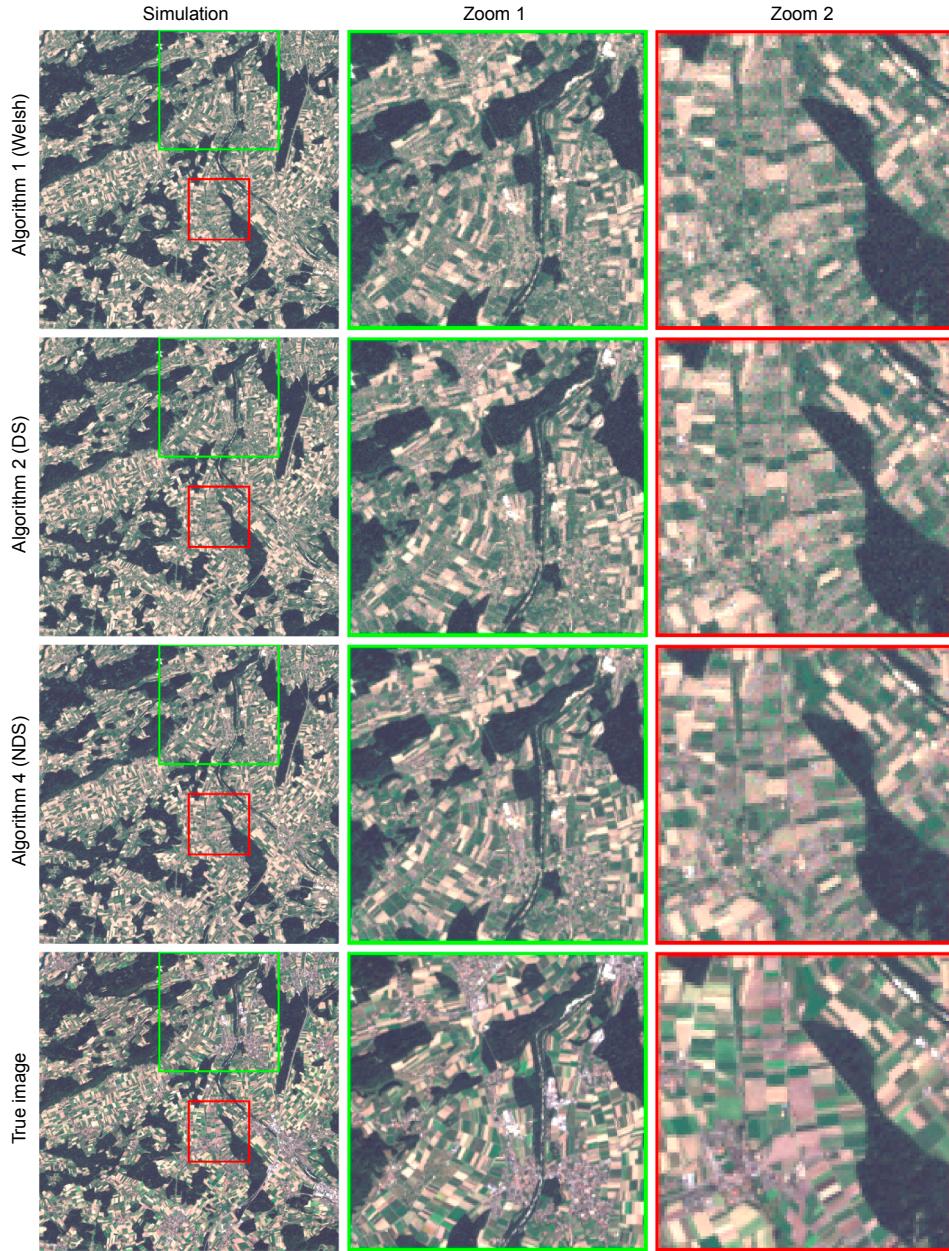


Fig. 4. Spectral disaggregation from PAN to RGB for all three algorithms.

uncertainty and improving the overall results. The guided simulation path is a variant of the strategy originally proposed by Liu and Journel (2004) in the context of geostatistical simulation of categorical variables.

2.1. General pattern matching strategy

All algorithms investigated in this paper use a look-up approach in a training image as a basis to generate new data. Given a target pattern to colorize, they identify another pattern in the training image from which to import the missing spectral information. In all cases, this operation is done by using spatial pattern matching, possibly in a large database of multi-temporal training images. In consequence, the algorithms described in the coming sections will largely use the concept of pattern matching, which require a few preliminary definitions and notations:

\mathbf{u} denotes a position in the training image (the source image with complete multispectral information), and \mathbf{v} denotes a position in the target image (the old image with only a limited number of bands being informed). \mathbb{B} is the set of all spectral bands considered. $\wp \in \mathbb{R}^2$ represents the extent of a pattern, which is defined by a series of lag vectors $\mathbf{h} \in \wp$ that describe the relative position between the central position of the pattern (\mathbf{u} in the training image or \mathbf{v} in the target image) and each position in the pattern. $w_i(\mathbf{h})$ is a kernel function that weights the importance of a matching error for a given lag $\mathbf{h} \in \wp$ and band $i \in \mathbb{B}$.

Each value in an image is defined by a spatial position and a spectral band. We denote $B_i(\mathbf{x})$ the value from the band $i \in \mathbb{B}$ at the position \mathbf{x} . By extension, $B(\mathbf{x})$ is the vector containing all spectral values at position \mathbf{x} . A neighborhood $N_i(\mathbf{x})$ around \mathbf{x} is defined as the association of the lag vectors and the band values from \wp where a value exists. It is denoted:

Table 1

Disaggregation error for each algorithm, shown separately for the Blue, Green and Red bands. For RMSE, smaller is better, and for CC and SSIM closer to 1 is better. The best result for each metric is shown in bold.

		RMSE	CC	SSIM
Algorithm 1 (Welsh)	Blue	0.1808 ± 0.0049	0.9251 ± 0.0036	0.7292 ± 0.0005
	Green	0.1719 ± 0.0020	0.9581 ± 0.0010	0.7462 ± 0.0008
	Red	0.2063 ± 0.0019	0.9602 ± 0.0007	0.7542 ± 0.0006
Algorithm 2 (DS)	Blue	0.1635 ± 0.0076	0.9378 ± 0.0050	0.7577 ± 0.0011
	Green	0.1621 ± 0.0032	0.9629 ± 0.0013	0.7681 ± 0.0007
	Red	0.1925 ± 0.0032	0.9655 ± 0.0011	0.7765 ± 0.0006
Algorithm 4 (NDS)	Blue	0.1385 ± 0.0053	0.9546 ± 0.0033	0.8259 ± 0.0010
	Green	0.1338 ± 0.0018	0.9744 ± 0.0007	0.8430 ± 0.0004
	Red	0.1542 ± 0.0017	0.9778 ± 0.0005	0.8447 ± 0.0007

$$N_i(\mathbf{x}) = \{\mathbf{h} \in \wp | B_i(\mathbf{x} + \mathbf{h}) \text{ is known}\} \quad (1)$$

By extension $N(\mathbf{x})$ is the set of $N_i(\mathbf{x})$ for all possible bands i . As $N_i(\mathbf{x})$ contains both spectral and spatial information, we call it a spatio-spectral pattern, and it is the entity used for defining matches between training and target images.

Typically, pattern matching involves computing a mismatch between a pattern found in the target image and another pattern coming from the training image. The mismatch between two patterns can be measured in a variety of ways (e.g. structure of the pattern, position of the information). In this paper we use a combination of pixel-wise errors, which we denote $\varepsilon(\mathbf{u}, \mathbf{v})$. It is computed as follows:

$$\varepsilon(\mathbf{u}, \mathbf{v}) = \sqrt[\beta]{\sum_{i \in B} \sum_{h \in N_i(u) \cap N_i(v)} w_i(\mathbf{h}) |B_i(\mathbf{u} + \mathbf{h}) - B_i(\mathbf{v} + \mathbf{h})|^\beta} \quad (2)$$

Many possible functions can be used for the kernel w (Silverman, 1986). In this paper we use an exponential function $w_i(\mathbf{h}) = e^{-\alpha_i \|\mathbf{h}\|_2}$ where α_i controls the shape of the kernel and is specific to each band i . $\alpha \rightarrow 0$ results in uniform weights for all neighbors. $\alpha \rightarrow \infty$ only considers the central pixel, meaning that the spatial dependence between neighboring pixels is ignored. β is typically 1 or 2.

2.2. The computer graphics approach

For over a decade, the computer graphics community has faced the need to colorize grayscale images or movies, and several approaches and algorithms were developed to this end. One of the simplest methods consists in assigning a single color for each grayscale value using a mapping table (Pratt, 2007). Charpiat et al. (2009) determines the color probability at each location using previously learned gradients on a training image and then use a graph cut approach to generate a colorization having the highest global coherence. Another approach uses deep convolutional neural networks with a training phase from a large image bank, and then predicts the color channel by channel (Hwang and Zhou, 2016). Semary et al. (2007) use segmentation techniques to separate images in homogeneous areas, which are then colorized independently. Levin et al. (2004) propose to use seed points where the color is set manually by the user. The color is then diffused based on the grayscale gradient, which has the effect of propagating it along smooth areas. Noda et al. (2006) propose to formulate the problem as a Bayesian inference, and perform the colorization using maximum a posteriori estimation. So far, these methods have only been used for the colorization of grayscale images to visible color (RGB), and mostly applied to photography. Furthermore, many of these approaches work only for a given color space, and may fail when dealing with spectral bands other than RGB.

Of particular interest, Welsh (2002) proposed an algorithm that uses a training image where the color and the grayscale information are known. For each pixel \mathbf{v} of the target, a neighborhood $N_i(\mathbf{v})$ is considered in a window of 5 by 5 pixels. The algorithm then searches for a similar location in the training image by pattern matching (based on computing $\varepsilon(\mathbf{u}, \mathbf{v})$ for 200 randomly sampled locations \mathbf{u}). The color at the location with lowest mismatch is then transferred to the target image.

This approach is interesting for processing old satellite products in at least two aspects:

- (1) It uses a training image deemed similar to the target image, which can be used to retrieve the missing multispectral information.
- (2) It exploits the spatial information contained in the grayscale spatial patterns to inform the missing spectral bands. The underlying assumption is that of a correspondence between a given spatial signature and a given color (or multispectral signature).

While Welsh (2002) did tests on satellite sensor data, these were

Table 2

Extrapolation error for all three algorithms, shown separately for the PAN → NIR and for RGB → NIR. For RMSE, smaller is better, and for CC and SSIM closer to 1 is better. The best result for each metric is shown in bold.

			RMSE	CC	SSIM
Algorithm 1 (Welsh)	PAN → IR	0.8712 ± 0.0019	0.3555 ± 0.0017	0.1097 ± 0.0009	
	RGB → IR	0.4886 ± 0.0006	0.7763 ± 0.0005	0.4072 ± 0.0008	
Algorithm 2 (DS)	PAN → IR	0.7747 ± 0.0021	0.4009 ± 0.0028	0.1202 ± 0.0015	
	RGB → IR	0.4485 ± 0.0007	0.8048 ± 0.0007	0.4238 ± 0.0014	
Algorithm 4 (NDS)	PAN → IR	0.7364 ± 0.0018	0.4609 ± 0.0048	0.1589 ± 0.0020	
	RGB → IR	0.4110 ± 0.0008	0.8376 ± 0.0006	0.4733 ± 0.0014	

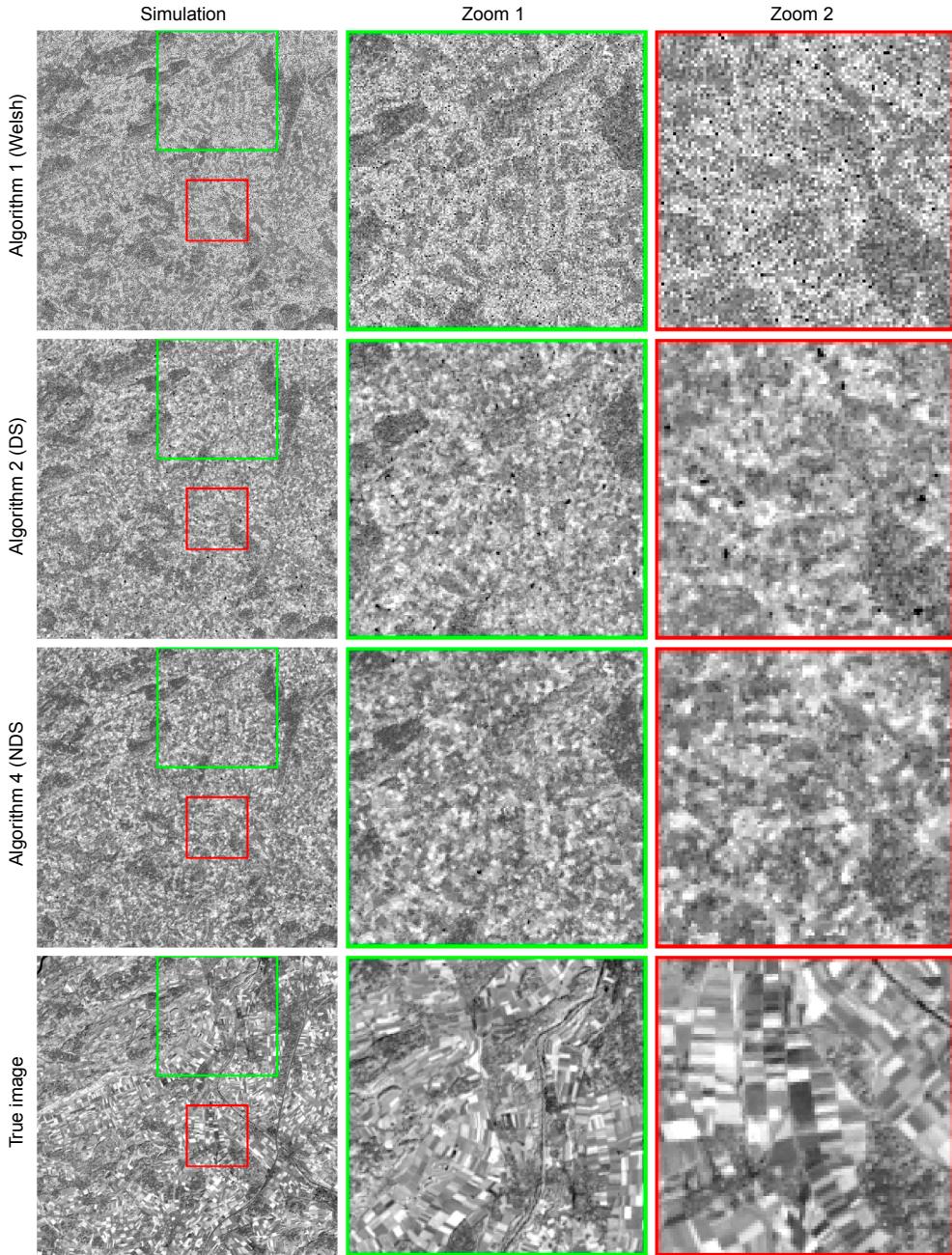


Fig. 5. Spectral extrapolation PAN → NIR for all three algorithms.

very limited and restricted to the colorization from grayscale to color images (i.e. not considering multispectral images). In this paper, we use this approach as a benchmark against other algorithms. The pseudocode of the algorithm of Welsh (2002) is given in Algorithm 1.

Pseudocode of Algorithm 1 (modified implementation of Welsh (2002))

- 1 Welsh_Simulation (training image, target image, threshold t , exploration ratio f)
- 2 **For** each pixel $B(v)$ of the target image, **do independently**
- 3 Extract the neighborhood $N(v)$ of the target image around $B(v)$
- 4 Find a candidate $B(u)$ in the training image that satisfies $\varepsilon(u, v) < t$ by scanning a fraction f of the training image
- 5 Import the color from the candidate $B(u)$ at location $B(v)$ in the target image.
- 6 **End**
- 7 **Return** target image

Note that this algorithm is extremely easy to parallelize since each pixel is simulated independently.

For computational reasons, the original algorithm of Welsh (2002) searches for a matching pattern over a limited subset of the training image. Here, we modified it by using a strategy that keeps searching until a candidate with a mismatch under a given threshold is found. Therefore, if no candidate is found with an error below t , best candidate is accepted. Once a candidate is found, the values of the additional bands are imported from the training image to the target image.

2.3. The geostatistical approach

In the field of geostatistics, the multiple point simulation approach (MPS) uses the concept of analogues, which is similar as the training image used by Welsh (2002). The goal of MPS is to generate random fields of unobserved variables, based on a training image, which represent a similar process as the one to model. In its most general

application, MPS produces images by resampling patterns found in the training image and assembling them in a coherent way such as to obtain a new image that is either entirely random, or conditioned to a small number of known measurements (Guardiano and Srivastava, 1993).

There have been applications of MPS in different fields, including weather generation (Jha et al., 2015; Oriani et al., 2017), subsurface modeling (Kessler et al., 2012; Pirot et al., 2014; Strebelle, 2002), and notably in remote sensing with downscaling (Boucher, 2009; Tang et al., 2015), classification (Ge, 2013; Ge and Bai, 2011) or gap-filling of Landsat-7 images (Yin et al., 2017). Recent developments have extended the approach to condition the produced simulations to one or several exhaustively known covariates, thus extending its use for new applications such as geophysics, where indirect information is commonly available throughout the studied domain (Lochbühler et al., 2013). Over the last few years, several methods have been developed, aimed at simulating both categorical and continuous variables. Among the approaches suited for the simulation of continuous variables, Direct Sampling (DS) (Mariethoz et al., 2010) is interesting in the context of multispectral images because it allows multivariate simulation, that is, using training images consisting of multiple co-registered variables. This functionality can potentially be used for colorizing grayscale or multispectral images by considering each spectral band as a covariate.

The major difference between DS and the approach of Welsh is its sequential character. The DS algorithm first generates a path P defining the order in which pixels will be simulated sequentially. Then, the value of each pixel is simulated conditionally to the previous ones based on a neighborhood $N(x)$. This allows considering the previously simulated values when determining the value of an uninformed pixel. This link between simulated values increases their spatial coherence and improves the simulation output. Here, we use a random path P , which is common in geostatistics.

Pseudocode of Algorithm 2 (Direct Sampling)

```

1 DS_Simulation (training image, target image, threshold t, exploration ratio f, path P)
2 For each pixel  $B(v)$  of the target image do following sequentially the path  $P$ 
3 Extract the neighborhood  $N(v)$  of the target image around  $B(v)$  taking into account previously simulated values
4 Find a candidate  $B(u)$  in the training image that satisfies  $\varepsilon(u, v) < t$  by scanning a fraction  $f$  of the training image
5 Import the color from the candidate  $B(u)$  at location  $B(v)$  in the target image.
6 End
7 Return target image

```

Note that the threshold t is the same as in Algorithm 1.

2.4. Simulation by narrow distribution selection (NDS)

We develop here the Narrow Distribution Selection (NDS) algorithm, which is partly related to an idea proposed by Liu and Journal (2004) for the simulation of categorical variables using multi-grids. While many geostatistical methods work by computing a conditional probability density function (CPDF) for a simulated value $B(v)$, Algorithms 1 and 2 do not require this CPDF. Instead, these algorithms sample only a single value for each colorized pixel. The absence of a CPDF is computationally advantageous; however, a CPDF can offer additional possibilities, which we exploit in the NDS approach. In NDS, the selection of a pattern is not only conditioned to a match with its neighborhood: an additional criterion is included, which is that the CPDF of the value for a target pixel must be narrow. Indeed, using only a best pattern match criterion (as in Algorithms 1 and 2) can yield a set of candidates with different pixel values, all with a low mismatch. Such poorly determined situations result in pixels being assigned with an irrelevant value, creating artifacts that are thereafter propagated throughout the sequential simulation. The addition of a narrowness criterion results in only simulating the pixels whose values can be determined unambiguously. Therefore, like DS, this algorithm is

sequential, however the simulation path P is not random: it follows an order defined by the narrowness criterion.

Pseudocode of Algorithm 3 NDS

```

1 NDS_Simulation(training image, target image, best pattern match criterion, narrowness criterion)
2 Until all positions are simulated do
3 For all non-simulated pixels  $B(v)$  of the target image do
4 Extract the neighborhood  $N(v)$  of the target image around  $B(v)$  taking into account previously simulated values
5 Find all candidates in the training image satisfying the best pattern match criterion
6 Compute the CPDF of  $B(v)|N(v)$ 
7 If the CPDF satisfies the narrowness criterion do
8 Import the color from a random candidate  $B(u)$  sampled from the CPDF, at location  $B(v)$  in the target image.
9 Endif
10 End
11 End

```

Note that if many bands are simulated at the same time, there is one narrowness per band. In that case, all narrowness are aggregated into a single value by using a weighted average, the weights corresponding to the central value of the kernel for each band.

This approach avoids the propagation of prediction errors caused by the simulation of uncertain values. Instead, it only propagates the information contained in values with a narrow distribution (hence highly certain). A benefit of the narrowness criterion is that it is based on multiple possible candidates, which offers more information than Algorithms 1 and 2 that only retain a single candidate.

Algorithm 3 is computationally expensive. We propose an efficient implementation of it, described in Algorithm 4. It produces similar results while limiting redundant computations. Algorithm 4 is also geared towards implementation on modern hardware such as MIC (Many Integrated Core) or GPUs (graphics processing units) by separating the parts of the code that can be run in parallel. To approximate the CPDF and to determine a narrowness criterion, we use pattern matching to obtain a set of candidate pixels from the training image. For each simulated pixel, once a sufficient number of candidates is available, the CPDF of the target value can be determined empirically. In practice, we define the set of candidates by taking the k candidates that have the smallest matching error. Then, we formulate the narrowness criterion for the CPDF as a constraint on the interquartile range (IQR).

Pseudocode of Algorithm 4: an efficient implementation of Algorithm 3

```

1 Parallel_NDS_Simulation(training image, target image, number of candidates to estimate the narrowness k)
//Initialization loop
2 For each pixel  $B(v)$  of the target image do independently
3 Extract the neighborhood  $N(v)$  of the target image around  $B(v)$ 
4 Find the set of  $k$  candidates in the training image with lowest  $\varepsilon(u, v)$ 
5 Compute and store the narrowness based on the candidates in  $E(v)$ ; select and store one candidate  $B(u)$  randomly chosen among the set in  $C(v)$ .
6 End
//Simulation loop
6 Until all positions are simulated do
7 Find the position  $v$  with the lowest narrowness  $E(v)$  among all non-simulated pixel in  $E$ .
8 Import the color from  $B(C(v))$ , at location  $B(v)$  in the target image.
9 For all non-simulated pixels  $B(v)$  of the target image that are directly adjacent to the simulated pixel (including diagonals) do independently
10 Extract the neighborhood  $N(v)$  of the target image around  $B(v)$  taking into account previously simulated values
11 Find the set of  $k$  candidates in the training image with lowest  $\varepsilon(u, v)$ 
12 Compute and store the narrowness based on the candidates in  $E(v)$ ; select and store one candidate  $B(u)$  randomly chosen among the set in  $C(v)$ .
13 End

```

The two main algorithmic parameters are the kernel and the number of candidates k .

Note that in this algorithm most computations are done in the

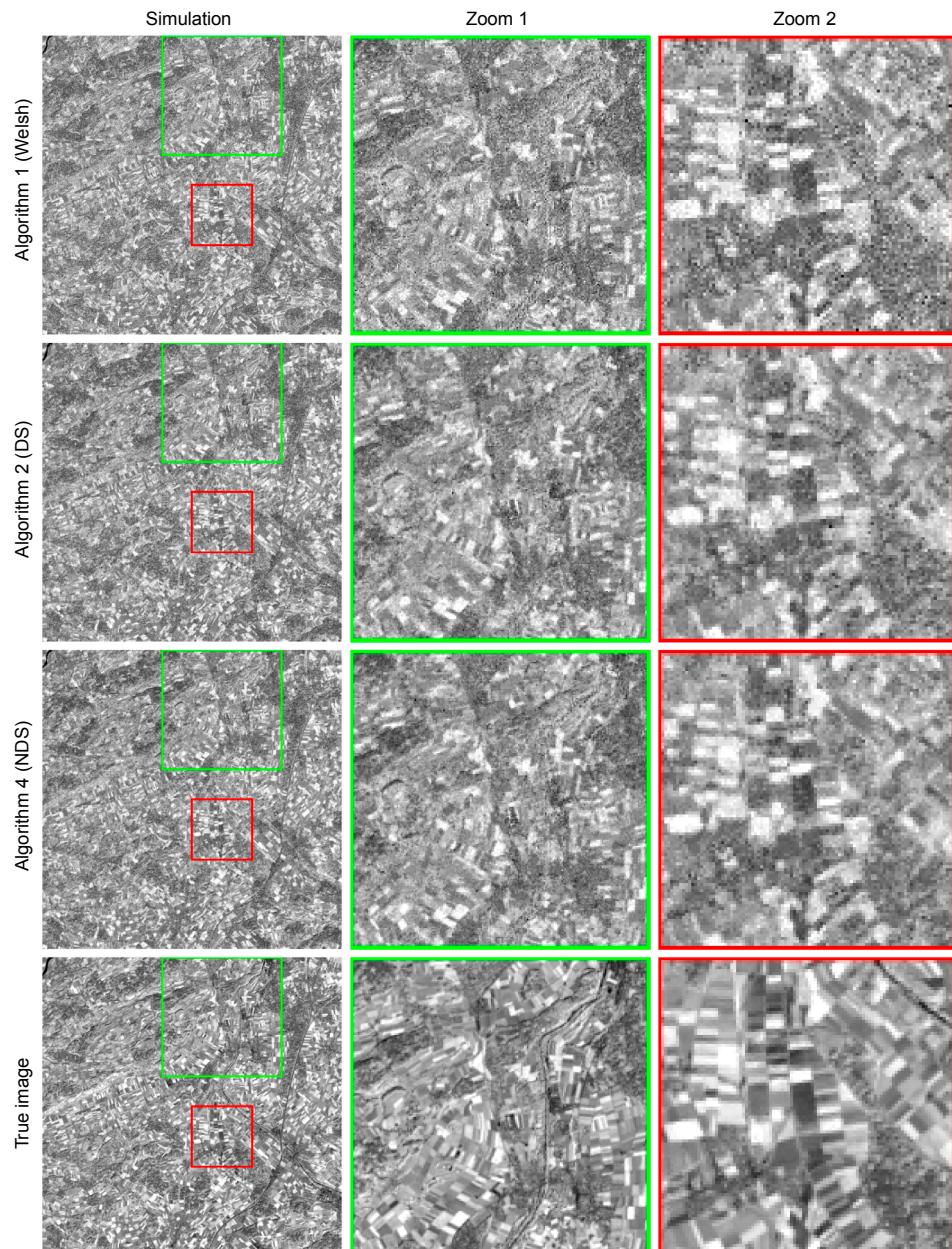


Fig. 6. Spectral extrapolation RGB → NIR for all three algorithms.

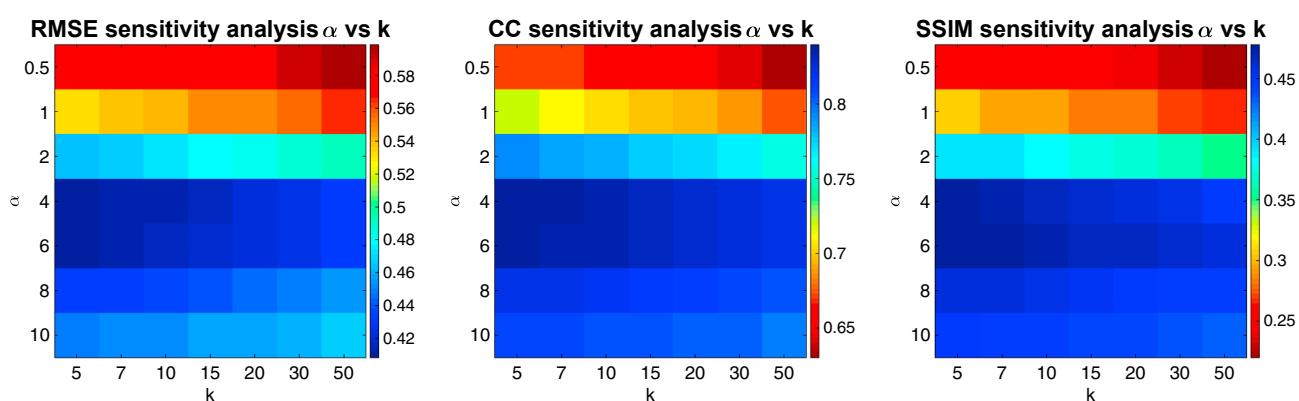


Fig. 7. Sensitivity of the NDS parameters using three different error metrics.

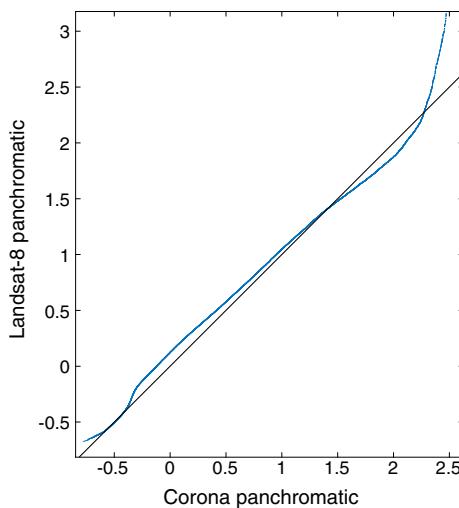


Fig. 8. q-q plot between corona and Landsat 8 panchromatic bands.

initialization part [Line 2–5 Algorithm 4] and updating part [Line 9–12 Algorithm 4], where operations on each pixel are independent and can be parallelized.

3. Numerical tests

In this section, we test the colorization algorithms presented in Section 2 in different scenarios. In the first example, the target multispectral image is known, so that it provides a reference against which the results can be validated. This case is used to study both the spectral disaggregation ($\text{PAN} \rightarrow \text{RGB}$) and spectral extrapolation ($\text{PAN} \rightarrow \text{NIR}$) problems. In the second case study, we test the colorization ($\text{PAN} \rightarrow \text{RGB}$) of some of the oldest satellite acquisitions, the Corona images, for which no reference color is available. For all the following tests, we generate 10 independent realizations for each technique, using an exponential kernel with $\alpha = 4.5$. For NDS $k = 10$, and for DS, $f = 0.3$ and $t = 0$.

The validation of the results is performed based on well-established metrics: the root mean square error (RMSE), correlation coefficient (CC) and Structural Similarity (SSIM) (Wang et al., 2004).

3.1. Landsat-8 test case

To test and compare the methods presented in Section 2 in the context of colorization, we apply them for the colorization of a raw OLI (Operational Land Imager) Landsat-8 scene of the Swiss plateau, acquired on August 30th, 2015. The scene is separated in two subareas of 500×500 pixels, which are shown in Fig. 3: the first one is used as training image and the second one as target, where we use the approaches presented to restore missing spectral bands. Both images present similar features, including mostly small urban areas, forests, and crops.

Usually, satellite sensor images are acquired under varieties of conditions, such as Sun position. Many of these variations can be modelled with a linear transformation. For practical reasons, each band is standardized (mean = 0 and standard deviation = 1) separately, which results in a homogeneous representation between bands. If required, the back-transformation is trivial to apply.

3.2. Spectral disaggregation

The spectral disaggregation is applied with each of the three algorithms presented in Section 2. The training image is composed of the panchromatic (PAN), red (R), green (G) and blue (B) channels. The target image contains only the PAN channel, which corresponds to the

band 8 on Landsat 8. This band originally has a 15 m spatial resolution, but in this example, it was interpolated to 30 m using bicubic interpolation to match the spatial resolution of the color bands.

The output of the simulations, shown in Fig. 4, are visually satisfying for all three algorithms. However, when zooming on the colorized images it is possible to observe that algorithm 4 produces less artifacts or noise than algorithms 1 and 2. In detail, it is clear that NDS produces more structured textures. For example, the NDS results depict a more accurate reconstruction of the forest colors (the bottom right of the zoom in Fig. 4) and a better preservation of features' edges; the roads are more clearly visible, for example (Zoom Fig. 4). These visual interpretations are confirmed by the error metrics. This is confirmed by the quantitative comparison results given in Table 1.

3.3. Spectral extrapolation

For spectral extrapolation, we use the same training and target images. Two series of tests are shown. In the first series, the unknown NIR band is simulated based on the known PAN band, with all three algorithms ($\text{PAN} \rightarrow \text{NIR}$). In the second series, we simulate again the NIR, but this time, based on the three RGB color bands ($\text{RGB} \rightarrow \text{NIR}$). The results are reported in Table 2 and Fig. 5 for $\text{PAN} \rightarrow \text{NIR}$ (black in Table 2) and Fig. 6 for $\text{RGB} \rightarrow \text{NIR}$ (red in Table 2).

3.4. Sensitivity of the NDS parameters

In order to better understand the behavior of the proposed NDS algorithm, we perform a sensitivity analysis to determine the influence of each parameter on the simulation performance. The sensitivity analysis is based on the case of the $\text{RGB} \rightarrow \text{NIR}$ simulation. The most influential parameters in the NDS algorithm are k and α , hence the analysis is based on these only. A series of values for k and α were initially predefined, and we ran a single simulation for each parameter combination because the variability between realizations is negligible when compared to the magnitude of the errors as showed Table 2. Each set of parameters is then assessed by computing the RMSE, CC and SSIM metrics.

Fig. 7 shows that regardless of the metric, the sensitivity to parameter α is dominant compared to parameter k . While the error does not change significantly between $k = 5$ and $k = 20$, the error space seems convex along α . It is also clear that, for this case study, the calibration of α is more important than that of k . Both variables are globally independent, thus we advise for practical applications to perform in priority a sensitivity analysis over α .

3.5. Corona test case

As a real-life test of our approach for enriching old satellite data, we applied it to the colorization of declassified Corona spy imagery (USGS, 2008). Such images were originally acquired on an argentic film, which was sent back to Earth in capsules, and digitalized many years later. Only the panchromatic band is available, with a variable spatial resolution between missions. The images themselves present a number of artifacts due to their age, or to the recuperation process, or artifact (lines) that are introduced by the digitalization process. Our goal here is to reconstruct a full RGB and NIR image with comparable spectral and spatial resolutions as Landsat 8 (Fig. 9 and Fig. 10). We use a Corona image acquired on the 13th of July 1975 over the same type of Swiss plateau landscape as in the case study of Section 3.1. As a consequence, the same training image as in Section 3.1 is used (Fig. 3). We compared the performance of algorithms 1, 2 and 4 in this exercise. It is important to note that it is impossible to validate such simulations quantitatively because there is no reference data to compare with, however a qualitative evaluation is possible.

Due to the specific nature of Corona images (based on film photography, with problems related to over and under-exposition), some

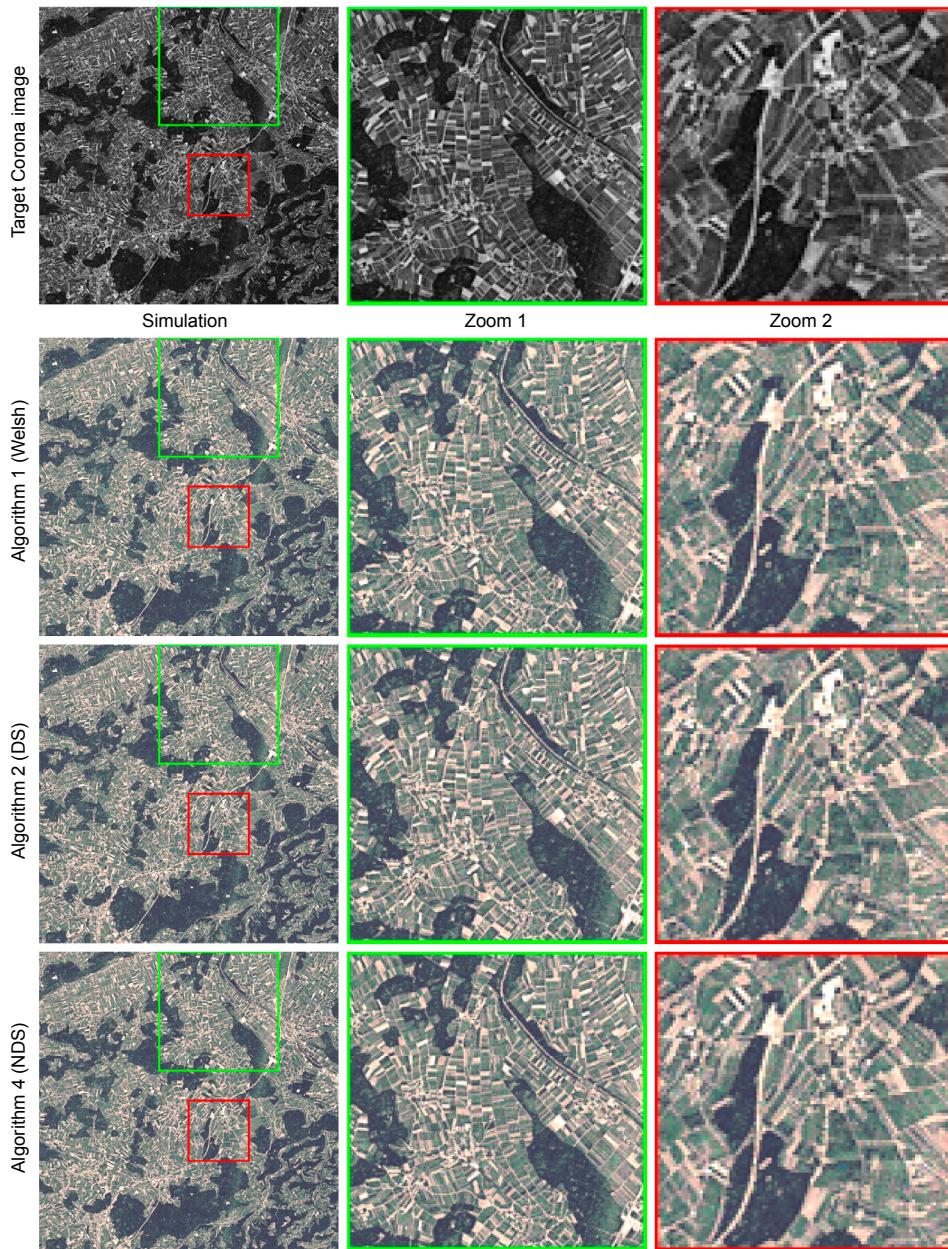


Fig. 9. Spectral disaggregation from PAN to RGB for all three algorithms on Corona images.

preprocessing steps had to be applied before the colorization process:

- The Corona image that has originally a higher spatial resolution was upscaled using a bicubic interpolation to a spatial resolution of 30 m in order to correspond to the resolution of the Landsat-8 image.
- The histogram of the values in the grayscale Corona image did not correspond to the histogram of the panchromatic Landsat-8 image, due to a saturation of the high and low values. Since all colorization approaches rely on corresponding training /target images pairs, we performed a histogram adjustment of the Landsat-8 panchromatic band based on a quantile-quantile mapping, as illustrated in Fig. 8. This allows having corresponding patterns in the Corona and in the Landsat-8 panchromatic band.

4. Discussion

All three algorithms show acceptable results for the case of disaggregation, which can be explained by considering that the color

information is present in a convoluted way: $PAN = \int RGB$. This results in a strong dependence between the known and unknown variables. While all methods produce color images that are reasonably similar, NDS results are systematically more accurate than DS, and Algorithm 1 shows systematically the lowest performance. However, the accuracy of the results is inversely proportional to the algorithm computational cost: NDS simulations are usually one order of magnitude slower (about 30 min on a bi-Xeon machine) than Algorithm 1 (about 3 min). The cost of algorithm 2 (DS) lies in-between, and varies depending on the parameterization. Accordingly, we recommend to always start by applying Algorithm 1 for disaggregation as this method often provides acceptable results at a relatively low cost. The NDS approach is recommended only if very high-quality spectral enhancement is required.

The recommendation is different for applications dealing with spectral extrapolation, where Algorithm 1 results in highly noisy images. This can be explained by the fact that spectral extrapolation is challenging, as can be seen in Fig. 11. It is clear that the dependence between the PAN and NIR bands is non-linear, whereas the dependence

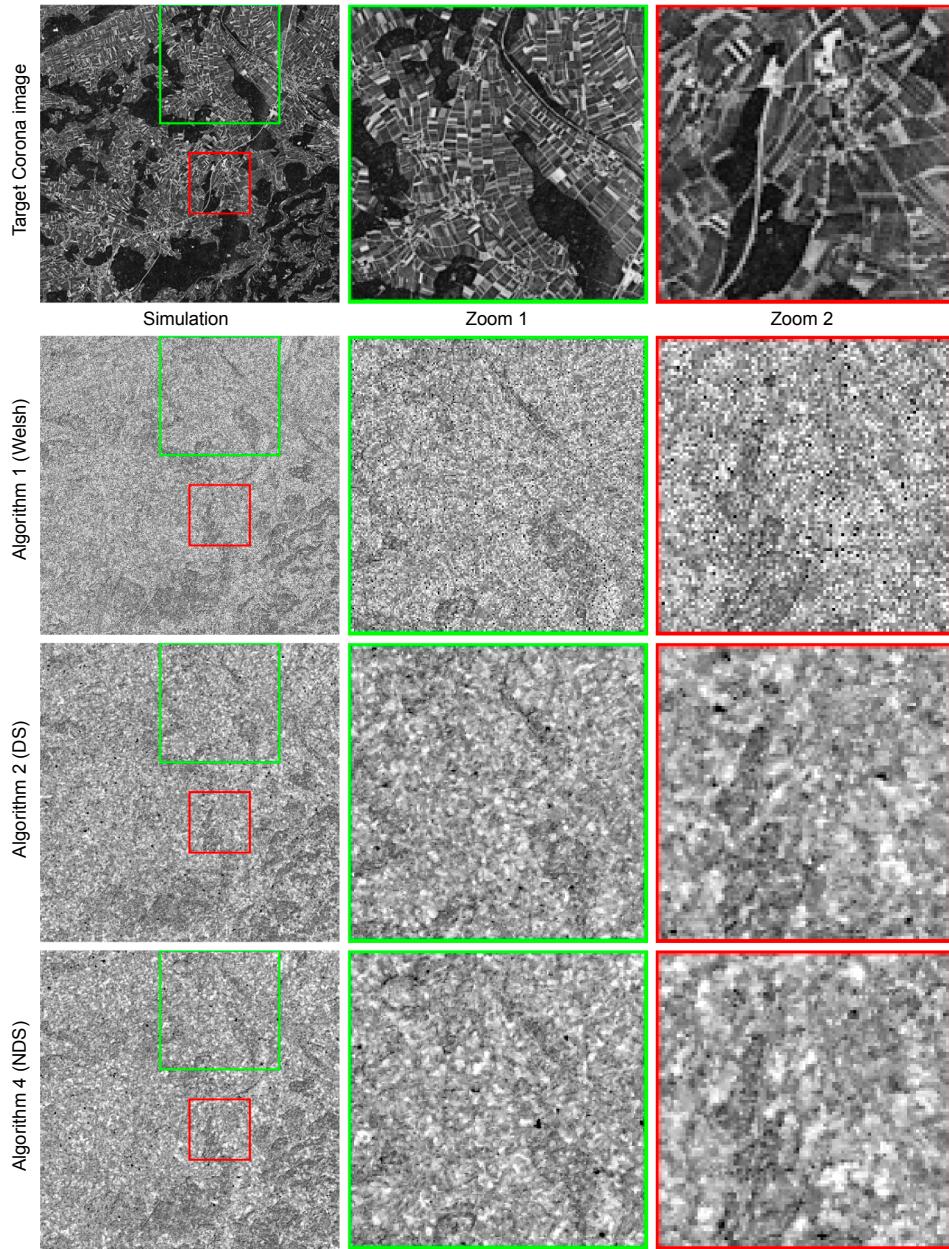


Fig. 10. Spectral extrapolation from PAN to NIR for all three algorithms on Corona images.

between the PAN and Green bands is close to linear.

For spectral extrapolation, NDS results are clearly more accurate than Algorithm 1. This is explained by the fact that NDS takes into account the dependence between neighborhoods, which allows constructing coherent textures, and thus reducing noise. This link between neighbors results in the successive simulation of nearby pixels (see video in supplementary material). Compared to DS, the narrowness constraint of NDS reduces the occurrence of incoherent neighborhoods, which results in increased colorization accuracy of features edges and textures. The result of the NIR colorization of Corona imagery clearly show that NDS can capture more complex structures, as visible in the zooms of Fig. 10 (especially in zoom 2).

The most important parameter, which is common to all three algorithms considered, is α which controls the shape of the kernel by setting, as seen at Fig. 7. Indeed, it defines the relative importance between the spectral constraint on the pixel and the spatial constraint. Therefore, α is a very important parameter that needs to be set carefully. Currently, cross validation on the training image appears to be the

best approach to this end. Our tests show that $\alpha=2$ is usually a good starting point, which attributes approximately half of the sum of weights to the central pixel. However, this value is not universal as for the example in Section 3.1, $\alpha = 4.5$ provided the best setup.

NDS has significantly more accurate results, however it requires two additional parameters. The first one is k , the number of candidates to consider for computing the narrowness. The minimum value for k is around 5 to allow a proper measurement of the narrowness of the CPDF. If the training image is large and presents enough repetition of patterns, k can be increased to 30 to allow for a more reliable narrowness measurement. The second parameter refers to the statistical dispersion metric used to compute the narrowness. In this paper, we use the interquartile range, which is a robust non-parametric dispersion measure, and not very sensitive to extreme values.

A common challenge of the proposed methods is to find a compatible training image. Our tests show that a good starting point is to use a present-day acquisition of the same or a neighboring area. In the particular case of important changes in the landscape configuration (such

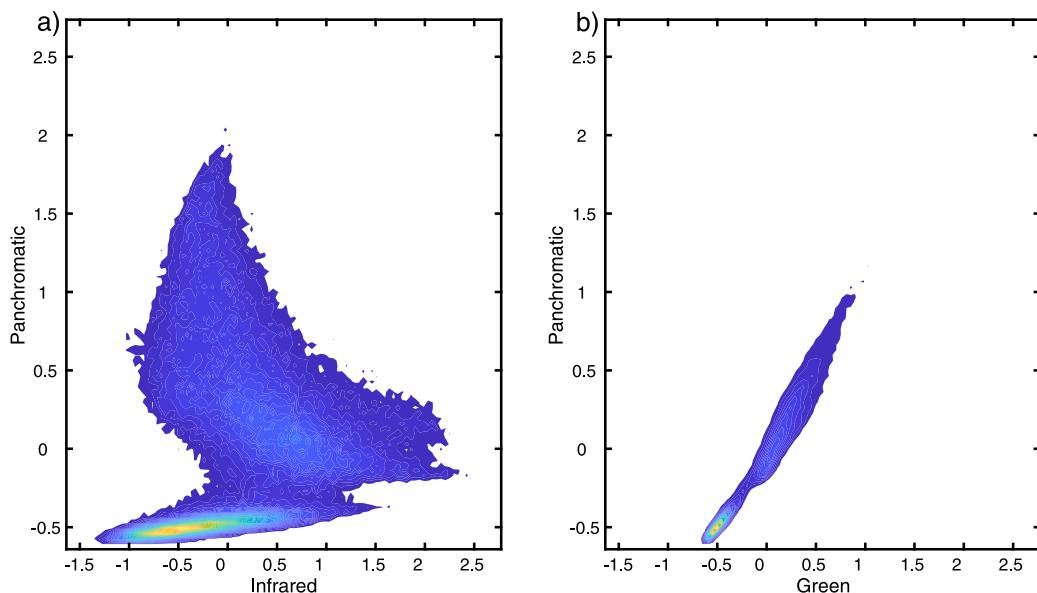


Fig. 11. Joint probability distributions of Landsat 8 data. (a) Green and Panchromatic bands, (b) NIR and Panchromatic bands.

as for example urban growth, deforestation or disasters such as forest fires or tsunamis), expert knowledge is required to find a suitable analogue area to use as training image.

The generality of the proposed framework allows envisioning new possibilities related to the spectral readjustment of sensor, like to convert the blue-green and green-red bands of Landsat 1 to the blue, green and red bands of Landsat 8. More generally it can be extended to new data source such as archive aerial photography, or for the disaggregation of near infrared on filterless standard cameras, which could be used to measure NDVI at low-cost.

5. Conclusion

Remotely sensed data are available for more than 45 years; however, only data from the last 30 years is being widely used. Analogue-based colorization algorithms provide an opportunity to use data from very early sensors, allowing for the integration of archive data that were previously unused, such that they can be exploited and compared to recent data sources. In this paper, propose a novel and general method with high computational cost (Algorithm 4), and the use of another much faster approach (Algorithm 1), but only reliable in the case of spectral disaggregation.

These methods can potentially be generalized to produce data with much higher spectral resolution, if compatible training images are available. For example, the methods presented in this paper could be applied to data acquired from hyperspectral airborne or satellite-based imagery such as Hyperion EO-1 data. As such, training-based colorization could offer an interesting alternative to spectral band reconstruction approaches based on regression, which do not account for spatial patterns (Hoang and Koike, 2018, 2017). In addition, the proposed algorithms can be useful to improve image classification techniques, by synthesizing the missing spectral bands of the target image. Finally, of the methodology can be potentially extended for downscaling of both spatial and temporal information simultaneously.

6. Code availability

Source code and documentation of the NDS simulation algorithm is available at: <https://github.com/GAIA-UNIL/G2S>

Platform: Linux / macOS / Windows 10

Language: C/C++

Interfacing functions in MATLAB and Python3

Acknowledgments

We are thankful to Thomas Hansen, Peter Atkinson, Julien Straubhaar and Raphaël Nussbaumer for their insightful comments that helped improve this paper. We also thank two anonymous reviewers for their constructive comments. This research was funded by the Swiss National Science Foundation, grant number 200021_162882.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.isprsjprs.2018.11.003>.

References

- Alphan, H., Yilmaz, K.T., 2005. Monitoring environmental changes in the Mediterranean coastal landscape: the case of Cukurova, Turkey. Environ. Manage. 35, 607–619. <https://doi.org/10.1007/s00267-004-0222-7>.
- Beck, R., 2010. EO-1 User Guide v. 2.3, 2003. <https://doi.org/10.1016/j.rse.2004.11.012>.
- Boucher, A., 2009. Sub-pixel mapping of coarse satellite remote sensing images with stochastic simulations from training images. Math. Geosci. 41, 265–290. <https://doi.org/10.1007/s11004-009-9215-7>.
- Bruzzone, L., Demir, B., 2014. A review of modern approaches to classification of remote sensing data. In: Manakos, I., Braun, M. (Eds.), Land Use and Land Cover Mapping in Europe, Remote Sensing and Digital Image Processing. Springer, Netherlands, Dordrecht, pp. 127–143. https://doi.org/10.1007/978-94-007-7969-3_9.
- Charpiat, G., Bezrukova, I., Hofmann, M., 2009. Machine Learning Methods for Automatic Image Colorization 1–20. <https://doi.org/10.1201/b10284-16>.
- Cheng, G., Han, J., 2016. A survey on object detection in optical remote sensing images. ISPRS J. Photogramm. Remote Sens. 117, 11–28. <https://doi.org/10.1016/j.isprsjprs.2016.03.014>.
- Cohen, W.B., Spies, T.A., Alig, R.J., Oetter, D.R., Maiersperger, T.K., Fiorella, M., 2002. Characterizing 23 Years (1972–95) of stand replacement disturbance in western Oregon forests with Landsat imagery. Ecosystems 5, 122–137. <https://doi.org/10.2307/3659042?ref=search-gateway:e1903d9c988897adcb86d25852b606a6>.
- Dewidar, K.M., 2002. Landfill detection in Hurghada, North Red Sea, Egypt, using Thematic Mapper images. Int. J. Remote Sens. 23, 939–948. <https://doi.org/10.1080/01431160110070645>.
- Fraser, R.H., Olthof, I., Carrière, M., 2011. Detecting long-term changes to vegetation in northern Canada using the Landsat satellite image archive. Environ. Res. Lett. 6, 045502. <https://doi.org/10.1088/1748-9326/6/4/045502>.
- Ge, Y., 2013. Sub-pixel land-cover mapping with improved fraction images upon multiple-point simulation. Int. J. Appl. Earth Observ. Geoinf. 22, 115–126. <https://doi.org/10.1016/j.jag.2012.04.013>.
- Ge, Y., Bai, H., 2011. Multiple-point simulation-based method for extraction of objects with spatial structure from remotely sensed imagery. Int. J. Remote Sens. 32, 2311–2335. <https://doi.org/10.1080/014311601003698278>.
- Guardiano, F.B., Srivastava, R.M., 1993. Multivariate geostatistics: beyond bivariate moments. In: Geostatistics Tróia '92, Quantitative Geology and Geostatistics. Springer, Dordrecht, pp. 133–144. https://doi.org/10.1007/978-94-011-1739-5_12.

- Hansen, M.C., Stehman, S.V., Potapov, P.V., Loveland, T.R., Townshend, J.R.G., DeFries, R.S., Pittman, K.W., Arunarwati, B., Stolle, F., Steininger, M.K., Carroll, M., Dimiceli, C., 2008. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. Proc. Natl. Acad. Sci. USA 105, 9439–9444. <https://doi.org/10.1073/pnas.0804042105>.
- Hoang, N.T., Koike, K., 2018. Comparison of hyperspectral transformation accuracies of multispectral Landsat TM, ETM, OLI and EO-1 ALI images for detecting minerals in a geothermal prospect area. ISPRS J. Photogramm. Remote Sens. 137, 15–28. <https://doi.org/10.1016/j.isprsjprs.2018.01.007>.
- Hoang, N.T., Koike, K., 2017. Transformation of Landsat imagery into pseudo-hyperspectral imagery by a multiple regression-based model with application to metal deposit-related minerals mapping. ISPRS J. Photogramm. Remote Sens. 133, 157–173. <https://doi.org/10.1016/j.isprsjprs.2017.09.016>.
- Hwang, J., Zhou, Y., 2016. Image Colorization with Deep Convolutional Neural Networks.
- Jha, S.K., Mariethoz, G., Evans, J., McCabe, M.F., Sharma, A., 2015. A space and time scale-dependent nonlinear geostatistical approach for downscaling daily precipitation and temperature. Water Resour. Res. 51, 6244–6261. <https://doi.org/10.1002/2014WR016729>.
- Kessler, T.C., Comunian, A., Oriani, F., Renard, P., Nilsson, B., Klint, K.E., Bjerg, P.L., 2012. Modeling fine-scale geological heterogeneity-examples of sand lenses in tills. Groundwater 51, 692–705. <https://doi.org/10.1111/j.1745-6584.2012.01015.x>.
- Levin, A., Lischinski, D., Weiss, Y., 2004. Colorization using optimization. ACM Trans. Graph. 23, 689–694. <https://doi.org/10.1145/1015706.1015780>.
- Liu, Y., Journel, A., 2004. Improving sequential simulation with a structured path guided by information content. Math. Geol. 36, 945–964. <https://doi.org/10.1023/B:MATG.0000048800.72104.de>.
- Lochbühler, T., Pirot, G., Straubhaar, J., Linde, N., 2013. Conditioning of multiple-point statistics facies simulations to tomographic images. Math. Geosci. 46, 625–645. <https://doi.org/10.1007/s11004-013-9484-z>.
- Lu, D., Mausel, P., Brondízio, E., Moran, E., 2004. Change detection techniques. Int. J. Remote Sens. 25, 2365–2401. <https://doi.org/10.1080/0143116031000139863>.
- Mariethoz, G., Renard, P., Straubhaar, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resour. Res. 46. <https://doi.org/10.1029/2008WR007621>.
- Munyati, C., 2010. Wetland change detection on the Kafue Flats, Zambia, by classification of a multitemporal remote sensing image dataset. Int. J. Remote Sens. 21, 1787–1806. <https://doi.org/10.1080/014311600209742>.
- Noda, H., Korekuni, J., Niimi, M., 2006. A colorization algorithm based on local MAP estimation. Pattern Recognition 39, 2212–2217. <https://doi.org/10.1016/j.patcog.2006.03.015>.
- Oriani, F., Ohana-Levi, N., Marra, F., Straubhaar, J., Mariethoz, G., Renard, P., Karnieli, A., Morin, E., 2017. Simulating small-scale rainfall fields conditioned by weather state and elevation: a data-driven approach based on rainfall radar images. Water Resour. Res. 15, 265. <https://doi.org/10.1002/2017WR020876>.
- Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-Resolution Mapping of Global Surface Water and its Long-term Changes. Nature Publishing Group, pp. 418–422. <https://doi.org/10.1038/nature20584>.
- Pirot, G., Straubhaar, J., Renard, P., 2014. Simulation of braided river elevation model time series with multiple-point statistics. Geomorphology 214, 148–156. <https://doi.org/10.1016/j.geomorph.2014.01.022>.
- Pratt, W.K., 2007. Digital Image Processing. Wiley-Interscience.
- Semary, N.A., Hadhoud, M.M., El Kilani, W.S., Ismail, N.A., 2007. Texture recognition based natural gray images coloring technique. In: Presented at the National Radio Science Conference, NRSC, Proceedings. IEEE, pp. 1–12. <https://doi.org/10.1109/NRSC.2007.371374>.
- Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. CRC Press.
- Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Math. Geol. 34, 1–21. <https://doi.org/10.1023/A:1014009426274>.
- Tang, Y., Atkinson, P.M., Zhang, J., 2015. Downscaling remotely sensed imagery using area-to-point cokriging and multiple-point geostatistical simulation. ISPRS J. Photogramm. Remote Sens. 101, 174–185. <https://doi.org/10.1016/j.isprsjprs.2014.12.016>.
- USGS, 2008. U.S. Geological Survey Fact Sheet 2008-3054, Declassified Intelligence Satellite Photographs, pp. 1–2.
- Vaduva, C., Costachioiu, T., Patrascu, C., Gavat, I., Lazarescu, V., Datcu, M., 2013. A latent analysis of earth surface dynamic evolution using change map time series. IEEE Trans. Geosci. Remote Sens. 51, 2105–2118. <https://doi.org/10.1109/TGRS.2012.2219316>.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612. <https://doi.org/10.1109/tip.2003.819861>.
- Welsh, T., Ashikhmin, M., Mueller, K., 2002. Transferring color to greyscale images. SIGGRAPH 277. <https://doi.org/10.1145/566570.566576>.
- Weng, Q., 2001. A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China. Int. J. Remote Sens. 22, 1999–2014. <https://doi.org/10.1080/01431160118847>.
- Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. Nat. Climate Change 3, 875–883. <https://doi.org/10.1038/nclimate1908>.
- Yin, G., Mariethoz, G., McCabe, M., 2017. Gap-filling of landsat 7 imagery using the direct sampling method. Remote Sens. 9, 12. <https://doi.org/10.3390/rs9010012>.
- Zanter, K., 2005. Landsat 8 (L8) data users handbook. LSDS-1574 Version.