

## FUSION OF HYPERSPECTRAL AND PANCHROMATIC DATA BY SPECTRAL UNMIXING IN THE REFLECTIVE DOMAIN

Y. Constans<sup>1,2</sup>, S. Fabre<sup>1</sup>, M. Seymour<sup>3</sup>, V. Crombez<sup>3</sup>, X. Briottet<sup>1</sup>, Y. Deville<sup>2</sup>

<sup>1</sup> DOTA, ONERA, 31055 Toulouse, France - (yohann.constans, sophie.fabre, xavier.briottet)@onera.fr

<sup>2</sup> IRAP, UPS-CNRS-OMP-CNES, Université de Toulouse, 31400 Toulouse, France - (yconstans, ydeville)@irap.omp.eu

<sup>3</sup> AIRBUS Defence and Space, 31400 Toulouse, France - (michael.seymour, vincent.crombez)@airbus.com

**KEY WORDS:** Image fusion, panchromatic, hyperspectral, pansharpening, spectral unmixing, urban, reflective domain

### ABSTRACT:

Earth observation at the local scale implies working on images with both high spatial and spectral resolutions. As the latter cannot be simultaneously provided by current sensors, hyperspectral pansharpening methods combine images jointly acquired by two different sensors, a panchromatic one providing high spatial resolution, and a hyperspectral one providing high spectral resolution, to generate an image with both high spatial and spectral resolutions. The main limitation in the fusion process is in presence of mixed pixels, which particularly affect urban scenes, and where large fusion errors may occur. Recently, the Spatially Organized Spectral Unmixing (SOSU) method was developed to overcome this limitation, delivering good results on agricultural and peri-urban landscapes, which contain a limited number of mixed pixels. This article presents a new version of SOSU, adapted to urban landscapes. It is validated on a Toulouse (France) urban dataset at a 1.6 m spatial resolution acquired by the HySpex instrument from the 2012 UMBRA campaign. A performance assessment is established, following Wald's protocol and using complementary quality criteria. Visual and numerical (at the global and local scales) analyses of this performance are also proposed. Notably, in the VNIR domain, around 51 % of the mixed pixels are better processed by the presented version of SOSU than by the method used as a reference. This ratio is improved regarding shadowed areas in the reflective (52 %) and VNIR (57 %) domains.

### 1. INTRODUCTION

At the local scale, most remote sensing applications for Earth observation require both high spatial resolution, to accurately depict the geometry of the observed scene, and high spectral resolution, to extract information about its state (Sabins, 2007). However, sensor characteristics are often limited as they cannot simultaneously provide high spatial and spectral resolutions (Lier et al., 2012). A solution is to combine images jointly acquired by two different sensors. On the one hand, panchromatic (PAN) images provide high spatial resolution with one broad spectral band in the visible range [0.4  $\mu\text{m}$  – 0.8  $\mu\text{m}$ ]. On the other hand, hyperspectral (HS) images provide numerous spectral bands covering the reflective range [0.4  $\mu\text{m}$  – 2.5  $\mu\text{m}$ ], with a lower spatial resolution. The combination of HS and PAN images to generate a new HS image with high spatial and spectral resolutions represents a case of image fusion called hyperspectral pansharpening, or hypersharpening.

The various methods presented in the literature for spatio-spectral image fusion can be classified into several main classes depending on the processing strategy, each of them having its own advantages and drawbacks. Component substitution, multiresolution analysis and hybrid approaches are adapted to the fusion of multispectral (MS) and PAN images (Vivone et al., 2014), whereas Bayesian and matrix factorization methods are suited to the fusion of HS and MS images (Yokoya et al., 2017). The HS and PAN fusion (abbreviated as HS+PAN) is based on methods from these different classes. A comparative study performed in the reflective domain (Loncan et al., 2015) highlighted the main limitations of these HS+PAN methods:

- Better preservation of one type of information (spatial or spectral) at the expense of the other one;

- Limitation of the HS/PAN spatial resolution ratio (denoted as  $r$ );
- Spectral distortions in the HS range not included in the PAN domain, mostly visible beyond 1  $\mu\text{m}$ ;
- Exploitation of limited spectral ranges: visible domain for the PAN image, at best reflective domain for the HS image, and generally HS and PAN images covering the same spectral domain;
- Errors from scenes with high spatial variability, inducing HS pixels associated with several materials and called mixed pixels (Constans et al., 2020), which are not well processed by most of the existing fusion methods. In urban areas, 40 to 50 % of the pixels are mixed at a 4 m spatial resolution (Wu, 2009);
- Errors from shadows, particularly present in urban areas (due to the 3D building shapes);
- Intra-class variability: a single material can have very different reflectance values from one location to another, depending on various parameters (age, orientation of the objects...).

The Spatially Organized Spectral Unmixing HS+PAN method (SOSU-2019) has been designed (Loncan, 2016) and recently improved (Constans et al., 2020) to minimize the limitations stated above (particularly the preservation of the spatial and spectral contents) in the reflective domain for complex areas (mixed pixels). It is based on an existing fusion process for spatial information preservation, which is called Gain. The latter is inspired from the Brovey transform in the RGB+PAN case (Vivone et al., 2014) and is considered as the reference method. SOSU-2019 supplements Gain with a preprocessing based on spectral unmixing and spatial reorganisation (steps identified in green in Fig. 1), to detect mixed pixels and better handle them. This preprocessing locally extracts pure spectra consti-

tuting the scene (also called endmembers) from the HS image, and judiciously assigns them to the different pixels of the resulting image, to follow the spatial organisation of the scene at the finer PAN spatial resolution. SOSU-2019 has been validated on agricultural and peri-urban scenes, providing better results than Gain (Constans et al., 2020). However, it still had to be extended to real urban scenes, meaning adapting the method to process a larger proportion of mixed pixels while handling significant shadowed areas.

Thus, the aim of this article is to present the extended method, with its evolutions and latest enhancements for urban environments, which is denominated as SOSU. Its upgrades include in particular a new combined strategy for the segmentation (Section 2.1.1), mixed pixel detection (Section 2.1.3) and spatial reorganisation (Section 2.1.5) steps. In addition, a complete performance assessment of SOSU (in comparison with Gain) is established, which is composed of visual and numerical analyses. The evaluation method, following Wald's protocol (Section 2.2), involves relevant and complementary quality criteria (Section 2.2.2), including spatial, spectral and global measures. These criteria are applied at the global image scale, as well as the local pixel scale to generate error maps depicting the spatial variations of the error.

## 2. METHODOLOGY

### 2.1 Presentation of SOSU

The proposed SOSU method includes six main steps (Figure 1), each of them summarized below. The improvements are identified and further detailed in this article. The Gain fusion process (spatial information adding step), on which SOSU is based, is described in detail in (Loncan, 2016) and (Constans et al., 2020).

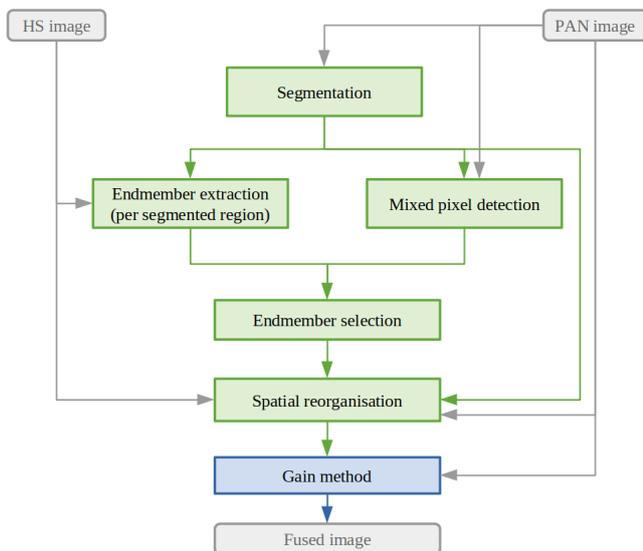


Figure 1. Flow chart of the SOSU method (the preprocessing steps are identified in green, the fusion step in blue)

The presented method assumes the following hypotheses:

- The HS and PAN images are fully registered. In addition, the HS/PAN spatial resolution ratio, referred to as  $r$ , is an integer;

- The HS and PAN images respectively cover the reflective and visible spectral domains;
- All images are expressed in spectral radiance.

In the sequel, we use the following terms:

- Subpixels: pixels at the finer PAN spatial resolution. A single HS pixel covers the same area as  $r \times r$  PAN subpixels; it is *associated* with this group of subpixels.
- Pure pixel: an HS pixel whose spectral signature is the spectrum of a single material. Hypothesis: a HS pixel is pure if the corresponding PAN subpixels have radiance values characterized by low variability.
- Mixed pixel: an HS pixel which is not pure. Its spectrum is a supposedly linear combination of several pure material spectra, and we assume that the associated PAN subpixels have heterogeneous radiance values.
- Endmember spectrum (abbreviated as endmember): also called pure spectrum, it is the spectral signature of a single material. In particular, the spectrum of a pure pixel is an endmember. Yet, endmembers can also be extracted from a given group of spectra by unmixing methods.

**2.1.1 Segmentation of the PAN image (upgraded in this paper):** This step aims to split the PAN image (which contains the spatial information) into several regions (also called segments) of homogeneous radiance values. The ideal segmentation method should couple one single region of PAN pixels with one single material (from ground truth). This ensures the relevance of the endmembers which will be extracted from these regions (Section 2.1.2), and then the spatial accuracy of the reorganised image (Section 2.1.5).

The quality of the segmentation is thus crucial, yet over-segmentation is not a problem and is preferred to under-segmentation. Indeed, even if the former case can cause issues in terms of implementation (unnecessary additional regions and associated extracted endmembers increase the number of combinations to be tested, which can lead to implementation time issues), in terms of fusion accuracy, the quality of the final result will be strictly enhanced (additional combinations to test and finer regions).

Hence, we retain clustering-based methods, to preserve as many spatial details as possible without caring about over-segmentation. The retained methods are: k-means clustering (Kettaf et al., 1996), meanshift clustering (Comaniciu, Meer, 2002) and the meanshift-based segmentation method, EDISON (Christoudias et al., 2002). The two raw clustering methods (k-means and meanshift) are specific to this presented version of SOSU. For these two methods, the input data consist of a list of points with three coordinates: the radiance value plus the two spatial coordinates of each PAN pixel. This way, we favor spatially compact clusters. Then, we convert the clusters into regions by separating clusters composed of non-adjacent groups of pixels into distinct regions.

**2.1.2 Endmember extraction per region:** To extract all the endmembers corresponding to each segment, we refer to the spectra of all the HS pixels which cover (even partially) this segment at the PAN spatial resolution. Several methods have been tested (Constans et al., 2020), but the Vertex Component Analysis (VCA) method (Nascimento, Dias, 2005) has been selected as the most accurate one for this purpose. The number of endmembers to be returned can be estimated for each segment

by the Hyperspectral Signal Identification by Minimum Error (HySime) method (Bioucas-Dias, Nascimento, 2008), but it can also be a fixed parameter.

**2.1.3 Mixed pixel detection (upgraded):** This step aims to discriminate mixed pixels from pure pixels. Indeed, SOSU preprocessing only applies to mixed pixels.

To this end, we refer, for a given HS pixel, to the group of  $r \times r$  PAN subpixels covering the same area, and we evaluate its homogeneity. Unlike the previous versions of SOSU, we choose to refer to the segmentation map (Section 2.1.1), and check if the corresponding area of analysis is associated with one or several segments. Thus, if we count more than one segment, we consider that the corresponding HS pixel is mixed. This simple method tends to underestimate the number of pure pixels (since over-segmentation is favored, as detailed in Section 2.1.1), but the latter are more reliable, which avoids the opposite problem, i.e. mixed pixels detected as pure (and therefore not processed). This is the reason why this method has been chosen.

**2.1.4 Endmember selection (upgraded):** For each mixed HS pixel, we assemble a list of pure spectra to be assigned to its corresponding subpixels in the resulting image. This list gathers:

- the endmembers extracted from all segments included (at least partially) in this pixel at the PAN spatial resolution;
- the spectra of pure pixels present in a given neighbourhood.

In both cases, the selection is local (neighbourhood or adjacent regions), to avoid the intra-class variability of a single material present at different locations of the image.

As this gathered list might be redundant (i.e. it might contain several spectra per class), we can reduce it by a correlation test (Constans et al., 2020), but this might also decrease the quality of the result (fewer endmembers to test), which is why we do not use this option in the sequel.

**2.1.5 Spatial reorganisation (upgraded):** This last preprocessing step aims to attribute, for each mixed pixel, the right endmembers to the right subpixels. It has to find the best arrangement according to the spatial variability of the PAN image, as depicted in Figure 2 (Loncan, 2016).

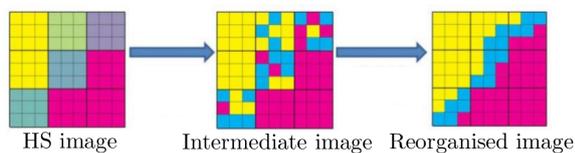


Figure 2. Description of the spatial reorganisation step

The proposed method (Constans et al., 2020) relies on a combinatory analysis, which consists in testing all the possible combinations of pairs constituted of one region (included in the processed HS pixel) and one possible endmember (from the associated list, established in Section 2.1.4). Note that, when a tested endmember is associated with a region, it is attributed to all subpixels covering that region, so that we can generate the reorganised pixels associated with each tested combination. The chosen reorganised pixel is the one which minimizes the reconstruction error, which is defined in this paper as  $\sqrt{NRMSE_{PAN}^2 + NRMSE_{HS}^2}$ , with:

- $NRMSE_{PAN}$ : the normalized RMSE (see Section 2.2.2) between the spectra of the tested pixel integrated over the PAN domain and the corresponding values in the PAN image;
- $NRMSE_{HS}$ : the normalized RMSE between the averaged spectrum of the tested pixel and the corresponding spectrum in the HS image.

To limit the number of tested combinations, which can lead to a significant need of memory or computation time (depending on the implementation choice), two approaches are proposed in this paper:

- Full combinatory analysis: we test all the region/endmember combinations, as explained above (i.e.  $n_{endmembers} \times n_{regions}$  combinations per mixed pixel);
- Alternative method: we still retain one endmember per region, but these regions are processed one by one (i.e.  $n_{endmembers} \times n_{regions}$  combinations only per mixed pixel). For each region, we test all the proposed endmembers, then we keep the endmember minimizing the  $NRMSE_{PAN}$ , as defined above except that it is limited to the subpixels belonging to the processed region only.

For each mixed pixel, the selection of the reorganisation method is done by comparing the planned number of combinations (for the full combinatory analysis) with an empirical threshold. If the former is higher than the fixed threshold, the full combinatory analysis is not performed, and the non-exhaustive alternative approach is used instead.

**2.1.6 Gain method:** Once SOSU preprocessing has been done, the last step is to apply the Gain fusion process, to inject the spatial information from the PAN image into the reorganised one. The method is detailed in (Constans et al., 2020). The advantage of this method is that, on the one hand, the spectral content of the preprocessed HS image is fully preserved (only scale factors are applied to spectra), and, on the other hand, the spatial content from the PAN image is fully added into the preprocessed HS image (integrating the fused image over the PAN spectral domain results in the original PAN image).

## 2.2 Performance assessment protocol

**2.2.1 Wald's protocol:** To evaluate the SOSU performance, we use simulated images obtained by degrading (spatially to get the input HS image, and spectrally to get the input PAN image) a real reference HS image. Measuring the gap between the reference image and the fused image (which should have the same dimensions), by using adapted quality criteria (Section 2.2.2), is a relevant evaluation of the fusion process known as Wald's protocol (Wald et al., 1997). The closer the fused and reference images, the more relevant the fusion process. Wald's protocol is used here to set a systematic comparison between the proposed method without preprocessing (Gain) and with it (SOSU).

**2.2.2 Quality criteria:** Image Fusion Quality Metrics (IFQMs) are quality criteria adapted to image fusion (Jagalingam, Hegde, 2015), measuring the proximity between two input (reference and fused) images. They can be spatial, spectral and global (Loncan et al., 2015). The selected IFQMs are the RMSE (global), ERGAS (global), SAM (spectral) and CC (spatial), as defined below and further detailed in previous work (Constans et al., 2020). These IFQMs are complementary, widely used in image fusion, and among the most reliable criteria (Pei et al., 2012).

- Root Mean Squared Error (RMSE): this error takes into account all spatial and spectral dimensions in the same way. The higher the RMSE, the higher the error, 0 being the ideal value. A normalised RMSE (NRMSE) can be obtained by dividing the result by the mean of the input reference data.
- Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS): this error is designed to be independent of the units of measurement, the number of spectral bands and the spatial resolution ratio (Wald, 2000). The higher the ERGAS, the higher the error, a 0 value meaning that the two compared images are equal.
- Spectral Angle Mapper (SAM): this measure represents the angular deviation, for a given pixel, between its estimated and reference spectra (Kruse et al., 1993). It compares spectral shapes, and is independent of scale factors. The SAM of the complete image is then computed by averaging the values spatially obtained for each pixel. The higher it is, the more the spectral signatures of the compared spectra differ, a 0 value being ideal.
- Cross Correlation (CC) – non-centered spatial version : it measures the geometric distortion between two single-band images (Yoo, Han, 2009). The CC of the complete image is then achieved by averaging the results obtained for each spectral band. The higher the CC, the higher the similarity between the two images. It ranges from 0 to 1, a value of 1 meaning a maximal correlation.

**2.2.3 Analysis method:** Comparing the performance of SOSU to that of Gain for complex environments, and particularly urban landscapes, requires performing different types of analyses (visual, numerical), to dispose of complementary evaluation processes, and thus establish a complete performance assessment of the compared methods. To this end, quality criteria (Section 2.2.2) need to be applied to specific spectral domains and at different spatial levels.

Spectrally, each quality criterion can be obtained by taking into account the reflective domain, or by focusing on specific spectral ranges. In the sequel, we notably refer to the Visible and Near-Infrared (VNIR) domain, which covers the  $[0.4 - 1.0 \mu m]$  spectral interval, and the Short-Wave Infrared (SWIR) domain, which covers the  $[1.0 - 2.5 \mu m]$  spectral interval.

Spatially, the quality criteria can be performed at different levels. On the one hand, at the global level, they can be applied to the entire image or a specific pixel group (for example: shadowed pixels, mixed pixels), to get a general numerical evaluation of the fusion methods. In particular, we mainly focus on mixed pixels (and associated subgroups) by leaving aside the pure pixels because, by definition, Gain and SOSU are strictly the same process in this latter case (see Section 2.1.3). On the other hand, at the local pixel level, the spectral criteria can generate error maps, providing the spatial distribution and variations of the error according to the chosen spectral metric. At both local and global levels, we pay particular attention to the SAM criterion, which is a crucial measure because it only focuses on spectral shapes (Section 2.2.2) and thus indicates if materials are correctly assigned with SOSU preprocessing.

### 3. DATASET

#### 3.1 Real image

The chosen urban dataset represents the city center of Toulouse (France), in spectral radiance. It has been acquired at

a 1.6 m spatial resolution in the reflective range by the HySpex instrument from the 2012 UMBRA (ONERA-IGN) airborne campaign (Adeline et al., 2013), in the context of the HYPXIM/HYPEX-2/BIODIVERSITY hyperspectral mission (Briottet et al., 2017). This complete dataset ( $1417 \times 1417$  pixels) contains 408 spectral bands covering the reflective domain ( $[0.4 - 2.5 \mu m]$ ).

#### 3.2 Reference image

The reference image extracted from this dataset covers a reduced scene ( $96 \times 96$  pixels), which represents the "Halle aux Grains" and neighbour buildings (Figure 3(a)). The spectral bands whose wavelengths correspond to an atmospheric transmission coefficient lower than 90 % have been removed, therefore the reference image contains 234 spectral bands covering the  $[0.5 - 2.4 \mu m]$  domain.

The scene involves complex and close structures, which is why the proportion of mixed pixels is estimated as being higher than 90 % with all the methods of mixed pixel detection. A significant proportion of pixels is also affected by shadows: by using a simple and efficient R-G-B-NIR (Red-Green-Blue-Near Infrared) literature index (Nagao et al., 1979), we detect 24 % of shadowed pixels in the image (Figure 3(b)). They are mainly present in the lower street, and in the right part of the scene.

In addition, many high-radiance artifacts are visible, which are due to reflective materials (cars, roof tiles, glass). Such pixels can deteriorate the associated larger pixels in the simulated HS image, identifiable with a green hue because of the viewing adjustment (high radiance thresholding) of the images (Figure 3(d)). These affected HS pixels can in turn deteriorate larger areas in the fused image, because their spectra may be detected as endmembers.

#### 3.3 Simulation of the HS and PAN degraded images

According to Wald's protocol (Section 2.2.1), the degradation to get the PAN image merely consists in spectrally averaging all the spectral bands of the reference image included in the visible domain (Figure 3(c)), while the degradation to get the HS image consists in spatially averaging all the  $r \times r$  subpixels associated with each HS pixel (Figure 3(d)). Here, an  $r$  spatial resolution ratio of 4 has been chosen. Hence, the spatial resolution of the generated HS image is 6.4 m, while the spatial resolution of the generated PAN image remains unchanged (i.e. 1.6 m).

## 4. RESULTS AND DISCUSSION

In this section, fusion results, obtained by SOSU and Gain with the Halle dataset, are presented, compared and discussed. To this end, we use the analysis method as well as the quality criteria presented in Section 2.2.

#### 4.1 Set of parameters

SOSU has been tested on the presented urban dataset with the following parameters:

- Segmentation method: Meanshift with over-segmentation allowed (quantile = 0.25 ; Number of samples = 30);
- Mixed pixel detection method: homogeneity of the segmentation map (see Section 2.1.3);

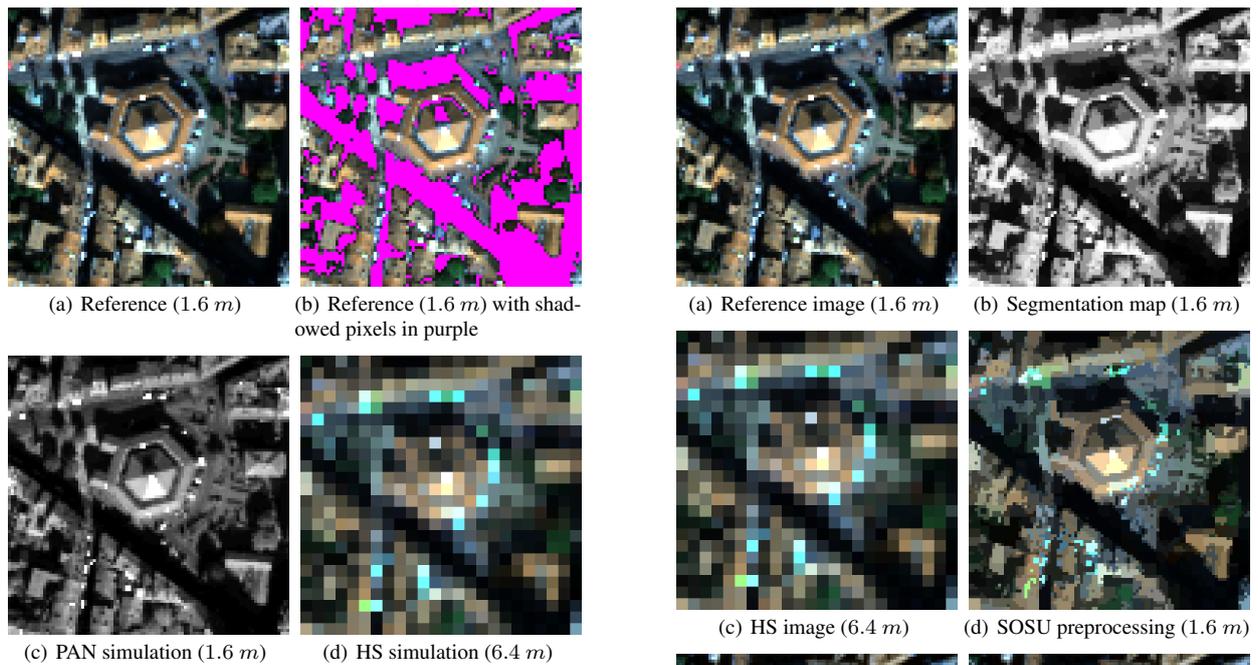


Figure 3. "Halle" dataset (spatial resolutions in brackets)

- Number of endmembers extracted per segment: 2 (due to the limited average size of each segment, no additional endmembers are needed);
- Neighbourhood of pure spectra selection: 2 (see Section 2.1.4);
- Reorganisation approach: all segment/endmember combinations are tested if the calculated number of combinations is lower than the combination threshold, otherwise the alternative method is performed;
- Combination threshold:  $10^6$  (see Section 2.1.5).

#### 4.2 Visual analysis

Fig. 4 visually compares the images fused by Gain and SOSU (preprocessing and full process).

Spatially, the different structures and buildings present in the reference image are overall well reconstructed by SOSU preprocessing, contrary to the input HS image. The delineations are accurate and broadly respect the spatial organisation of the original scene. This concerns in particular the central Halle building, the shadowed areas as well as the different streets of the scene.

For example, fig. 5 shows a pixel whose spectral signature is well reconstructed by SOSU due to a relevant spatial reorganisation, whereas Gain provides non-representative information of the reference spectral behaviour in the VNIR domain (information related to the HS spectra with a scale factor) and underestimates the radiance values in the SWIR domain. Hence, the average normalised gap in the reflective domain between the reference spectrum and the reconstructed spectrum is 10 % for SOSU (with only 2 % in the VNIR domain), as compared with 33 % for Gain (with 31 % in the VNIR domain). By also comparing the spectral shapes, the SAM applied to the reflective domain between the reference spectrum and the reconstructed spectrum is  $2^\circ$  for SOSU, as compared with  $19^\circ$  for Gain. This case thus reveals the interest of the added SOSU preprocessing, which aims at recovering the appropriate materials at the PAN resolution, before applying the Gain process.

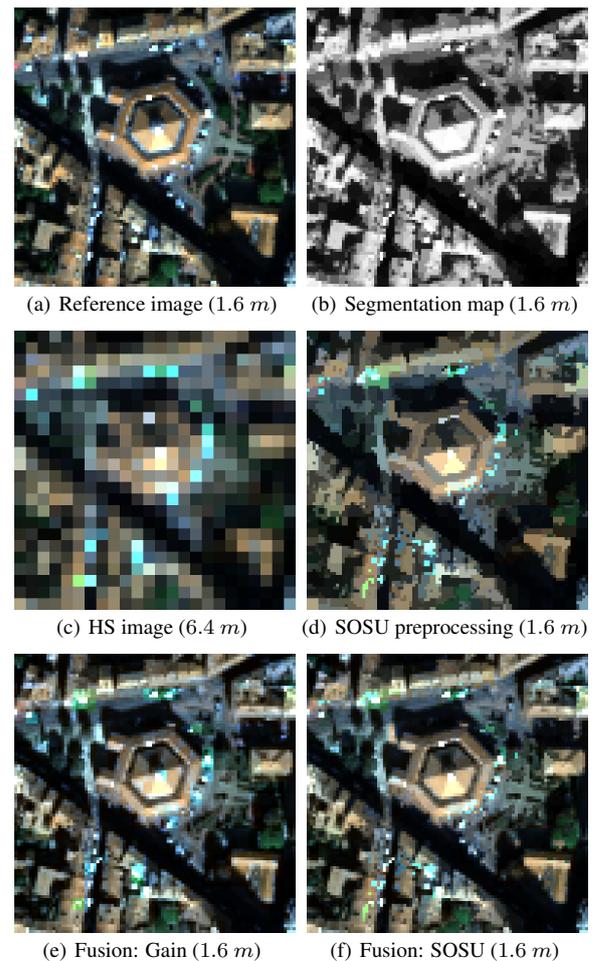


Figure 4. "Halle" dataset (spatial resolutions in brackets)

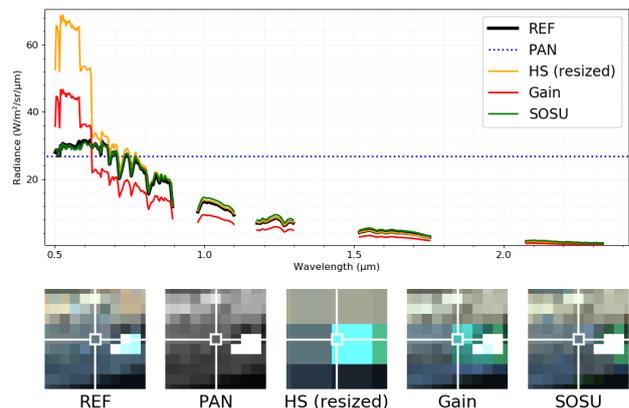


Figure 5. Subpixel better reconstructed with SOSU than Gain

However, local reorganisation errors remain (Fig. 4), which means some endmembers have been poorly assigned. This notably concerns small sized reflective objects (car parts, roof tiles, glass), and their associated subpixels. Let us consider such a subpixel: its spectral signature, due to high radiance values, significantly alters the spectrum of the corresponding HS pixel. Thus, this particular HS spectrum can be extracted from each segment included (at least partially) in this HS pixel by the VCA method, and might be wrongly attributed to the subpixels belonging to all the mixed HS pixels joining each of these segments. That is why this issue can even affect subpixels remote

from reflective sources: even if a HS pixel covers no reflective subpixel in the reference image, the presence of reflective shapes in its neighbourhood is sufficient to affect the endmembers extracted and then selected during the process.

Fig. 6 shows the spectra associated with such a subpixel. The average normalised gap in the reflective domain between the reference spectrum and the reconstructed spectrum is 33 % for SOSU, as compared with 6 % for Gain. By comparing the spectral shapes, the SAM applied to the reflective domain between the reference spectrum and the reconstructed spectrum is 24° for SOSU, as compared with 6° for Gain. Yet, the correct endmember has been extracted (it has been correctly attributed to a neighbour subpixel), therefore it is clear that the issue comes from the spatial reorganisation step. Notably, the corresponding HS pixel has not been reorganised with the full combinatory analysis method (too many combinations to test), as confirmed in Fig. 7 in the next section.

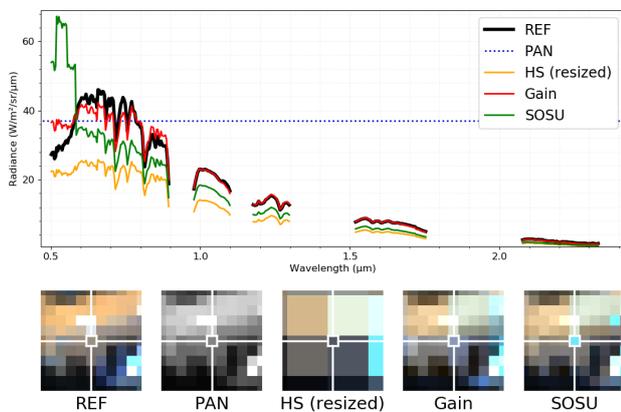


Figure 6. Subpixel degraded by SOSU preprocessing

### 4.3 Analysis of global quality criteria

Tables 1 to 4 compare the quality criteria between Gain and SOSU (in the reflective, VNIR and SWIR domains), respectively calculated for the full image, the mixed pixels reorganised with and without combinatory analysis, and the shadowed pixels.

		SAM (°)	RMSE	ERGAS	CC
Reflective	Gain	<b>8.1</b>	<b>4.2</b>	<b>8.1</b>	<b>0.93</b>
	SOSU	8.6	5.1	10.2	0.90
VNIR	Gain	<b>6.9</b>	<b>5.3</b>	<b>6.6</b>	<b>0.96</b>
	SOSU	7.1	6.6	8.1	0.94
SWIR	Gain	<b>1.9</b>	<b>2.8</b>	<b>9.2</b>	<b>0.91</b>
	SOSU	2.4	3.5	11.8	0.86

Table 1. Criteria computed on the full image

Overall, global criteria are quite close between Gain and SOSU, regardless of the spectral domain or the group of pixels, although Gain generally provides slightly better results than SOSU. One can still notice that SOSU is closer to Gain in the VNIR domain than in the SWIR domain. This is because several steps of the method rely on the PAN image as the only information source at the targeted spatial resolution (notably the spatial reorganisation step with the PAN reconstruction error, and the final fusion step with the gain derived from the PAN image). Yet, this PAN image is defined over the visible domain only, which almost entirely covers the VNIR domain. Thus, SOSU cannot systematically rely on the information provided

by the SWIR domain, taking better account of the VNIR domain.

		SAM (°)	RMSE	ERGAS	CC
Reflective	Gain	<b>7.1</b>	<b>2.4</b>	<b>7.0</b>	<b>0.96</b>
	SOSU	7.2	2.7	8.0	0.95
VNIR	Gain	<b>6.0</b>	<b>3.0</b>	<b>5.5</b>	<b>0.97</b>
	SOSU	<b>5.8</b>	3.3	6.0	<b>0.97</b>
SWIR	Gain	<b>1.8</b>	<b>1.7</b>	<b>8.0</b>	<b>0.95</b>
	SOSU	2.0	2.0	9.5	0.94

Table 2. Criteria computed on mixed pixels reorganised with combinatory analysis

		SAM (°)	RMSE	ERGAS	CC
Reflective	Gain	<b>8.7</b>	<b>5.1</b>	<b>8.0</b>	<b>0.91</b>
	SOSU	9.7	6.3	10.3	0.85
VNIR	Gain	<b>7.5</b>	<b>6.5</b>	<b>6.6</b>	<b>0.95</b>
	SOSU	8.0	8.1	8.2	0.92
SWIR	Gain	<b>1.9</b>	<b>3.4</b>	<b>9.1</b>	<b>0.87</b>
	SOSU	2.7	4.2	11.9	0.80

Table 3. Criteria computed on mixed pixels reorganised without combinatory analysis

Results on pure pixels are not displayed as Gain and SOSU are the same process in this case (see Section 2.2.3). Regarding the mixed pixels, however, it can be noticed that the ones reorganised with the complete combinatory analysis, although they are a minority (41% of the mixed pixels, see Fig. 7), imply better results than the ones reorganised with the alternative simpler approach (methods defined in Section 2.1.5). In the former case, the SAM, which we consider as the most relevant criterion (see Section 2.2.3), even provides better results with SOSU than with Gain in the VNIR domain (5.8° as compared with 6.0°).



(a) Pixels reorganised by combinatory analysis in purple (41 % of the mixed pixels) (b) Pixels reorganised by the alternative method in purple (59 % of the mixed pixels)

Figure 7. SOSU preprocessed image – localisation of each reorganisation method

This improvement induced by the complete combinatory analysis is confirmed by referring to Fig. 7, where almost all visible reorganisation errors come from the pixels reorganised with the alternative method. One must however keep in mind that this observation may be slightly biased since the alternative method is applied to a large number of mixed pixels, including most of the reflective subpixels and complex spatial structures of the image. Therefore, the choice of the reorganisation method is not the only factor explaining the lower results in this pixel group.

A last point concerns the shadowed pixels: according to the quality criteria, SOSU performs slightly better than Gain in the reflective domain with SAM (0.5° gap, i.e. a 5 % improvement). This enhancement is related to the better performance of SOSU in the VNIR domain, mainly for SAM and RMSE (respectively 13 % and 8 % improvements). This is an important

		SAM (°)	RMSE	ERGAS	CC
Reflective	Gain	9.4	<b>1.1</b>	<b>9.5</b>	<b>0.82</b>
	SOSU	<b>8.9</b>	1.2	10.8	0.76
VNIR	Gain	8.2	1.2	<b>5.3</b>	<b>0.89</b>
	SOSU	<b>7.1</b>	<b>1.1</b>	<b>5.3</b>	0.88
SWIR	Gain	<b>2.4</b>	<b>1.1</b>	<b>12.0</b>	<b>0.77</b>
	SOSU	2.7	1.2	14.0	0.65

Table 4. Criteria computed on shadowed pixels

result because shadowed areas represent not only a significant proportion of urban areas (here, 24 % of the image), but also a specific and challenging aspect of this type of scenes.

#### 4.4 Analysis of local quality criteria

Fig. 8 depicts four error maps calculated with the SAM criterion in the VNIR and SWIR domains, for both images fused by Gain and SOSU. They are used to locally check if the correct materials have been assigned to each subpixel of the resulting images (see Section 2.2.3).

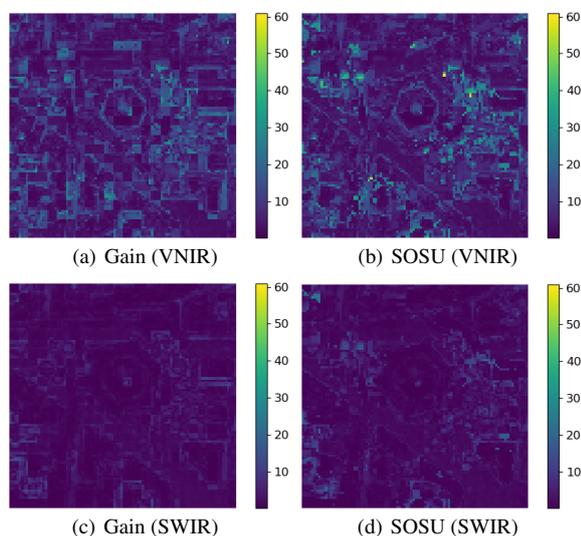


Figure 8. SAM error maps (in degrees) obtained in VNIR and SWIR for images fused by Gain and SOSU

Regarding the VNIR domain, both SOSU and Gain error maps contain 14 % of pixels with a negligible SAM value (lower than 2°). For high SAM values (more than 10°), the ratios still remain close between Gain and SOSU, with a very slight advantage for SOSU (22 % of the image, as compared with 23 % for Gain). By visually comparing the VNIR error maps, one can notice that the errors are more localised with SOSU, but with, punctually, some remarkable and very high values. The latter correspond to non-reflective pixels associated with reflective materials, or to reflective pixels associated with non-reflective materials.

Regarding the SWIR domain, results are more distinct between both methods. The negligible SAM values represent 69 % of the image for Gain, as compared with 63 % only for SOSU. This gap is even more important with the high SAM values, which represent less than 1 % of the image processed by Gain, in comparison with almost 3 % of the image processed by SOSU. This confirms that SOSU performs slightly better in the VNIR domain than in the SWIR domain, as already established by the global analysis (Section 4.3).

It is possible to determine the ratios of mixed pixels that have been better processed by SOSU than by Gain (i.e. strict improvement ratios), by comparing the corresponding SAM values for each given mixed pixel, in the chosen spectral domain. Thus, 48 % of the mixed pixels are better processed by SOSU in the reflective range, 50 % in the VNIR range, but only 43 % in the SWIR range. These ratios are enhanced if we only focus on mixed pixels reorganised by the complete combinatory analysis method: we get 49 % of improved pixels in the reflective range, 51 % in the VNIR range, and 46 % in the SWIR range. Eventually, if we focus on shadowed pixels, we get even better ratios, with 52 % of improved pixels for the reflective range, 57 % for the VNIR range and 46 % for the SWIR range.

#### 4.5 Synthesis of the analyses

On the one hand, the visual analysis highlighted the interest of SOSU preprocessing, by revealing an accurate reconstruction of the scene at the PAN spatial resolution before using the Gain process. On the other hand, numerical results from both global and local analyses are not different enough to rank SOSU and Gain, even if, by strictly referring to quality criteria values, Gain remains slightly better. Nevertheless, these numerical analyses still reveal several advantages for SOSU. First of all, specific pixels largely represented in urban landscapes, like shadowed areas, are in favor of SOSU. Then, we highlighted that, in the VNIR range, SOSU performs better than in the SWIR range (by considering the closeness of results between Gain and SOSU), and is generally more relevant (both visually and numerically) than Gain.

### 5. CONCLUSION AND FUTURE WORK

SOSU, an HS pansharpener method, has been presented. It is an evolution of SOSU-2019 from previous work (Constans et al., 2020), which had been tested on relatively homogeneous areas (agricultural and peri-urban scenes). The general method is based on a fusion process of the literature preserving the PAN spatial information (Gain), to which we added preprocessing steps (segmentation, mixed pixel detection, endmember extraction, endmember selection, and spatial reorganisation) to improve the spectral content at the PAN resolution. SOSU-2019 still required enhancements to be applied to complex environments like urban scenes. These enhancements mainly concerned the spatial reorganisation step.

In this article, SOSU has been tested on an urban dataset provided by the hyperspectral HySpex instrument (UMBRA 2012) at a spatial resolution of 1.6 m and acquired on the Toulouse city. To evaluate the fusion results, a performance assessment protocol, based on Wald's protocol and using spatial, spectral and global quality criteria specifically adapted to complex areas, has been proposed. It has been used to compare SOSU with the reference method, Gain, via visual and numerical analyses. The chosen criteria have been applied to different spectral ranges (reflective, VNIR, SWIR), and spatially to different pixel groups (full image, mixed pixels with each reorganisation method, shadowed pixels) and at different scales (global measures and local error maps), to localise the error sources and thus refine the performance assessment of SOSU.

These complementary analyses revealed very close performance for Gain and SOSU. However, some advantages of SOSU have been highlighted, including a greater fidelity to the spatial organisation of the scene at the PAN spatial resolution (from

the visual analysis), encouraging numerical results in the VNIR range (50 % of the mixed pixels are improved by SOSU, and 51 % if we focus on mixed pixels reorganised by the optimal approach), as well as a better shadowed pixel processing (52 % of improvement in the reflective range, and 57 % in the VNIR range).

Nevertheless, this performance can still be improved by enhancing SOSU, to definitely outperform Gain. An important point of consideration to refine SOSU results in the reflective domain will be taking better account of the SWIR range, by applying for example a normalisation method to the HS spectra, to balance the contributions of all spectral bands to SOSU preprocessing. Then, our future work will include evaluating the final method for different spatial resolutions (varying sampling rates and modulation transfer functions) and for HS/PAN resolution ratios from 2 to 10.

## REFERENCES

- Adeline, K., Le Bris, A., Coubard, F., Briottet, X., Paparoditis, N., Viallefont, F., Riviere, N., Papelard, J.-P., David, N., Déliot, P. et al., 2013. Description de la campagne aéroportée UMBRA: Étude de l'impact anthropique sur les écosystèmes urbains et naturels avec des images THR multispectrales et hyperspectrales. *Revue française de photogrammétrie et de télédétection*, 202, 79–92.
- Bioucas-Dias, J. M., Nascimento, J. M., 2008. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8), 2435–2445.
- Briottet, X., Feret, J.-B., Jacquemoud, S., Lelong, C., Rocchini, D., Schaepman, M. E., Sheeren, D., Skidmore, A., Somers, B., Gomez, C., Chabrillat, S., Schmid, T., Stenberg, B., Marion, R., Carrère, V., Delacourt, C., Chami, M., McKee, D., Bajjouk, T., Gege, P., Minghelli, A., Gamba, P., Weber, C., Shimoni, M., Foucher, P.-Y., Chanussot, J., Lopez, S., Lefevre, M.-J., Binet, R., 2017. European Hyperspectral Explorer: HYPEX-2 – A new space mission for vegetation biodiversity, bare continental surfaces, coastal zones and urban area ecosystems. *10th EARSeL SIG imaging spectroscopy workshop*.
- Christoudias, C. M., Georgescu, B., Meer, P., 2002. Synergism in low level vision. *null*, IEEE, 40150.
- Comaniciu, D., Meer, P., 2002. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 603–619.
- Constans, Y., Fabre, S., Brunet, H., Seymour, M., Crombez, V., Chanussot, J., Briottet, X., Deville, Y., 2020. Fusion of hyperspectral and panchromatic data by spectral unmixing in the reflective domain. working paper or preprint.
- Jagalingam, P., Hegde, A. V., 2015. A review of quality metrics for fused image. *Aquatic Procedia*, 4, 133–142.
- Kettaf, F., Bi, D., de Beauville, J. A., 1996. A comparison study of image segmentation by clustering techniques. *Proceedings of Third International Conference on Signal Processing (ICSP'96)*, 2, IEEE, 1280–1283.
- Kruse, F. A., Lefkoff, A., Boardman, J., Heidebrecht, K., Shapiro, A., Barloon, P., Goetz, A., 1993. The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3), 145–163.
- Lier, P., Valorge, C., Briottet, X., 2012. *Satellite Imagery From Acquisition Principle to Processing of Optical Images for Observing the Earth*. CEPADUES Editions: Toulouse, France.
- Loncan, L., 2016. Fusion of hyperspectral and panchromatic images with very high spatial resolution. PhD thesis, Université Grenoble Alpes.
- Loncan, L., De Almeida, L. B., Bioucas-Dias, J. M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G. A., Simoes, M. et al., 2015. Hyperspectral pansharpening: A review. *IEEE Geoscience and remote sensing magazine*, 3(3), 27–46.
- Nagao, M., Matsuyama, T., Ikeda, Y., 1979. Region extraction and shape analysis in aerial photographs. *Computer Graphics and Image Processing*, 10(3), 195–223.
- Nascimento, J. M., Dias, J. M., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(4), 898–910.
- Pei, W., Wang, G., Yu, X., 2012. Performance evaluation of different references based image fusion quality metrics for quality assessment of remote sensing image fusion. *2012 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2280–2283.
- Sabins, F. F., 2007. *Remote sensing: principles and applications*. Waveland Press.
- Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., Restaino, R., Wald, L., 2014. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5), 2565–2586.
- Wald, L., 2000. Quality of high resolution synthesised images: Is there a simple criterion ? T. Ranchin, L. Wald (eds), *Third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, SEE/URISCA, Sophia Antipolis, France, 99–103.
- Wald, L., Ranchin, T., Mangolini, M., 1997. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric engineering and remote sensing*, 63(6), 691–699.
- Wu, C., 2009. Quantifying high-resolution impervious surfaces using spectral mixture analysis. *International Journal of Remote Sensing*, 30(11), 2915–2932.
- Yokoya, N., Grohnfeldt, C., Chanussot, J., 2017. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2), 29–56.
- Yoo, J.-C., Han, T. H., 2009. Fast normalized cross-correlation. *Circuits, systems and signal processing*, 28(6), 819.