MDPI

Article

Assessment of the usefulness of spectral bands for the next generation of Sentinel-2 satellites by reconstruction of missing bands

Jordi Inglada [1], Julien Michel[1], Olivier Hagolle[1]

[1] CESBIO, Université de Toulouse, CNES/CNRS/IRD/UPS, Toulouse, France
* jordi.inglada@cesbio.eu, Tel. +33.561.558.676, Fax. +33.561.558.500.

Abstract: The Sentinel-2 constellation provides high spatial, spectral and temporal resolution optical imagery of the continental surfaces since 2015. The spatial and temporal resolution improvements that Sentinel-2 brings with respect to previous systems has been demonstrated in both the literature and operational applications. On the other hand, the spectral capabilities of Sentinel-2 appear to have been exploited to a limited extent only. At the moment of definition of the new generation of Sentinel-2 satellites, an assessment of the usefulness of the current available spectral bands seems appropriate. In this work, we investigate the unique information contained by each 20 m resolution Sentinel-2 band. A statistical quantitative approach is adopted in order to yield conclusions which are application agnostic: multivariate regression is used to reconstruct some bands using the others as predictors. We conclude that, for most observed surfaces, it is possible to reconstruct the reflectances of most Red Edge or NIR bands from the rest of observed bands with an accuracy within the radiometric requirements of Sentinel-2. In some cases, it is possible to reconstruct a pair of non observed bands. We also identify mission scenarios for which several of the current Sentinel-2 bands could be removed for the next generation of sensors.

Keywords: spectral bands; Sentinel-2; regression

Citation: Inglada, J.; Michel, J.; Hagolle, H Reconstruction of missing Sentinel-2 spectral bands. *Remote Sens.* **2022**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Submitted to *Remote Sens.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/ 4.0/).

1. Introduction

The Sentinel-2 constellation constitutes a revolution in the remote sensing field in terms of data quantity, quality and availability. The high spatial and temporal resolutions of Sentinel-2 [1] have been demonstrated to be crucial for many applications that have been reported in the scientific literature and validated by operational applications covering a wide range of use cases like land cover mapping, snow extent mapping, biophysical parameter estimation, agriculture monitoring, etc.

Sentinel-2 provides 13 spectral bands with spatial resolutions from 10m to 60m and a 5-day revisit cycle.

The particularities of Sentinel-2 with respect to pre-existing comparable systems are:

- in the temporal domain, a systematic acquisition plan (unlike tasked satellites which acquire scenes on demand) with a high revisit frequency (5 days compared to the 16 days of Landsat);
- in the spatial domain, a higher resolution than Landsat (10 m to 20 m compared to 30 m);
- in the spectral domain, an increased number of bands with respect to both the classical Blue, Green, Red, NIR band set and Landsat (4 visible, 1 NIR, 2 SWIR), with the novelty of 3 red edge (RE) bands, although a lack of thermal band with respect to Landsat.

However, as we show in section 1.4, very few published works have made full use of the spectral richness of Sentinel-2 and, often, these uses have not been demonstrated to be the only way to extract the target information.

After 5 years of operations, the work on the new generation of Sentinel-2 satellites (S2NG) has started, and one of the tasks is to identify the set of spectral bands. The question of «which additional spectral bands could be put on board of S2NG» has to be balanced with the one of

the «S2 possible useless bands», that is, the current available bands which could be removed for S2NG. Adding spectral bands to a satellite bears a cost which could impact the trade-off with other mission requirements, like temporal revisit needing an additional satellite.

Of course, all current Sentinel-2 bands contain *potentially useful* information, since they sample different intervals of the electro-magnetic spectrum, and except for the pair B8-B8A (see section 1.1), there is no significant overlap between the different spectral ranges. However, since there exists a high amount of redundance in the underlying observed nature, one can expect high degrees of correlation between the different bands, allowing us to *question the true usefulness* of some bands.

With 5 years of data collection and exploitation, it is now possible to quantitatively assess the usefulness of the different bands on board of Sentinel-2. This could be done in terms of the quality of the results of downstream processing (biophysical parameter estimation, land-cover mapping, etc.) but this would need to address a huge number of application domains or review lots of literature without the guaranty of exhaustivity or chances of replicability.

On the other hand, if we address the problem from the *information content* point of view, we only have to deal with data at the sensor level. We therefore choose to pose the problem as a data reconstruction one: if one band can be reconstructed – within a predefined error margin – from the other bands, it can be removed from the satellite without quantitative loss of information.

One could argue that, what matters is the estimation of physical parameters and that an imperfect reconstruction of a particular band can have no impact for many applications. This would allow to go further in terms of spectral band removal. We agree with this point of view, but all downstream processing entails the use of (imperfect by construction) models, and, the closer we get to the sensor, the most application independent the conclusions of the study will be.

The aim of this paper is to leverage this interband correlation and assess which bands could be removed from future iterations of the Sentinel-2 constellation with a minimal impact on the usefulness of the acquired data. To do so, we *predict* the reflectances of missing bands with non-linear regression algorithms that use the other spectral bands as *predictors*. In order to produce results which are representative of real settings and which are generalizable, we build a data set by sampling pixels from Sentinel-2 acquistions with a wide variety of geographic areas and dates. We therefore take an empirical, data-driven approach.

We choose not to leverage the spatial and the temporal dimensions and carry out a mono-date, pixel-based analysis. We understand that temporal and spatial correlations would reduce the errors in the reconstruction of missing bands. The goal of the work is not to propose an optimal regression algorithm, but rather showing that band reconstruction is possible using regression. The results of this work can be seen as a lower bound in terms of reconstruction quality and therefore encourage the pursuit of further studies.

1.1. The Sentinel-2 spectral bands

Table 1 gives the name and the central wavelength for each band acquired by Sentinel-2. There are 4 bands at 10m resolution, the 3 usual visible bands (B2-B4) and a wide NIR band (B8). The 20m resolution bands are 3 narrow bands in the red edge (B5-B7), one narrow NIR (B8A) and 2 SWIR bands (B11, B12). Finally, the 60m resolution bands are aimed at radiometric corrections (B1 for aerosol content estimation, B9 for water vapor and B10 for cirrus detection). Figure 1 illustrates the relative spectral responses of the 10m, 20m and 60m resolution bands.

1.2. S2 Radiometric Requirements

The Sentinel-2 Mission Requirements Document (MRD) [2] states that for the applications covered by this mission, the radiometric accuracy at top of atmosphere (TOA) has to be not worse than 3% (goal) to 5% (threshold). For inter-band radiometric calibration, 3% accuracy is also required.

These requirements allow to define error bounds for the band reconstruction tasks that we assess in this work. For TOA reflectances, we can aim the 3% reconstruction error. In

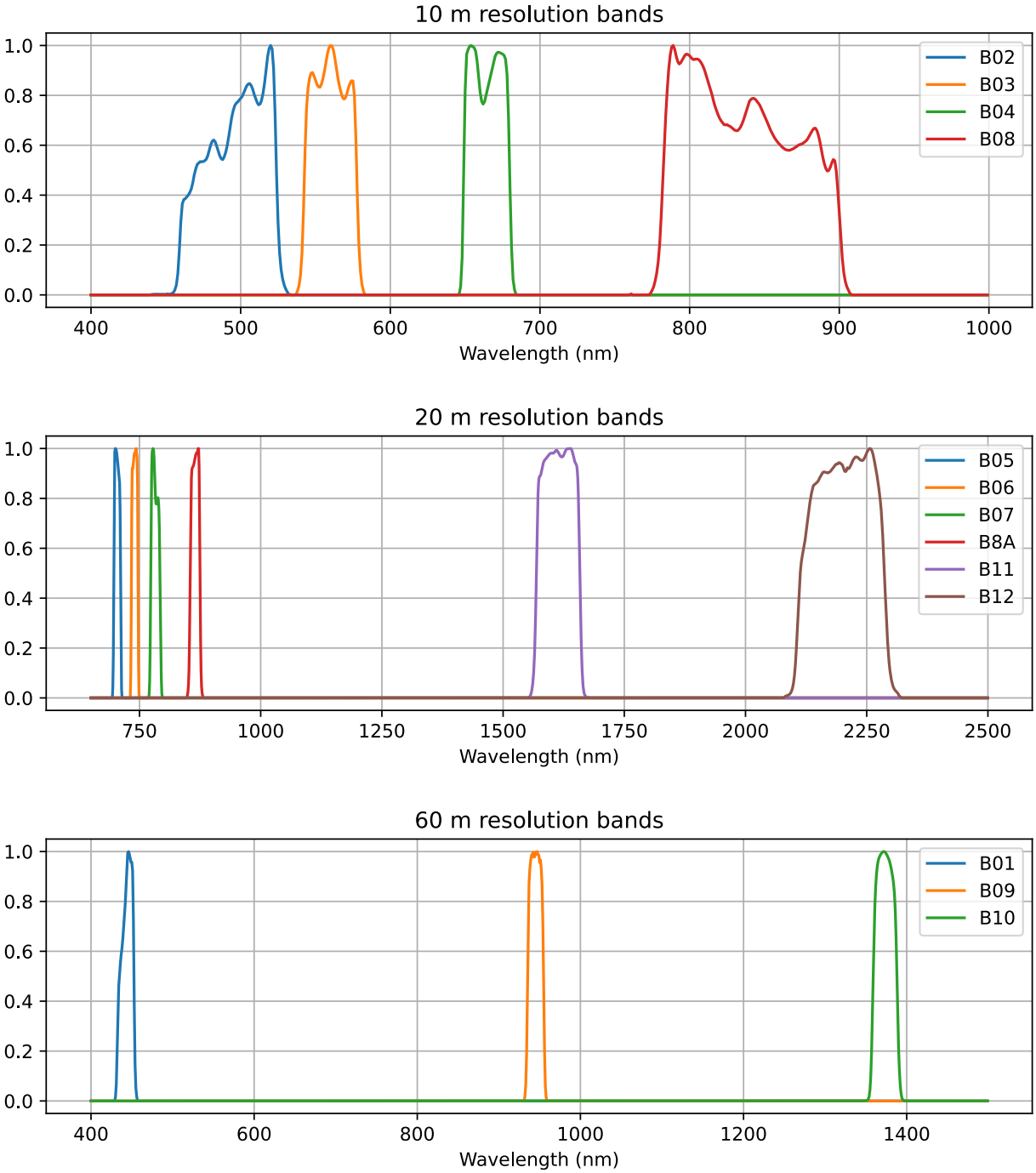


Figure 1. Sentinel-2 relative spectral responses

terms of surface reflectance, the accuracy of the MAJA processor [3][4] is 0.01 (not in %, but in reflectance values) and we can use this value as requirement.

Given the fact that there are other errors in the measure (geometric registration between bands, MTF differences, etc.), achieving these error bounds can be considered rather ambitious.

Other approaches to define the reconstruction requirements could be used. For instance [5] presents a Radiometric Uncertainty Tool which allows to estimate the radiometric uncertainty associated with each pixel of a Sentinel-2 image in the TOA images provided by ESA. The approach integrates all the errors from the TOA reflectance to the L1C product and typical

Table 1: Name and central wavelength for the Sentinel-2 spectral bands [1].

Band	Central wavelength (nm)	Spatial resolution (m)
1 – Coastal aerosol	442.7	60
2 – Blue	492.4	10
3 – Green	559.8	10
4 – Red	664.6	10
5 – Vegetation red edge	704.1	20
6 – Vegetation red edge	740.5	20
7 – Vegetation red edge	782.8	20
8 – NIR	832.8	10
8A – Narrow NIR	864.7	20
9 – Water vapour	945.1	60
10 – SWIR – Cirrus	1373.5	60
11 – SWIR	1613.7	20
12 – SWIR	2202.4	20

values are greater than 10% for open sea, 5% to 15% in rice fields covered by water and 2% to 4% in land areas. We see that the 3% specification is very strict.

1.3. Directional effects

Since the reflectance of surfaces depends on the observation and illumination directions [6], particular attention has to be payed to the acquisition geometry. Directional effects are specially important in (nearly) specular reflections, but also in the case of shadow or volume effects.

The MSI instrument is composed of 2 focal planes covering the VNIR and the SWIR channels respectively, each one having an array of 12 detectors. Due to the shifted positioning of the detectors along the track direction on the focal planes, angular differences between the two alternating odd and even clusters of detectors are induced in the measurements. The parallax Base/Height (B/H) ratio ranges between 0.022 and 0.059. A similar issue occurs between the VNIR and SWIR detectors, resulting in an inter-band B/H which is less than 0.01 for the VNIR channels and less than 0.018 for the SWIR.

The values of the solar and sensor angles on a 5 km grid are provided in the L1C product meta-data. We will leverage this information in the band reconstruction algorithms that will be used in this work.

1.4. Specific uses of S2 bands

The spectral bands of Sentinel-2 allow the computation of a large variety of spectral indices other than NDVI which are useful for different applications. Table 2 presents a selection of several of them.

The RE bands have been proposed for chlorophyll estimation, burnt severity assessment [7], LAI estimation [8] and non photosynthetic vegetation [9]. The SWIR bands have been proposed for dry mass vegetation [11] and water or moisture indices [13].

Although a thorough review of the literature is out of the scope of this paper, a bibliometric analysis shows that very few papers published after the launch of Sentinel-2 make an explicit use of the spectral particularities (RE and SWIR bands). Furthermore, a recent review about phenology monitoring using Sentinel-2 [14] shows that only 1 out of 4 published papers uses other spectral information than NDVI.

Some studies as for instance [15] claim that RE and SWIR bands during vegetation senescense appear as being important for machine learning based classification. The concept of importance has to be nuanced, since it measures the errors made when the reflectance of those bands are replaced by random values. In order to have an accurate assessment of the usefulness of those variables, the classifiers should have to be re-trained without them. On the other hand, the same work shows that PSRI, which is computed from red, green and NIR is

Table 2: Spectral indices leveraging Sentinel-2 spectral bands

Index	Formula	Application	Reference
$CI_{red-edge}$	$\left(\frac{B7}{B5}\right)-1$	Chlorophyll, burnt areas	[7]
CI_{green}	$\left(\frac{B7}{B3}\right)-1$	"	"
REP	$705+35\frac{\frac{(B4+B7)}{2}-B5}{B6-B5}$	"	"
$MTCI$	$\frac{B6-B5}{B5-B4}$	"	"
$NDRE1$	$\frac{B6-B5}{B6+B5}$	"	"
$NDRE2$	$\frac{B7-B5}{B7+B5}$	"	"
$TRBI$	$\frac{B12+B6}{B8A}$	LAI estimation	[8]
$NSSI$	$\frac{B8A-B7}{B8A+B7}$	Non photosynthetic vegetation	[9]
$PSRI$	$\frac{B4-B2}{B6}$	Senescent vegetation	[10]
STI	$\frac{B11}{B12}$	Tillage, dry vegetation	[11]
$NDWI$	$\frac{B3-B8A}{B3+B8A}$	Water bodies	[12]
$NDWI$	$\frac{B8-B11}{B8+B11}$	"	[13]
$NDWI$	$\frac{B8-B12}{B8+B12}$	"	"

also *important*, which may indicate a high correlation (and therefore redundancy) with RE bands.

Another work supporting the interest of RE and SWIR bands is [16], where they are shown to be useful for Gross Primary Productivity estimation in grasslands. Using regression approaches, the authors show that those bands are useful to predict the target variable, but do not study whether using more complex regressors, the error without those bands could be reduced.

It is interesting to note that other works like for instance [17], show that NDVI is best suited to monitor grass phenology than more sophisticated VIs using RE and SWIR bands. Another example is [18], where it is shown that the RE bands of Sentinel-2 do not perform well for the estimation of chlorophyll content changes in certain crops. One should note that, before the launch of Sentinel-2 the same community had great expectations for these bands and for the same application [19]. However, at the time, the authors already suggested that using the green band in CI_{green} also seemed very promising and therefore further research was required.

The apparent contradictions between these different works are likely due to the fact that different experimental settings, different data and different applications were involved.

Also, we find that the works on the usefulness of spectral bands are usually addressed only from the point of view of demonstrating that a particular phenomenon has a signature in a particular band. For instance, a recent publication [20] proposed additional bands in the SWIR in order to detect non photosynthetic vegetation and crop residues. The study shows indeed that these objects can not be detected with the SWIR bands of Landsat-8. However the cited work does not analyze how the complete set of Landsat-8 bands could be used to retrieve a signature of the phenomenon at hand.

At the moment of this writing, and to the best of our knowledge, the most thorough study of the *usefulness* of Sentinel-2's spectral bands is [21]. This reference is actually a detailed litterature review of the use of hyperspectral imagery with the goal of proposing synergies with Sentinel-2 in order to overcome the limitations of space-borne hyperspectral sensors (spatial resolution, revisit time and signal to noise ratio). Interestingly, the review shows how the current set of Sentinel-2 bands constitutes in itself a very wise choice for many applications. However, the limit of such a meta-analysis, is that there can't be a wholistic view of the problem, since the pertinence of each spectral range is performed in isolation in the reviewed litterature. Indeed, this prevents from discovering redundancies between different bands.

We think that this supports the idea of performing a pure data driven approach over a large data set and with an application agnostic point of view. However, the work presented in this paper is just a modest demonstration of what could be done by exploiting the existing Sentinel-2 archive.

Finally, we will stress again the fact that we don't claim that some Sentinel-2 bands do not contain useful information. We want to assess the possibility of reconstructing this information by leveraging redundancies among the complete set of spectral bands. This reconstruction will of course contain errors, and the goal here is to give bounds allowing to inform design trade-offs for future systems.

2. Data preparation

For this study, a set of 128 Sentinel-2 tiles was used. Figure 2 illustrates the geographic distribution of these tiles. For each tile, a single date was used and the selection was random on the period going from early 2016 to the end of 2020. The goal was to cover a wide range of geographic areas and seasons. For each acquisition, the data was obtained at 2 processing levels: 1C (from PEPS, CNES' mirror of Sentinel data) and 2A (from Theia's catalogue), the latter having been produced by the MAJA processor. This allows us to use accurate masks at the pixel level for clouds, cloud shadows and saturation effects.

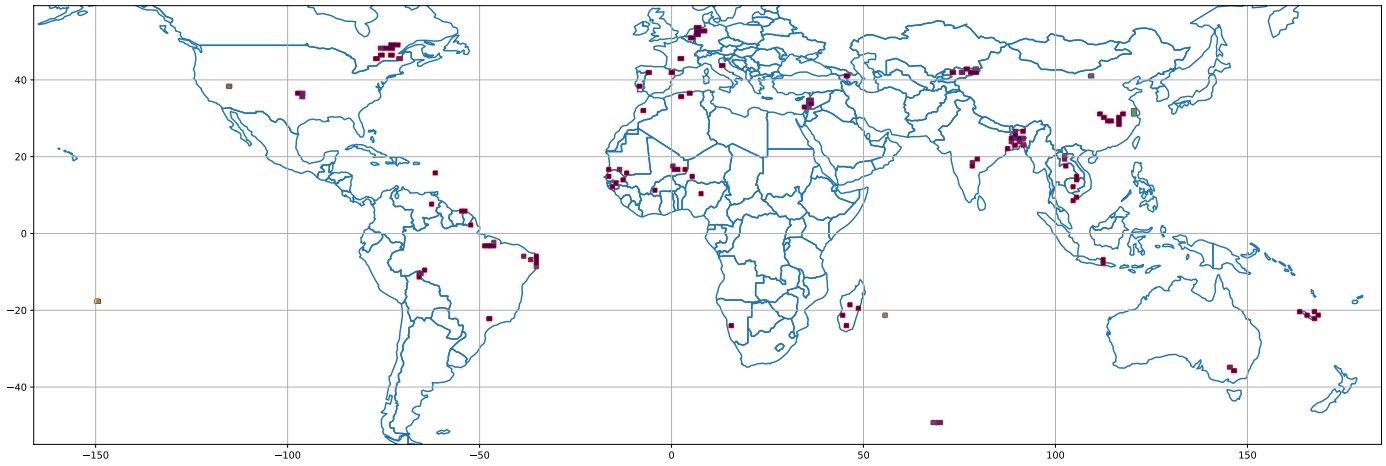


Figure 2. Geographic distribution of the tiles used for the study

For each acquisition, 100,000 pixels where sampled. Only non-saturated pixels were selected, regardless of their cloud or shadow status. Pixel positions were selected on the 20m resolution grid. For each 20m pixel position, the following information was recorded:

- whether the pixel was detected as a cloud or a shadow (without distinction of these 2 states),
- the reflectance in the 20m bands for levels 1C and 2A,
- the reflectance of the 4 corresponding pixels of each of the 10m resolution bands for levels 1C and 2A,
- the reflectance at the 20m pixel position of the 60m resolution bands after bicubic resampling for level 1C,
- the solar and viewing angles for each pixel.

For the analyses performed in the following sections, we split the data at the tile level. This means that all the pixels used for testing purposes (measures of accuracy of the reconstructions) belong to tiles for which no pixel was used for training or even intermediate validation.

In the experiences carried out in this work, we randomly select 100 tiles and we do a 80%/20% split at the tile level for training and testing purposes. This means that training and testing pixels come from different tiles and dates. The training set is further split in proper training samples (80%) and validation samples (20%), the latter being used for monitoring the convergence of the training. For each experiment (i.e. set of predicted bands and set of predictor bands) the experiment is repeated 10 times by selecting a different set of 100 tiles

among the 128 available. This allows to check for possible selection biases and allows to further assess the robustness of the regressions.

Also, only clear pixels (non cloudy nor shadow) are used for training and validating models. This reduces the number of available pixels. In average, each experiment uses 3.86928e+06 training samples, 967320 validation samples and 1.2582e+06 testing samples and is repeated 10 times.

The dataset has been made public [22] and is available for other researchers to reproduce and improve the work presented in this paper.

3. Regression model

As stated in the introduction, we aim at estimating a subset of the Sentinel-2 bands from the other ones. This estimation will be done using regression techniques. The regression algorithms will be calibrated and validated using the data described in section 2. In this section we describe the regression approach chosen.

3.1. Reflectance estimation with associated uncertainties

The regression problem is posed as the estimation of one or several spectral bands as a nonlinear combination of a disjoint set of the available bands. For the prediction of a single band, this can be written as:

$$\widehat{\rho_i} = f(\{\rho_{j\neq i}\}, \vec{\theta}),$$

that is, the prediction of the reflectance of band i is a function of the measured reflectances of the other bands and a set of parameters $\vec{\theta}$ containing other pertinent information, like solar and sensor angles. The regression can jointly estimate several spectral bands in a set I:

$$\{\widehat{\rho_i}\}_{i\in I} = f(\{\rho_j\}_{j\notin I}, \vec{\theta}) \tag{1}$$

The regression procedure should also produce a credibility interval[1] of the estimation of the target variable. In order to do this, instead of regressing over the expected mean, we can implement a regression of the mean and the variance of the target variable. Estimating a mean and a variance means that we are assuming a Gaussian error model.

At inference (estimation) time, the mean will be used as the variable estimation (in a Gaussian model the mean is the value with the highest probability), and the variance will be used to give the credibility interval.

Given a target value y (in our case that would be ρ_i) and the estimates $\hat{\mu}$ and $\hat{\sigma}$, the predictive likelihood of the target value given the estimates is the Gaussian distribution whose probability density function is

$$p(y|\hat{\mu},\hat{\sigma}) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} e^{\frac{(y-\hat{\mu})^2}{2\hat{\sigma}^2}}$$

We can therefore pose the regression problem as the minimization of a cost function given by the negative log-likelihood [23]. The log-likelihood takes the form:

$$log\big(p(y|\hat{\mu},\hat{\sigma})\big) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\hat{\sigma}^2 - \frac{(y-\hat{\mu})^2}{2\hat{\sigma}^2}$$

And therefore, after removing the constant term and a multiplicative factor, the cost function to be minimized is:

$$\mathscr{L} = \sum_i \log\hat{\sigma}_i^2 + \frac{(y_i-\hat{\mu}_i)^2}{\hat{\sigma}_i^2}$$

[1] We use the term *credibility interval* instead of *confidence interval* because we adopt a Bayesian point of view: we consider that the estimated value is a random variable and that the bounds of the interval al fixed, while the use of confidence intervals considers the bounds as random variables that result from repeated measures.

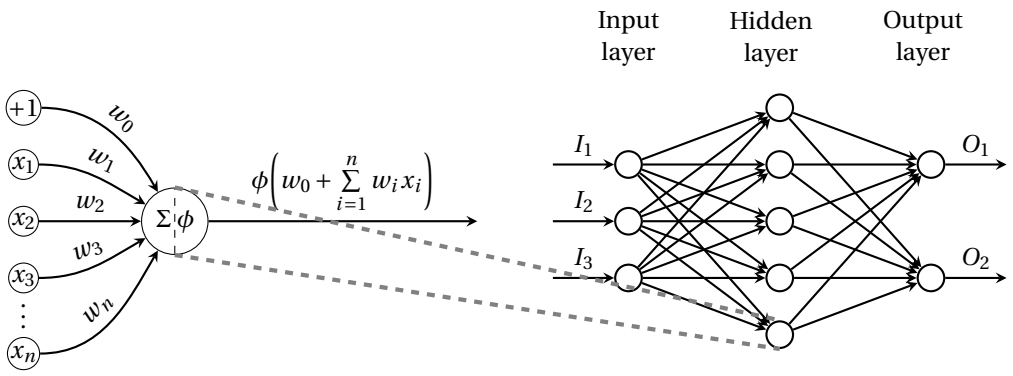


Figure 3. A multi-layer perceptron with 1 input layer, 1 hidden layer and 1 output layer. Diagram adapted from https://github.com/PetarV-/TikZ

where the sum is taken over the training samples.

Beyond being the correct theoretical loss under a Gaussian error model, this penalty function can be intepreted as follows :

- the term $(y_i - \hat{\mu}_i)^2$ penalizes the errors between the target value and the estimated mean;
- these errors, are weighted by the uncertainty estimation: larger errors will need larger values of σ_i to lower the penalty;
- in order to avoid allowing large errors on μ_i by always estimating large values of σ_i, large values of σ_i are also penalized by the first term in the loss.

3.2. Regression algorithm

The regression algorithm will have to find the approximation to the function f in equation 1 minimizing the cost function described above. Since we don't have prior knowledge about the shape of f, we choose to use a non parametric approach. Among the non parametric algorithms for regression, feed-forward neural networks (multi-layer perceptrons, MLP) seem a good choice because they are universal function approximators [24] which can be used in a multi-variate input and output setting and with custom cost functions. Conversely, other choices have limitations. For instance, linear and logistic regressions impose a strong prior on the shape of f and Random Forest regression can't predict several targets. The main drawback of neural networks is their lack of interpretability.

MLP are composed of fully connected linear layers (sets of neurons computing linear combinations of the inputs) followed by nonlinearities ϕ called activation functions. Figure 3 illustrates an MLP with a single hidden layer with 5 neurons. A large number of layers with different numbers of neurons can be used. Training such a network consists in finding the set of weights w_i that minimize the loss function for the set of training samples. The optimization is done by stochastic gradient descent.

Another interesting property of MLPs is that they can be combined as elementary bricks in more complex architectures. This will allow to introduce some structure in the processing, which brings interpretability and the possibility of introducing some prior knowledge. We will develop this point in the next section.

3.3. Network architecture

As stated above, the regression neural network will produce an estimation of the reflectances of the target bands using the reflectances of the other bands as predictors. All computations are performed for individual pixels. In order to take into account BRDF effects, the solar and sensor angles (both azimuth and zenith, as described in 1.3) are also used as predictors. More precisely, the sinus and cosinus of each angle are used.

Instead of using all predictors (reflectances and angles) together in a flat vector as input of an MLP as in figure 3, we use an attention mechanism where the angular information modulates the spectral values. This is implemented as illustrated in figure 4. First, the spectral and

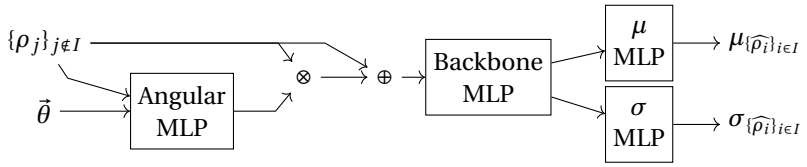


Figure 4. Overview of the nonlinear regression of a set of spectral bands using other bands and angular information as predictors assuming a Gaussian error model.

angular information are fed to the *Angular MLP* which is used to generate an attention mask. An attention mask is a vector of real numbers in $[0,1]$ with the same number of components as the data on which the attention is being applied. In our case, this is the vector of spectral bands. The *Angular MLP* is a standard MLP, with a single hidden layer containing 8 neurons and a *SoftMax* layer as output. The *SoftMax* function is an exponential normalization that maps a set of values to the unit interval (simplex in more than one dimension) $\sigma : \mathbb{R}^K \rightarrow [0,1]^K$ and is defined by:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad \text{for } i = 1,\dots,K \text{ and } \mathbf{z} = (z_1,\dots,z_K) \in \mathbb{R}^K,$$

where the z_i are the outputs of the layer preceding the *SoftMax*.

Therefore, the *Angular MLP* learns a set of multiplicative weights (this operation is represented by the $\otimes$ symbol in figure 4) that will be applied to the input reflectances in order to perform an angular correction. It is interesting to note that this angular correction takes into account the spectral information itself, that is, the reflectances and the angles are both used for the estimation of the correction. It is therefore a kind of self-attention mechanism [25].

A residual connection (a simple elementwise addition represented by $\oplus$ in figure 4) is used after the attention mask in order to keep spectral information that could be excessively removed by the attention mechanism before entering the *Backbone MLP*. The latter is used to embed the predictors into a feature space that will be used to feed the 2 modules used for the estimation of the target values and their uncertainties respectively.

The backbone part (a 3 hidden layer MLP with 10 neurons per layer) allows to model the correlation between the target variables and their uncertainties. The independent MLP branches (with the same architecture as the backbone) for μ and σ get specialized into the estimation of each of the informations. Performing the regression for several target variables with the same network is a kind of multi-task learning which is able to leverage the correlation between target variables and is more efficient than preforming single target regressions.

For numerical stability and positivity constraints, instead of estimating the σ or σ^2, we estimate $\log\sigma$.

The output activation functions for the mean and the variance estimations are hyperbolic tangents so that the values are contained in the $[-1,1]$ interval. The output value is then rescaled into a pre-defined interval which is $[-0.2,1.3]$ for μ and $[1e-5,1.5]$ for σ^2. The rescaling for μ allows to take into account the fact that L2A reflectances can sometimes be negative due to over-corrections. Reflectances can also be higher than 1 in specular conditions. The rescaling intervals could be learned from the data, but we set them for simplicity.

The regression of several bands simultaneously is done by a straightforward extension of the single target case. The output layers, for both the means and the variances will have as many neurons as target variables. The loss function is just de sum of the losses for each target variable.

The network is trained for 100 epochs using an Adam optimizer [26] with a learning rate of 0.001 and a batch size of 256.

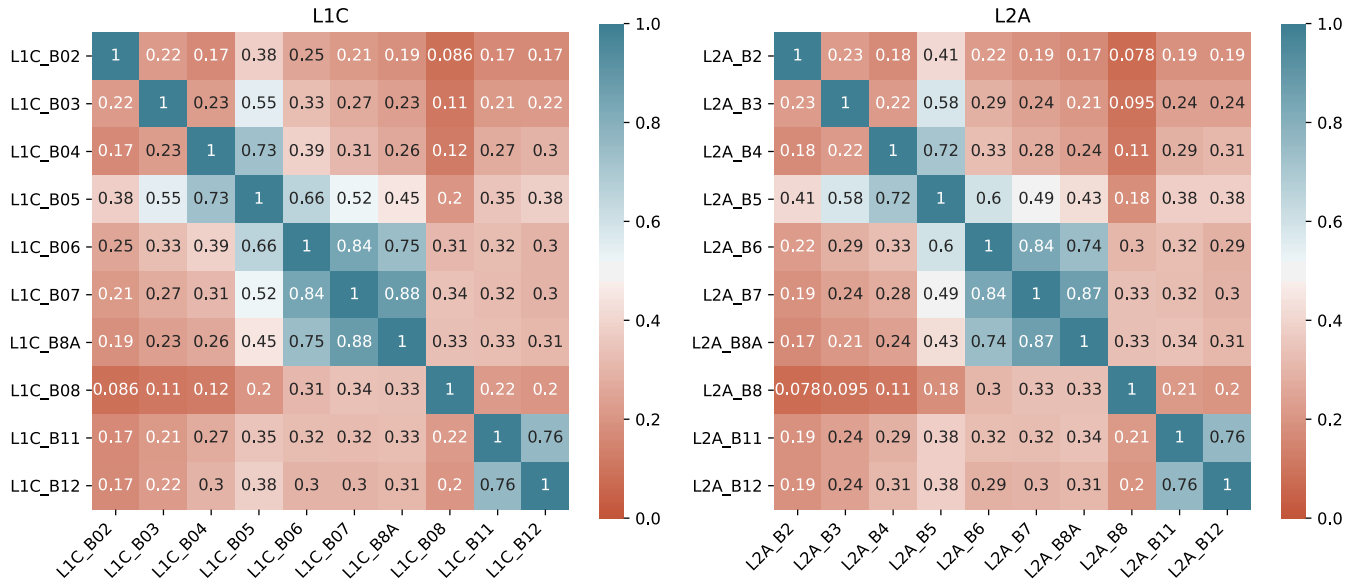


Figure 5. Normalized mutual information

4. Results

4.1. Redundancies in Sentinel-2 bands

Before assessing the quality of the spectral regression approaches, we analyze the statistical dependence between all the pairs of Sentinel-2 bands. Instead of measuring correlations, which are limited to linear (Pearson correlation) or monotonic (Spearman correlation) dependencies, we will use the mutual information, I. It measures a dissimilarity between the joint distribution of a pair of variables and the product of the marginals. It is therefore a measure of the distance to general statistical independence:

$$I(X;Y) = D_{KL}(P_{(X,Y)} \| P_X \otimes P_Y),$$

where D_{KL} is the Kullback-Leibler divergence. The mutual information can also be written in terms of entropies (H) as follows:

$$I(X;Y) = H(X,Y) - H(X|Y) - H(Y|X) = H(Y) - H(Y|X) = H(X) - H(X|Y),$$

and it is therefore a measure of the amount of uncertainty about one variable once the other is known. The mutual information is positive, but is not upper bounded. Therefore, we use a normalized version using the entropies of each variable:

$$I_{norm}(X;Y) = \frac{I(X;Y)}{\sqrt{H(X)H(Y)}}$$

We study this measure for both the L1C and the L2A data. Figure 5 displays the values of the normalized mutual information correlation for all the pairs of bands of L1C (left) and L2A (right) data.

Both levels of processing show the same patterns and nearly the same values, although L2A has slightly lower values of dependence. This may indicate that the atmospherical corrections are able to remove effects with high correlation across bands.

We observe high values for the red edge bands, between B5 and the red band, and between the 2 SWIR bands. Interestingly, B5 presents a relatively low dependence with respect to B6 and B7 and there is very small redundancy between B8 and B8A (it is for instance lower than between green and B5).

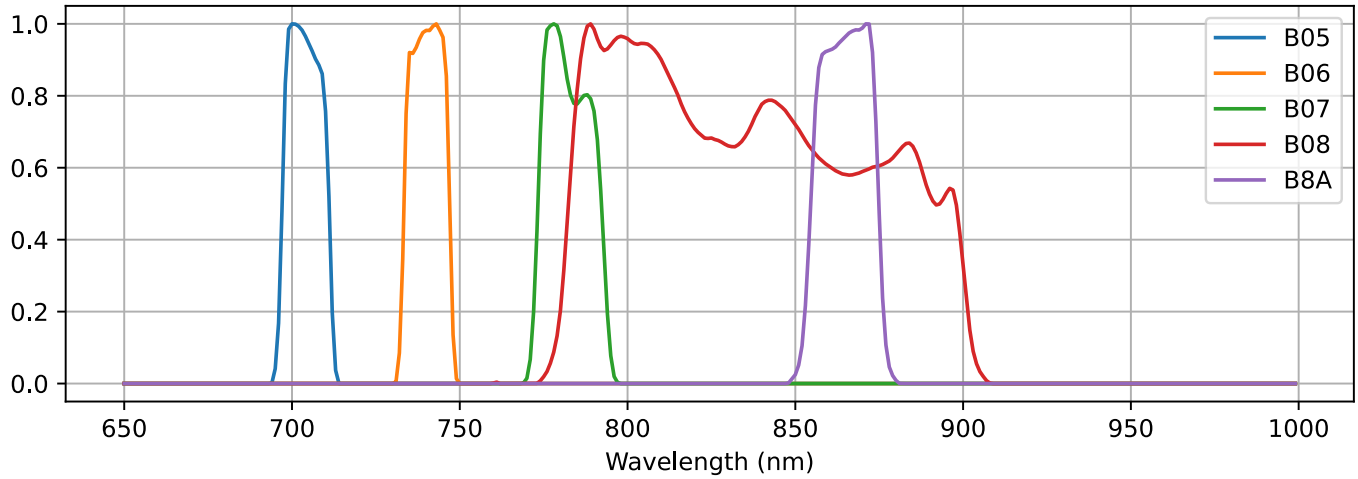


Figure 6. Red Edge and NIR bands

The highest values of mutual information are obtained between adjacent bands of the B6, B7, B8A triplet, B7 being the most similar to the others. B7 seems therefore a good candidate for reconstruction from other bands.

It is also interesting to note that B5 has all values higher than 0.4 (except for B8), which may indicate, either a possibility of reconstruction from the other bands, or conversely, being some sort of pivotal band to reconstruct the others.

The relatively low value of the mutual information between B8 and B8A may seem surprising since the latter is a subset of the former. Actually, this value is the same for B7 and B8, which are adjacent (see figure 6). However, B8A has a width of less than 20% of that of B8.

One limitation of this analysis is that only pairs of bands are compared, and therefore, it is impossible to assess if the redundancies between, for instance, B7 and B6 are complementary to those between B7 and B8A, which would allow a better reconstruction of B7 from the other 2 than if these redundancies were the same.

This means that these measures of mutual information are lower bounds of the amount of information that could be reconstructed from other bands.

4.2. Single band regression

We present in this section the performances of the reconstruction of each spectral band by applying the neural network regression algorithm described in section 3. As stated before, only the 20 m bands are reconstructed and the following data is used as predictors:

- the sinus and cosinus of the 4 observation angles
- all the 20 m bands except the target one
- the values of the 4 10 m pixels for B2, B3, B4 and B8 associated to the 20 m target pixel
- and only for L1C, the value of the 3 60 m bands interpolated (with a bicubic interpolator) to the coordinate of the center of the 20m pixel.

Each regression case is repeated 10 times using the protocol described at the end of section 2.

4.2.1. Analysis of errors

Validation metrics are computed across all experiments and reported on tables 3 and 4 for L1C and L2A data respectively. The tables present the root mean square error (RMSE), the mean absolute error (MAE), the relative error (RE) and the coefficient of determination (R^2). The rows of the tables are sorted by increasing values of RE for L1C and RMSE for L2A.

In section 1.2 we concluded that 3% error for L1C and 0.01 in surface reflectance values for L2A were good targets for band reconstruction. Of course, we are measuring reconstruction errors using data which itself may have errors, even if they are below the radiometric specifica-

Table 3: Single band regression results for L1C

Band	RMSE	MAE	RE	R^2
B07	7.17e-03	3.90e-03	2.96e-02	9.96e-01
B06	1.82e-02	4.77e-03	3.61e-02	9.88e-01
B8A	1.57e-02	5.33e-03	3.69e-02	9.91e-01
B05	1.57e-02	4.46e-03	3.79e-02	9.92e-01
B12	1.50e-02	9.18e-03	9.35e-02	9.83e-01
B11	1.83e-02	1.26e-02	1.51e-01	9.85e-01

Table 4: Single band regression results for L2A

Band	RMSE	MAE	RE	R^2
B5	7.33e-03	4.96e-03	2.07e-01	9.95e-01
B6	8.26e-03	5.04e-03	1.28e-01	9.96e-01
B7	8.42e-03	5.02e-03	1.18e-01	9.97e-01
B8A	1.11e-02	6.14e-03	2.23e-01	9.95e-01
B12	1.49e-02	9.49e-03	2.41e-01	9.75e-01
B11	2.06e-02	1.36e-02	4.05e-01	9.81e-01

tions. Therefore, the error bounds need not to be taken very strictly. Finally, Sentinel-2 can be considered to be over-specified in terms of radiometric quality for most applications, which makes using these error bands rather conservative from our point of view.

We see that, for L1C, only the reconstruction of B7 has a RE lower than 3%, although the other red edge and NIR bands are below 3.8%. For L2A, B5, B6 and B7 have an RMSE lower than 0.01, and B8A is only slightly above this level.

Estimating the noise in surface reflectances using the RMSE can suffer from strong outliers. The MAE gives a measure which is robust to these cases and shows that even B12 could be considered for reconstruction.

The error values presented on tables 3 and 4 are averages over the validation samples and don't show the proportion of pixels that do not fulfill the radiometric requirements. For this purpose, tables 5 and 6 show the percentage of pixels whose error is lower than a given threshold.

Table 5: Percentage of pixels beyond a given absolute error threshold (L2A)

Band	0.01	0.015	0.02	0.025
B5	13.09	4.76	1.94	0.91
B6	12.60	5.35	2.67	1.51
B7	12.47	5.31	2.60	1.39
B8A	19.19	8.73	4.40	2.47
B11	46.75	34.45	25.03	18.11
B12	33.81	22.00	14.96	10.41

Table 5 presents, for each L2A band, the percentage of pixels whose error is larger than a given threshold (from 0.01, which is the accuracy of the L2A processor, up to 0.025). We see, that even for the best predicted bands (in the Red Edge), less than 90% of the pixels fulfill the requirements. However, lowering the requirement accuracy to 0.015, a 95% compliance is achieved for these 3 bands.

Table 6 shows the same results for L1C data. The performaces seem to be much better than for L2A, but we must remember that the requirements for L1C are given as relative errors (the error must not exceed 3%).

Table 7 shows the percentage of validation pixels compliant with different error thresholds. We see that the requirement has to be lowered from 3% to 10% in order to get 95% compliance for the Red Edge and NIR bands. These bad performances are mainly due to high relative errors in the low reflectances. Tables 8 through 12 show the compliance with relative error

Table 6: Percentage of pixels beyond a given absolute error threshold (L1C)

Band	0.01	0.015	0.02	0.025
B05	6.86	2.37	1.21	0.78
B06	8.09	3.07	1.42	0.79
B07	8.95	3.15	1.19	0.51
B8A	13.49	4.83	1.87	0.82
B11	43.70	30.51	20.97	14.35
B12	30.20	20.38	14.14	9.75

Table 7: Percentage of pixels beyond a given relative error threshold (L1C)

Band	0.03	0.05	0.1
B05	37.23	20.44	5.34
B06	26.50	10.74	2.36
B07	21.99	8.97	2.10
B8A	30.11	13.16	3.66
B11	69.33	51.53	22.93
B12	73.79	58.25	30.10

thresholds for different intervals of reflectances. The results confirm that reflectances lower than 0.1 contain most of the errors.

Table 8: Percentage of pixels beyond a given relative error threshold for reflectances in [0,0.1] (L1C)

Band	0.03	0.05	0.1
B05	49.66	29.34	7.85
B06	39.48	16.89	2.25
B07	45.16	23.58	4.94
B8A	48.46	27.90	7.02
B11	75.16	60.25	32.95
B12	74.92	60.26	33.89

Figures 7 and 8 display scatterplots of predicted versus real reflectance values for the L1C and L2A bands respectively. For clarity in the visualization, these scatterplots are generated with a small random sample of the validation data. They show nevertheless the general behaviour and are coherent with the metrics presented in the tables above.

To complete the analysis of the errors, we present the histograms of the errors (true reflectance minus the predicted one) using the complete validation data set (about 5 million pixels). Figure 9 shows the histograms for the L1C bands and figure 10 for the L2A bands.

Table 9: Percentage of pixels beyond a given relative error threshold for reflectances in [0.1,0.25] (L1C)

Band	0.03	0.05	0.1
B05	35.01	17.53	3.58
B06	25.18	8.93	1.04
B07	21.63	8.11	1.05
B8A	30.70	12.17	1.69
B11	69.40	51.48	20.39
B12	73.14	57.22	27.93

Table 10: Percentage of pixels beyond a given relative error threshold for reflectances in [0.25,0.5] (L1C)

Band	0.03	0.05	0.1
B05	7.98	2.69	0.37
B06	16.71	5.98	0.42
B07	14.57	4.03	0.26
B8A	19.82	4.63	0.24
B11	65.31	45.43	15.98
B12	70.94	52.66	19.58

Table 11: Percentage of pixels beyond a given relative error threshold for reflectances in [0.5,0.75] (L1C)

Band	0.03	0.05	0.1
B05	10.06	3.15	0.29
B06	10.20	2.93	0.22
B07	17.32	4.88	1.13
B8A	20.40	5.26	0.25
B11	53.24	27.82	3.57
B12	52.10	31.80	5.49

Table 12: Percentage of pixels beyond a given relative error threshold for reflectances in [0.75,1] (L1C)

Band	0.03	0.05	0.1
B05	13.94	8.35	1.36
B06	5.30	1.17	0.07
B07	3.87	1.30	0.73
B8A	6.61	0.86	0.02
B11	85.90	75.64	51.28
B12	92.86	90.00	81.43

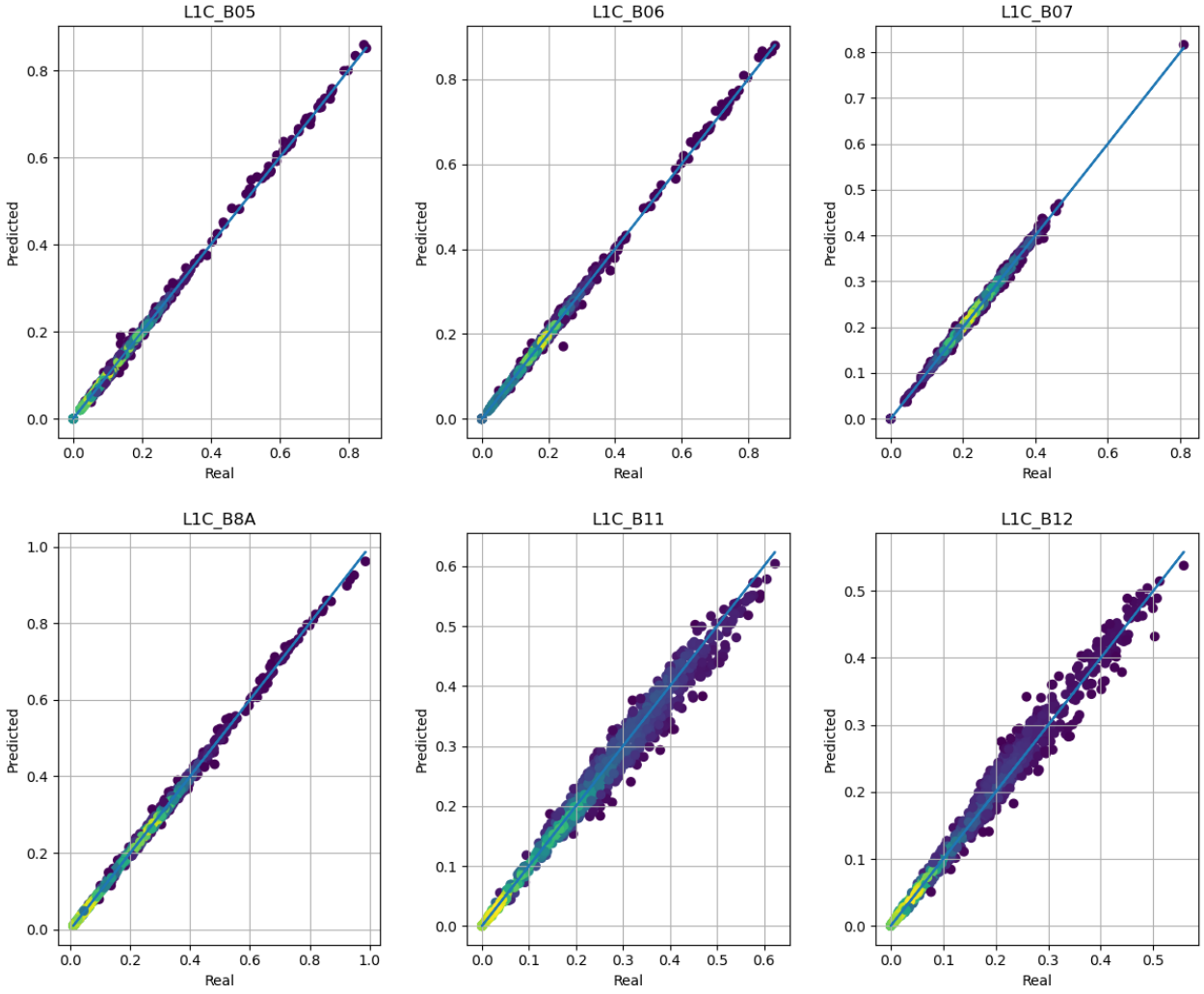


Figure 7. Scatterplots for the single band regression (L1C)

4.2.2. Analysis of the uncertainty estimation

As explained in section 3.1, the regression model is also able to estimate the uncertainty of the predicted value by associating a variance to it. Since this variance is an estimation itself, its meaningfulness needs to be assessed.

The loss function used to train the model was chosen assuming a Gaussian error model. The histograms in figures 9 and 10 show that the distributions of the errors are not Gaussian. However, these distributions are mono-modal, which may allow using the estimated variance as a good proxy for the uncertainty of the estimation. In order to check this hypothesis, we will measure the proportion of pixels having errors higher than a given proportion of the variance.

In the case of a Gaussian distribution, we have that $P(\mu-1\sigma \leq X \leq \mu+1\sigma) \approx 68.27\%$, $P(\mu-2\sigma \leq X \leq \mu+2\sigma) \approx 95.45\%$ and $P(\mu-3\sigma \leq X \leq \mu+3\sigma) \approx 99.73\%$.

We can therefore compute the proportion of pixels having an absolute error lower than σ, 2σ and 3σ and compare to the probability values above.

Tables 13 and 14 present the above-mentioned proportions of pixels whose errors are within the bounds given by the estimated sigma. We see that, although not identical, the proportions are relatively similar to what one should get in the Gaussian case.

It is important to understand that the value of σ is provided by the regression algorithm as a prediction. These results show that this prediction of σ is indeed a good proxy for the

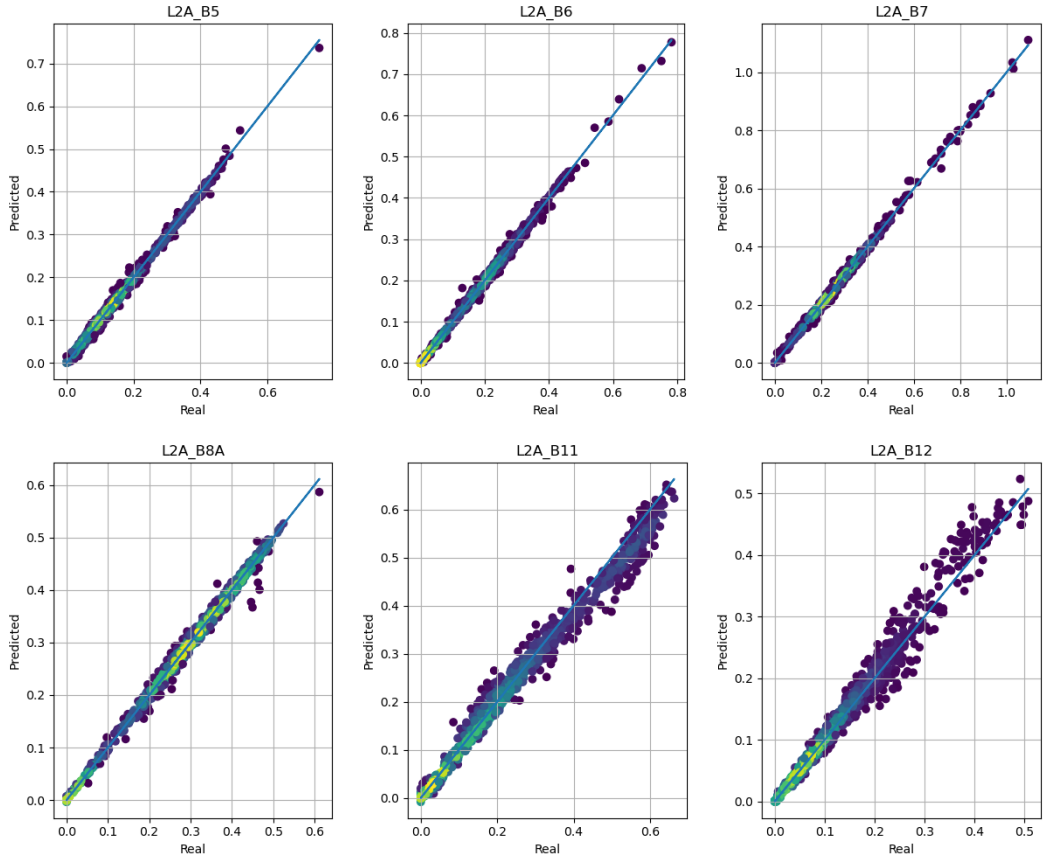


Figure 8. Scatterplots for the single band regression (L2A)

Table 13: Probability of the absolute error being lower than $n \times \sigma$ (L1C)

Band	σ (68.27%)	2σ (95.45%)	3σ (99.73%)
B05	68.80	92.26	98.29
B06	70.23	93.14	98.53
B07	70.82	93.27	98.43
B8A	70.36	91.99	97.57
B11	56.86	84.11	94.94
B12	61.57	87.50	96.23

Table 14: Probability of the absolute error being lower than $n \times \sigma$ (L2A)

Band	σ (68.27%)	2σ (95.45%)	3σ (99.73%)
B5	65.68	90.00	96.87
B6	71.84	93.85	98.70
B7	74.28	94.50	98.69
B8A	70.07	92.13	97.93
B11	56.18	81.51	93.04
B12	64.70	91.15	97.91

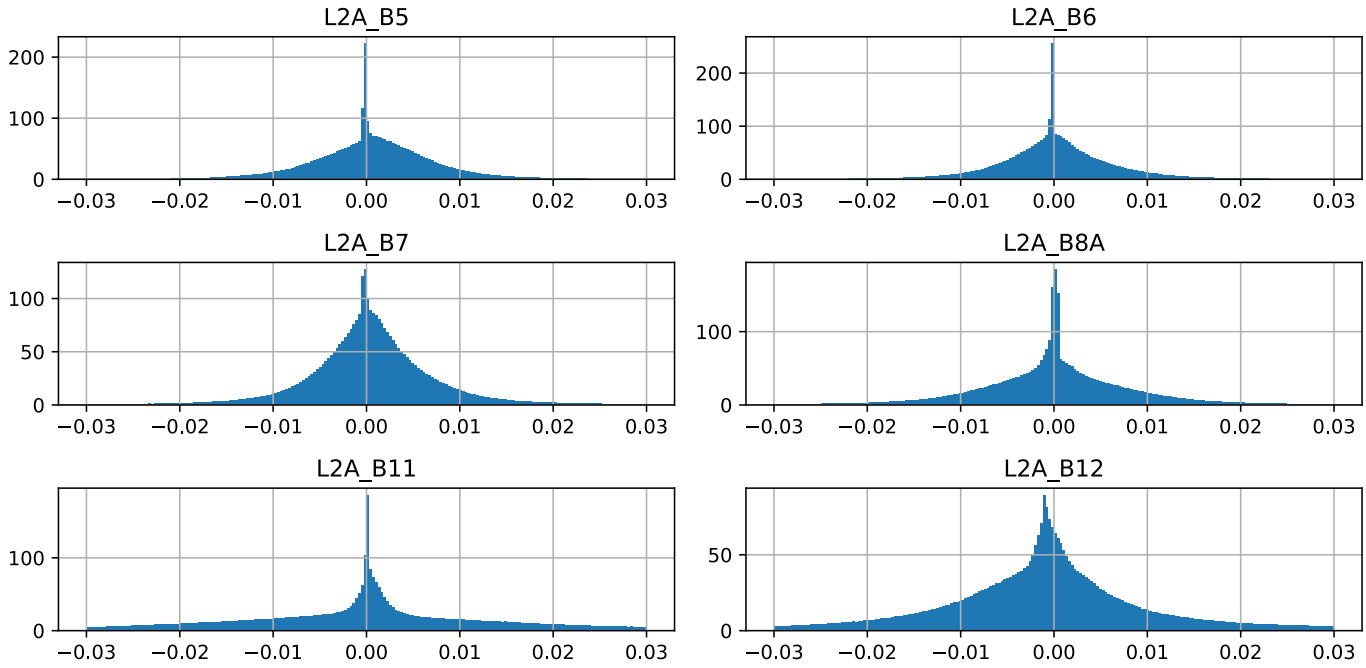


Figure 10. Histograms of the errors (true value minus prediction) for the L2A bands.

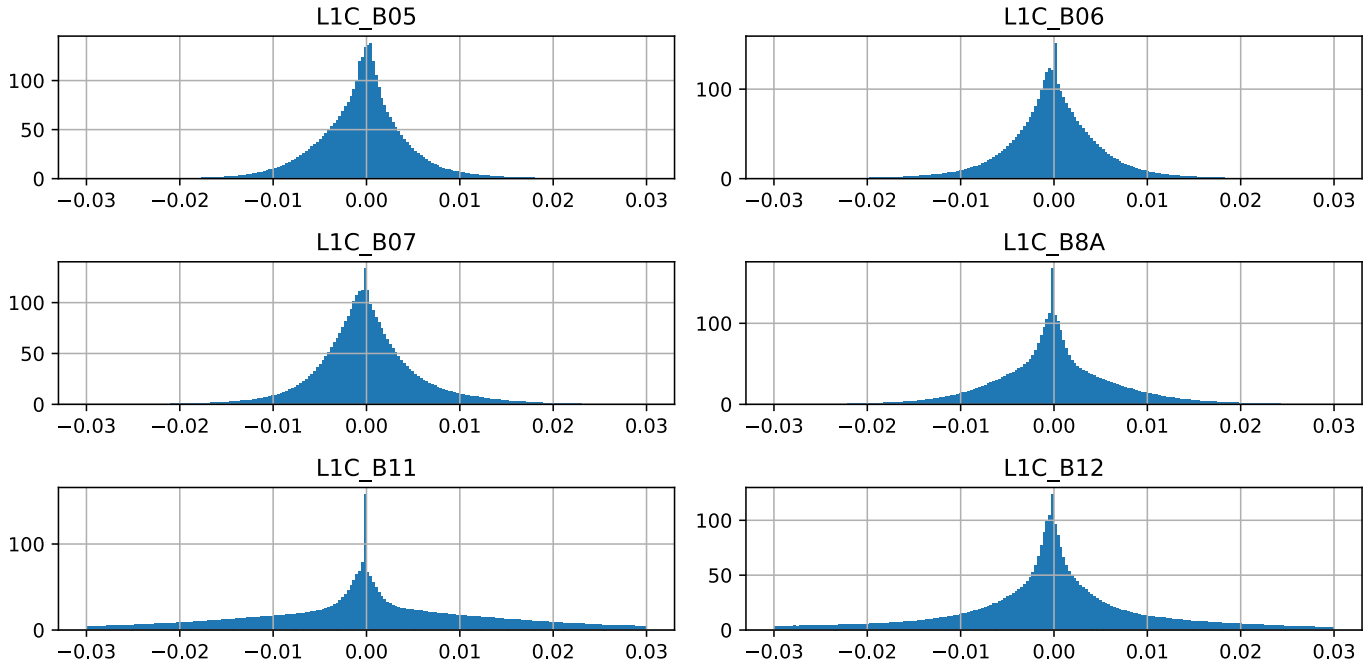


Figure 9. Histograms of the errors (true value minus prediction) for the L1C bands.

probability of the reflectance estimation of being in the predicted interval. Therefore, the estimation of σ can be thresholded and used as a validity mask for the estimations.

4.3. Double band regression

We present here the results for the case were 2 bands are predicted from the others. This case will of course produce higher estimation errors because for each predicted band there is one fewer predictor.

Table 1: Double band regression results for L1C

Band	RMSE	MAE	RE	R^2	Band	RMSE	MAE	RE	R^2
B06	9.28e-03	4.79e-03	2.67e-02	9.93e-01	B07	1.07e-02	4.93e-03	2.64e-02	9.93e-01
B07	1.68e-02	5.49e-03	3.47e-02	9.89e-01	B8A	1.95e-02	7.84e-03	4.34e-02	9.84e-01
B05	9.31e-03	3.96e-03	5.12e-02	9.94e-01	B06	1.01e-02	4.22e-03	4.07e-02	9.94e-01
B05	6.52e-03	3.58e-03	3.88e-02	9.94e-01	B11	1.66e-02	1.12e-02	7.57e-02	9.86e-01
B06	1.52e-02	4.85e-03	3.24e-02	9.92e-01	B11	1.75e-02	1.16e-02	9.04e-02	9.82e-01
B06	4.38e-02	1.58e-02	1.00e-01	9.47e-01	B8A	1.90e-02	5.88e-03	7.06e-02	9.92e-01
B05	8.20e-03	4.16e-03	3.64e-02	9.94e-01	B07	4.34e-02	1.66e-02	1.21e-01	8.41e-01
B05	3.79e-02	8.14e-03	9.38e-02	8.74e-01	B12	1.92e-02	1.07e-02	1.27e-01	9.66e-01
B05	9.64e-03	4.24e-03	3.93e-02	9.96e-01	B8A	4.63e-02	1.93e-02	1.31e-01	9.45e-01
B07	1.74e-02	4.45e-03	4.05e-02	9.88e-01	B11	5.92e-02	2.88e-02	1.97e-01	8.47e-01
B8A	2.66e-02	7.06e-03	5.44e-02	9.73e-01	B12	1.64e-02	1.00e-02	2.26e-01	9.72e-01
B07	1.93e-02	5.07e-03	4.74e-02	9.87e-01	B12	5.17e-02	2.00e-02	2.37e-01	7.91e-01
B06	7.30e-03	3.88e-03	3.07e-02	9.95e-01	B12	3.63e-02	1.78e-02	2.52e-01	8.59e-01
B8A	5.70e-02	1.47e-02	1.25e-01	8.87e-01	B11	6.20e-02	2.80e-02	2.79e-01	7.51e-01
B11	3.73e-02	2.45e-02	2.93e-01	9.18e-01	B12	3.26e-02	2.03e-02	2.24e-01	9.05e-01

Table 2: Double band regression results for L2A

Band	RMSE	MAE	RE	R^2	Band	RMSE	MAE	RE	R^2
B5	7.38e-03	4.96e-03	1.77e-01	9.95e-01	B8A	9.31e-03	6.22e-03	1.38e-01	9.96e-01
B5	1.11e-02	6.07e-03	1.80e-01	9.95e-01	B6	1.33e-02	6.37e-03	1.25e-01	9.94e-01
B6	8.68e-03	5.07e-03	1.91e-01	9.96e-01	B12	1.53e-02	9.59e-03	3.52e-01	9.77e-01
B6	1.24e-02	6.70e-03	5.37e-01	9.89e-01	B7	1.54e-02	7.70e-03	1.51e-01	9.85e-01
B5	1.26e-02	5.35e-03	2.03e-01	9.94e-01	B11	1.72e-02	1.15e-02	1.94e-01	9.85e-01
B7	7.30e-03	4.72e-03	1.23e-01	9.96e-01	B12	1.74e-02	1.11e-02	1.97e-01	9.78e-01
B7	1.13e-02	5.98e-03	1.10e-01	9.94e-01	B8A	1.74e-02	7.70e-03	1.65e-01	9.87e-01
B8A	1.21e-02	6.10e-03	1.91e-01	9.95e-01	B12	1.79e-02	1.08e-02	3.03e-01	9.75e-01
B8A	1.07e-02	6.74e-03	1.09e-01	9.94e-01	B11	1.93e-02	1.31e-02	2.22e-01	9.79e-01
B6	9.79e-03	5.60e-03	1.47e-01	9.95e-01	B8A	1.99e-02	8.58e-03	2.10e-01	9.81e-01
B11	4.14e-02	2.89e-02	3.29e-01	9.01e-01	B12	3.54e-02	2.36e-02	4.86e-01	8.90e-01
B7	3.74e-02	1.41e-02	5.18e-01	9.21e-01	B11	4.98e-02	2.50e-02	2.37e-01	8.56e-01
B5	5.80e-02	1.17e-02	2.11e-01	8.73e-01	B12	1.71e-02	1.09e-02	4.36e-01	9.76e-01
B6	6.55e-02	1.44e-02	7.58e-01	8.01e-01	B11	4.06e-02	1.88e-02	3.21e-01	8.75e-01
B5	7.70e-02	4.45e-02	2.87e-01	8.35e-01	B7	9.79e-03	5.83e-03	9.24e-02	9.96e-01

Tables 1 and 2 present the results for the L1C and the L2A data. Each row in the tables presents the results for a pair of bands jointly predicted. The same quality metrics as for the single band regression are presented. Each table has 15 rows, since we evaluate all the possible combinations of pairs of bands.

The rows in table 1 are sorted in increasing order of the maximum relative error of the pair of bands. This allows to quickly see that only one pair of L1C bands (B06 and B07) can be predicted with less than 3% error and that another pair (B07 and B8A) is slightly above this threshold.

For the L2A data presented in table 2, the rows are sorted by increasing RMSE using the maximum of the pair in each row. In this case, only the pair (B5, B8A) fulfills the 0.01 error threshold, although the pair (B5, B6) is not much above this threshold.

Figure 11 presents the scatterplots for the 2 best pairs of bands in the L1C case (the 2 first rows in table 1. Altough the scatterplots are generated by subsampling the test data set for readability, one can see that the estimations are unbiased and with a small dispersion around the regression lines. One can also see that part of the error comes from pixels with reflectances higher than 1, for which there is an underestimation. Since the regression algorithm is configured to yield reflectances in the $[-0.2, 1.3]$ interval, we can expect that the error in this interval is smaller than what is reported in the tables.

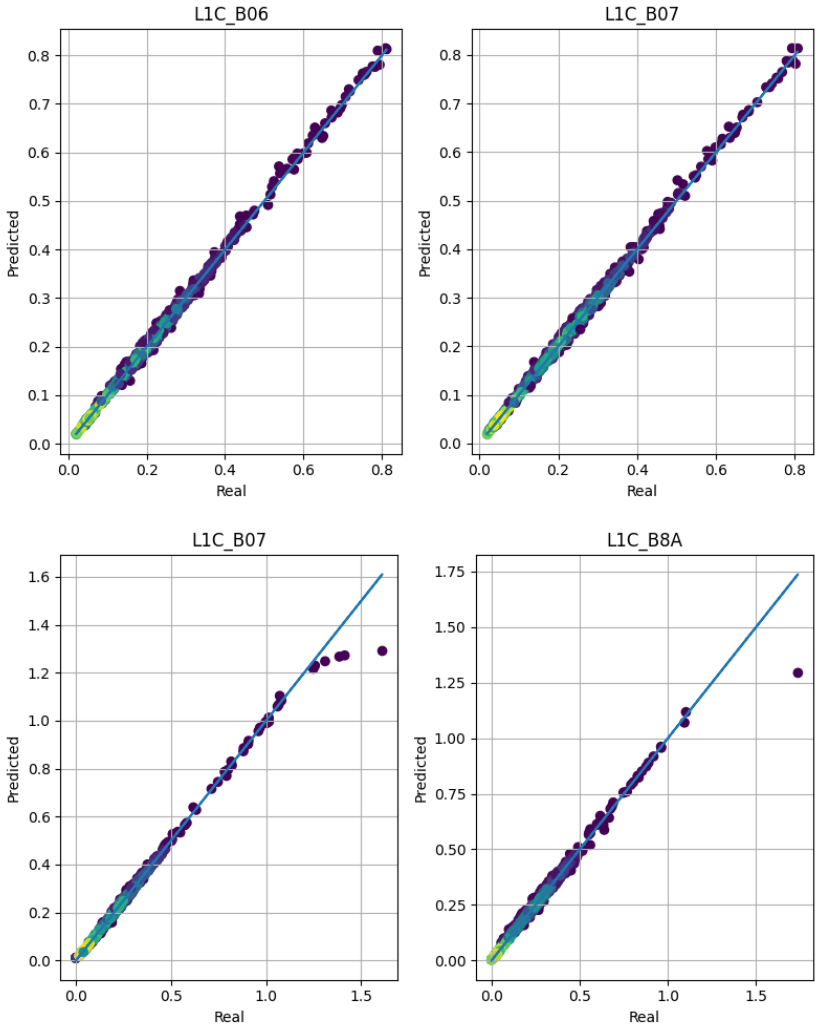


Figure 11. Scatterplots for the double band regression (L1C). Each row in the figure corresponds to a row in table 1

For L2A data, figure 12 presents the scatterplots for the pairs of bands in the 3 first rows of table 2. As for the L1C case, the scatterplots show unbiased estimations with small dispersions, except for the B12 band in the 3rd pair. The random sample of the test set used for generating these scatterplots does not contain pixels showing the underestimation of reflectances higher than 1, but they also exist.

It is difficult to give an explanation for these results. First of all, the pairs of bands that are predicted the *best* differ between L1C and L2A. This was already the case for the regression of a single band, but in that case, we could clearly define 2 groups, the Red-Edge-NIR and the SWIR. In the case of 2 bands, one could have expected that, for a pair of bands to be correctly reconstructed, they would have to be non-adjacent, so that the missing information could be reconstructed using the neighbouring bands. However, we see that the best pair in L1C is (B06, B07) and that the second best pair in L2C is (B5, B6).

With the same kind of reasoning, one could have expected that the pair (B11, B12) should be the one with the largest errors, since reconstructing the SWIR bands usin only the VIS-NIR range should be nearly impossible. This is the case in terms of relative error, but not in terms of RMSE, which makes the SWIR a better candidate for L2A reconstruction than more spectrally distant pairs.

Figure 13 presents the scatterplots for the prediction of the SWIR bands in L1C (top row) and L2A (bottom row). Although the dispersions are important, there is no systematic bias in the estimation, which confirms the redundancy of spectral information for most surfaces.

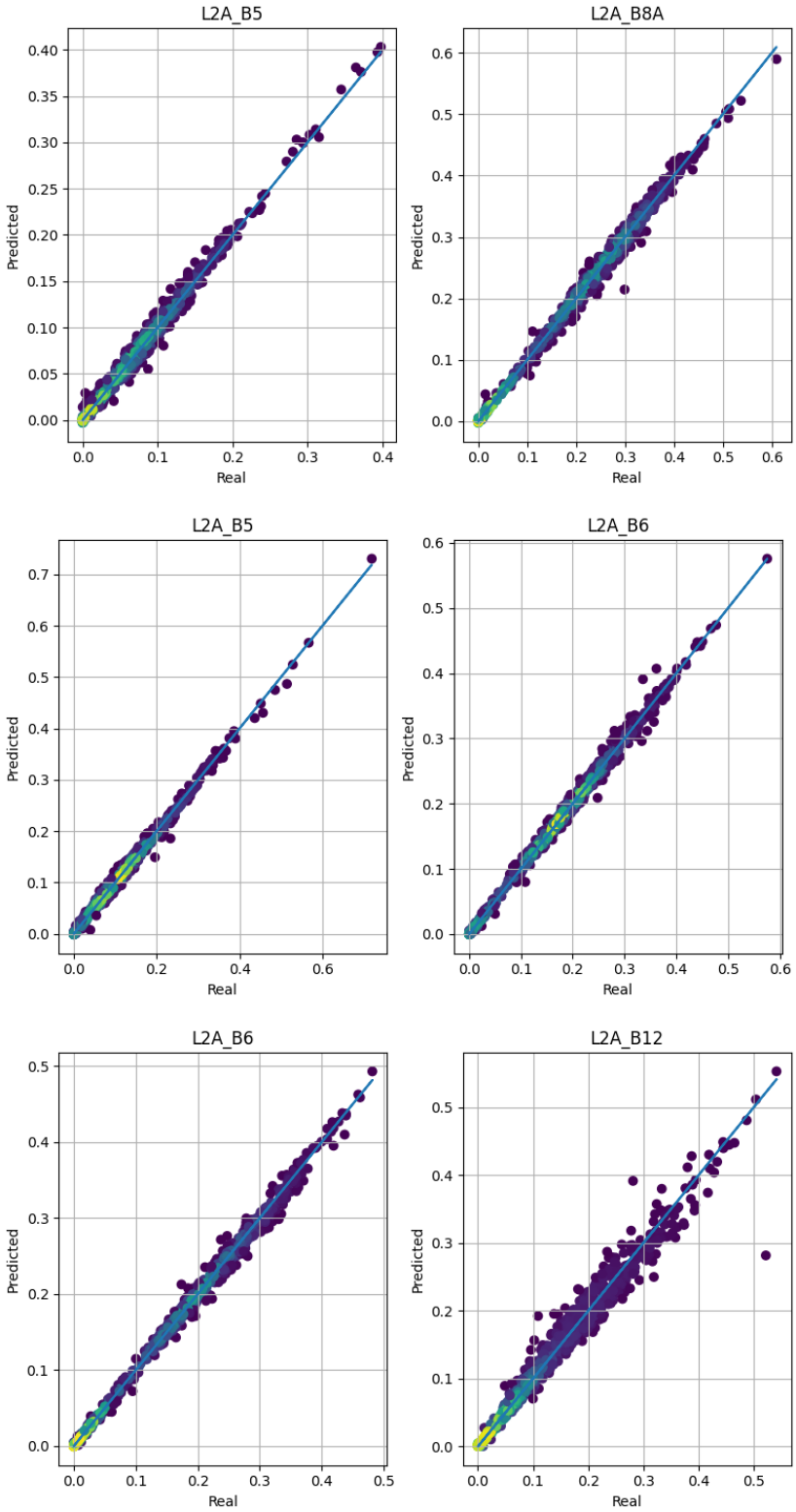


Figure 12. Scatterplots for the double band regression (L2A). Each row in the figure corresponds to a row in table 2

Actually, if the scatterplots of figure 13 were obtained for biophysical parameter estimations, as for instance LAI, chlorophyll, biomass, etc. they would be considered as very good (see for instance [27] or [18]). Of course, image quality criteria need to be more strict than those of downstream tasks, but this kind of result suggests that the impact of reflectance noise in downstream applications needs to be assessed.

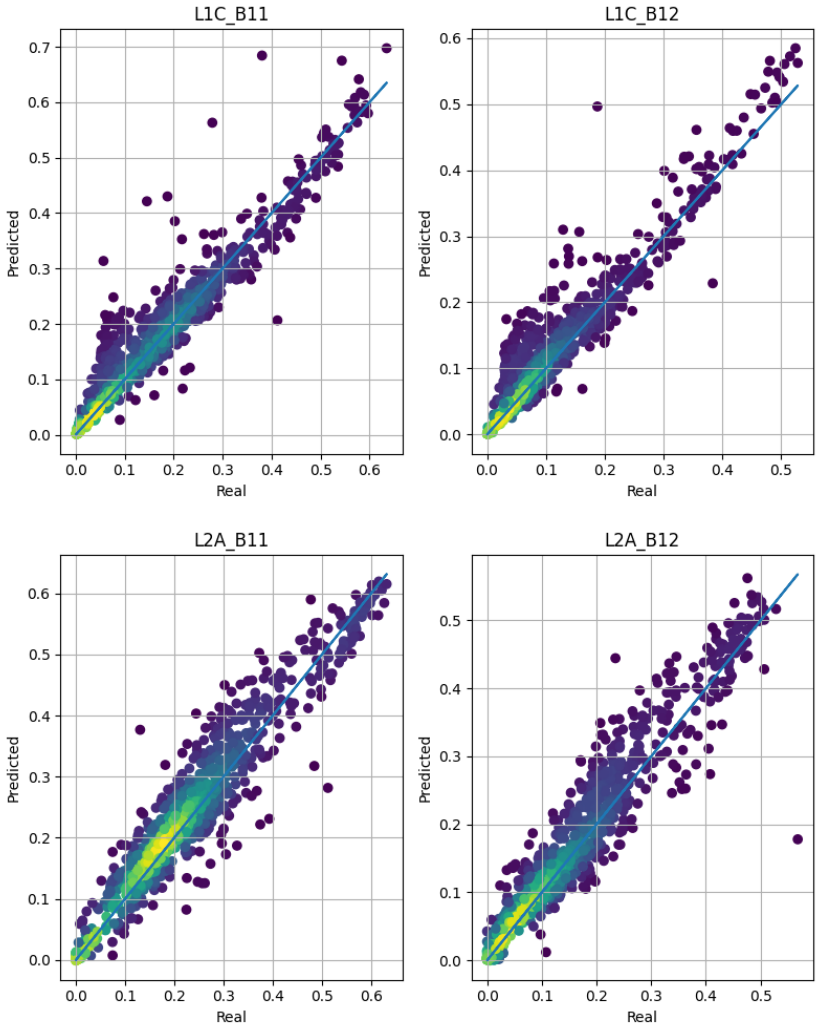


Figure 13. Scatterplots for the double band regression of the SWIR bands in L1C (top) and L2A (bottom).

5. Conclusions and further studies

In this paper, we have investigated the possibility of reconstructing one or two of the 20m resolution bands from Sentinel-2 using the remaining bands. The goal of the study was to assess the possibility of removing some of the current Sentinel-2 spectral bands for the next generations of similar satellites.

The interest of working on band reconstruction is that the approach is independent of the application. The main rationale is that, if a band can be reconstructed with errors which are within the radiometric requirements of the sensor, downstream applications can use a reconstructed band instead of a real measure.

Conclusions

The main findings of the study are that, at least, one of the bands among B5, B6, B7 and B8A could be removed from next generation sensors as all of them can be reconstructed with small errors when the others are available. Removing two bands could be possible at the cost of slightly higher reconstruction errors.

We have also shown that the estimation of a credibility interval for the predicted reflectances is possible and can therefore be used as a quality mask.

This study has however several limitations that would need to be addressed in the future.

First, an analysis of the errors per type of surface (material, land cover, vegetation status, etc.) should be carried out in order to assess the impact on different types of applications.

Although the spatial sampling of the data for this study contained enough variability for the results to be general, particular types of surfaces with specificities may need special attention. Furthermore, selecting the appropriate samples in the areas of most interest for particular applications can allow a fine tuning of the regression algorithm and improve the performances of the estimations.

A second limitation is related to the choice of regression algorithm for the study. The goal of the work was not to propose an optimal regression algorithm, but rather showing that band reconstruction was possible using regression. The choice of the neural network with a negative log-likelihood as a loss function was made for simplicity in terms of implementation, the possibility of performing multi-target regression and the generation of uncertainties associated to the estimations. Other approaches could yield better results and even produce a different set of bands candidate for removal.

All of the above suggests that replication of the study by other teams would be useful. For this purpose, the dataset has been published [22] and the source code is available for inspection and download[2].

A third limitation is the pixel-based approach taken here. Reconstructing a missing band from the reflectances of the other bands of the same pixel assumes unicity of the solution: one combination of observed bands can only correspond to one value of the missing band. Although the results of this study tend to show that this is the case, there are pixels for which the error is high. In the current setting, the regression algorithm is able to flag these pixels by reporting a high uncertainty, but this is not fully satisfactory. One way of lifting the ambiguity would be to add some spatial context for the regression, so that the observations of neighbouring pixels, and therefore the local texture, helps the prediction. This could be implemented with spatial convolutional layers in the regression algorithm.

In the same way, a multitemporal extension of the algorithm could improve the estimations. However, this extension is less straightforward than adding spatial context, since clouds and cloud shadows introduce an irregular temporal sampling that should be taken into account. Also, the relative geometric accuracy of multi-temporal series should be taken into account in this case.

Perspectives

If the next generation of Sentinel-2 had one or several bands removed, one could argue that the approach presented in this work couldn't be applied, since the regression calibration (i.e. the neural network training) needs the target band. Several responses can be given to this argument. If the bands used as predictors remain the same in the new sensor, the regressions calibrated with the current Sentinel-2 data should be applicable.

If the bands used as predictors for this study were not available in the next generation of satellites, one could constitute enough training data by using acquisitions from a hyperspectral mission like CHIME [28]. The appropriate spectral bands (predictors and targets) could be generated using the relative spectral responses of the next generation of Sentinel-2.

Finally, given the temporal revisit of Sentinel-2, it would be interesting to evaluate the possibility of having different bands in different satellites of the constellation, so that the band predictions could be temporally interleaved. For instance, with 2 satellites, one could imagine removing B6 in the A unit and removing B7 in the B unit. In this configuration, the reconstruction of B6 at a given date could use the other bands for this acquisition, as well as the most recent acquisiton of the other satellite for which B6 would have been observed. This kind of scenario would allow for interesting configurations where one satellite of the pair could have the SWIR bands absent. Indeed, the variance observed in figure 13 could be highly reduced if the 2 SWIR bands for a previous date were available. Of course, for surfaces where the SWIR signature can change quickly during cloudy periods (snow falls), the impact of this kind of settings should be studied. Fortunately, the available Sentinel-2 data in the archives allows to do that.

[2] http://gitlab.cesbio.omp.eu/Jordi/mmdc/blob/master/mmdc/spectral_regression.py

Other interesting possibility of the approaches presented in this paper is the *addition* of new bands, but only in some satellites of the constellation. On this topic, we should stress the comments on [29] we did in section 1.4: the fact that a particular phenomenon has a signature in a particular band, does not mean that this same phenomenon can not be detected by using a (non-linear) combination of other bands. The results presented in this paper indicate that the question can be reversed.

The attentive reader will have understood that many options are open to reduce costs and hardware complexity for the successors of the current Sentinel-2 system by leveraging spectral, spatial and temporal correlations of the observed surfaces through ground data processing.

This work is just an example of what could be done by using the richness of the Sentinel-2 archives. We think that, with the help of other scientists, further studies could be defined. For instance, a subset of geographic areas and dates for each target application, together with ground measures could be made available. This would allow the objective assessment of errors due to the lack of particular bands.



1. Drusch, M.; Bello, U.D.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; Meygret, A.; Spoto, F.; Sy, O.; Marchese, F.; Bargellini, P. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **2012**, *120*, 25–36. doi:10.1016/j.rse.2011.11.026.
2. Team, E.S.. GMES Sentinel-2 Mission Requirements Document. Technical report, European Space Agency, 2007.
3. Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A Multi-Temporal Method for Cloud Detection, Applied To Formosat-2, VENμS, Landsat and Sentinel-2 Images. *Remote Sensing of Environment* **2010**, *114*, 1747–1755.
4. Hagolle, O.; Huc, M.; Villa Pascual, D.; Dedieu, G. A Multi-Temporal and Multi-Spectral Method To Estimate Aerosol Optical Thickness Over Land, for the Atmospheric Correction of Formosat-2, Landsat, VENμS and Sentinel-2 Images. *Remote Sensing* **2015**, *7*, 2668–2691.
5. Gorroño, J.; Fomferra, N.; Peters, M.; Gascon, F.; Underwood, C.; Fox, N.; Kirches, G.; Brockmann, C. A Radiometric Uncertainty Tool for the Sentinel 2 Mission. *Remote Sensing* **2017**, *9*, 178. doi:10.3390/rs9020178.
6. Roujean, J.L.; Leroy, M.; Deschamps, P.Y. A Bidirectional Reflectance Model of the Earth's Surface for the Correction of Remote Sensing Data. *Journal of Geophysical Research: Atmospheres* **1992**, *97*, 20455–20468.
7. Fernández-Manso, A.; Fernández-Manso, O.; Quintano, C. Sentinel-2a Red-Edge Spectral Indices Suitability for Discriminating Burn Severity. *International Journal of Applied Earth Observation and Geoinformation* **2016**, *50*, 170–175. doi:10.1016/j.jag.2016.03.005.
8. Pasqualotto, N.; Delegido, J.; Van Wittenberghe, S.; Rinaldi, M.; Moreno, J. Multi-Crop Green Lai Estimation With a New Simple Sentinel-2 Lai Index (SeLI). *Sensors* **2019**, *19*, 904. doi:10.3390/s19040904.
9. Tian, J.; Su, S.; Tian, Q.; Zhan, W.; Xi, Y.; Wang, N. A Novel Spectral Index for Estimating Fractional Cover of Non-Photosynthetic Vegetation Using Near-Infrared Bands of Sentinel Satellite. *International Journal of Applied Earth Observation and Geoinformation* **2021**, *101*, 102361. doi:10.1016/j.jag.2021.102361.
10. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-Destructive Optical Detection of Pigment Changes During Leaf Senescence and Fruit Ripening. *Physiologia plantarum* **1999**, *106*, 135–141.
11. Jacques, D.C.; Kergoat, L.; Hiernaux, P.; Mougin, E.; Defourny, P. Monitoring Dry Vegetation Masses in Semi-Arid Areas With Modis Swir Bands. *Remote Sensing of Environment* **2014**, *153*, 40–49. doi:10.1016/j.rse.2014.07.027.
12. McFeeters, S. The Use of the Normalized Difference Water Index (ndwi) in the Delineation of Open Water Features. *International journal of remote sensing* **1996**, *17*, 1425–1432.
13. Gao, B.C. NDWI. A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment* **1996**, *58*, 257–266.
14. Misra, G.; Cawkwell, F.; Wingler, A. Status of Phenological Research Using Sentinel-2 Data: a Review. *Remote Sensing* **2020**, *12*, 2760. doi:10.3390/rs12172760.
15. Sitokonstantinou, V.; Papoutsis, I.; Kontoes, C.; Arnal, A.; Andrés, A.P.; Zurbano, J.A. Scalable Parcel-Based Crop Identification Scheme Using Sentinel-2 Data Time-Series for the Monitoring of the Common Agricultural Policy. *Remote Sensing, 10*.
16. Cerasoli, S.; Campagnolo, M.; Faria, J.; Nogueira, C.; Caldeira, M.d.C. On Estimating the Gross Primary Productivity of Mediterranean Grasslands Under Different Fertilization Regimes Using Vegetation Indices and Hyperspectral Reflectance. *Biogeosciences* **2018**, *15*, 5455–5471. doi:10.5194/bg-15-5455-2018.
17. Gómez-Giráldez, P.J.; Pérez-Palazón, M.J.; Polo, M.J.; González-Dugo, M.P. Monitoring Grass Phenology and Hydrological Dynamics of an Oak-Grass Savanna Ecosystem Using Sentinel-2 and Terrestrial Photography. *Remote Sensing* **2020**, *12*, 600. doi:10.3390/rs12040600.
18. Clevers, J.; Kooistra, L.; van den Brande, M. Using Sentinel-2 Data for Retrieving Lai and Leaf and Canopy Chlorophyll Content of a Potato Crop. *Remote Sensing* **2017**, *9*, 405. doi:10.3390/rs9050405.
19. Clevers, J.; Gitelson, A. Remote Estimation of Crop and Grass Chlorophyll and Nitrogen Content Using Red-Edge Bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation* **2013**, *23*, 344–351. doi:10.1016/j.jag.2012.10.008.
20. Hively, W.D.; Lamb, B.T.; Daughtry, C.S.; Serbin, G.; Dennison, P.; Kokaly, R.F.; Wu, Z.; Masek, J.G. Evaluation of Swir Crop Residue Bands for the Landsat Next Mission. *Remote Sensing* **2021**, *13*, 3718.

21. Transon, J.; d' Andrimont, R.; Maugnard, A.; Defourny, P. Survey of Hyperspectral Earth Observation Applications From Space in the Sentinel-2 Context. *Remote Sensing* **2018**, *10*, 157. doi:10.3390/rs10020157.
22. Inglada, J. Sentinel-2 L1C and L2A pixel samples for band regression. doi:10.5281/zenodo.5535821, 2021. doi:10.5281/zenodo.5535821.
23. Nix, D.; Weigend, A. Estimating the Mean and Variance of the Target Probability Distribution. *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)* **1994**. doi:10.1109/icnn.1994.374138.
24. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. doi:10.1038/nature14539.
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems 30*; Guyon, I.; Luxburg, U.V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds.; Curran Associates, Inc., 2017; pp. 5998–6008.
26. Kingma, D.P.; Ba, J. Adam: a Method for Stochastic Optimization. *CoRR* **2014**, [arXiv:cs.LG/1412.6980].
27. Camacho, F.; Fuster, B.; Li, W.; Weiss, M.; Ganguly, S.; Lacaze, R.; Baret, F. Crop Specific Algorithms Trained Over Ground Measurements Provide the Best Performance for Gai and Fapar Estimates From Landsat-8 Observations. *Remote Sensing of Environment* **2021**, *260*, 112453. doi:https://doi.org/10.1016/j.rse.2021.112453.
28. Earth.; Division, M.S. Copernicus Hyperspectral Imaging Mission for the Environment - Mission Requirements Document. Technical report, European Space Agency, ESTEC, 2019.
29. Hively, W.D.; Lamb, B.T.; Dennison, P.; Serbin, G. Reflectance spectra of agricultural field conditions supporting remote sensing evaluation of non-photosynthetic vegetation cover, 2021. doi:10.5066/P9XK3867.