

LEARNING HARMONISED PLEIADES AND SENTINEL-2 SURFACE REFLECTANCES

J. Michel*, J. Inglada

CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS, Toulouse, France
{julien.michel,jordi.inglada}@cesbio.cnes.fr

Commission III, WG III/6

KEY WORDS: Surface reflectance, cross-calibration, machine learning, Pleiades, Sentinel-2

ABSTRACT:

In this paper, we investigate the use of machine-learning techniques in order to produce harmonised surface reflectances between Sentinel-2 and Pleiades images, and reduce the impact of the differences in sensors, view conditions, and atmospheric correction differences between them. We demonstrate that if a simple linear regression considering Sentinel-2 surface reflectances as the target domain can overcome this problem when both images are calibrated to Top of Canopy reflectances, the non-linearity brought by a simple Multi-Layer-Perceptron is already useful when Pleiades is corrected to Top of Atmosphere level and contributions of the atmosphere need to be learned. We also demonstrate that learning a Convolution Neural Network instead of a plain MLP can learn undesired spatial effects such as mis-registration or differences in spatial frequency content, that will affect the image quality of the corrected Pleiades product. We overcome this issue by proposing an adhoc input convolutional layer that will capture those effects and can later be unplugged during inference. Last, we also propose an architecture and loss function that is robust to unmasked clouds and produces a confidence prediction during inference.

1. INTRODUCTION

1.1 Problem statement

The joint use of surface reflectances from different sensors can be challenging because of differences in sensor characteristics, ground segment algorithms and their parameters, and exogenous data used for corrections. All those factors can affect the coherency between surface reflectances and therefore create unwanted artefacts in operations leveraging direct comparison or statistical learning. To tackle this issue, the standard physics-based method includes careful modelling of effects induced by sensor differences as well as the use of common algorithms and parameters for both sensors, as demonstrated in (Claverie et al., 2018) for the constitution of an harmonised Sentinel-2 and Landsat 8 dataset.

When it comes to jointly using Pleiades and Sentinel-2, differences in sensors characteristics include (i) a factor of 5 in spatial resolution, (ii) different viewing angles for each acquisition compared to fixed nadir view, which causes BRDF variations, and (iii) differences in spectral sensitivities. Additional differences exist in ground segment algorithms, especially when using L2A Sentinel-2 products, for which endogenous atmospheric correction parameters are estimated from dedicated spectral bands, unavailable on sensors like Pleiades. This results in potentially non-linear reflectance discrepancies, as shown in figure 1.

One could of course benefit from the parameters estimated during Sentinel-2 atmospheric corrections, but many of those parameters vary rapidly in time, while Pleiades and Sentinel-2 pairs can be several days apart in time. Nonetheless, some Pleiades data providers deliver images that have already been processed for atmospheric corrections, with undisclosed algorithms and parameters.

* Corresponding author

In our case, level 2A Sentinel-2 time series contain fairly accurate estimations of surface reflectances with cloud and snow screening (Lonjou et al., 2016). Instead of trying to get coherent reflectances from Pleiades using traditional, model-based corrections which will inevitably suffer from all the effects that can not be modelled, we propose to use a model estimated by means of machine learning.

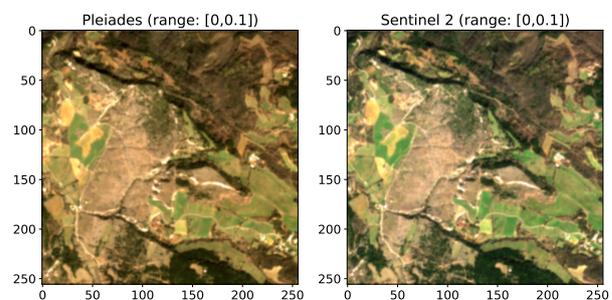


Figure 1. RGB composition of Top of Canopy surface reflectances from Pleiades and Sentinel-2, on the same acquisition date with same scaling.

1.2 Our contributions

In this work, we evaluate the ability of 3 different Neural Network architectures to learn the mapping of Top of Canopy or Top of Atmosphere reflectances from a Pleiades image to Top of Canopy reflectances from a Sentinel-2 L2A contemporary image. Most of the effects that we are trying to learn, for instance Pleiades viewing angle or the aerosol content of the atmosphere during Pleiades acquisition, are specific to each pair of acquisitions. We therefore seek for networks that train fast and apply only to those data, without looking for broader generalisation. Of course we still evaluate the performances on a set of patches that is not used during training to prevent over-fitting.

Site	PHR	Level	S2	training	testing
Aude 1	19.02.22	TOC	+0d	8856	2213
Aude 2	19.08.22	TOA	-6d	8204	2050
Bretagne 1	18.08.06	TOC	-1d	2280	570
Bretagne 2	18.05.06	TOC	+13d	2280	570

Table 1. Characteristics of the datasets used for the experiments. Data downloaded from Dinamis portal (PHR, Aude), Kalideos portal (PHR, Bretagne) and Theia portal (Sentinel-2, all sites). All Pleiades images displayed in this paper are ©CNES (2018, 2019), Distribution AIRBUS DS, all rights reserved, commercial use forbidden.

In order to evaluate the benefits of leveraging the Neural Network learning mechanisms, we quantitatively evaluate the gain in reflectance accuracy with respect to simple linear regression, a simple method used in similar problems (Luo et al., 2020), with four standard metrics. We also propose an additional prediction error output that is driven by a dedicated loss and can act as a simple cloud masking algorithm during inference.

2. EXPERIMENTAL SETUP

2.1 Datasets

For all the experiments, we use 4 pairs of Pleiades and Sentinel 2 L2A images, over two different sites in France. For each pair, Pleiades images were first projected in the Sentinel-2 coordinate reference system using phased re-sampling grids, on the geographic overlap of the two images. We use the four spectral bands of Pleiades and their closest match in Sentinel-2 images (i.e. B2, B3, B4 and B8), which are available at 10m resolution. Corresponding non-overlapping patches of 32x32 pixels at 10m resolutions were then extracted, and patches containing no-data values in either images or clouds according to Sentinel-2 cloud mask were discarded. Patches were then shuffled and split into 10% testing and 90% training sets. Pleiades patches are blurred with a Gaussian filter tuned to the Sentinel-2 modulation transfer function (MTF) values and down-sampled from 2m to 10m, except for the third network which uses full resolution Pleiades patches as input. Complete characteristics can be found in table 1.

2.2 Network architectures

The reference method consists in a separate linear-regression model for each output band, with the 4 input bands as features. It is trained using the exact same training set.

Its first competitor is the Multi-Layer Perceptron (MLP) network, presented in figure 2(a), which consists in batch normalisation followed by two fully connected hidden layers with 320 units with a leaky-ReLU activation, and an output layer with 4 units and a hyperbolic tangent activation, followed by a skip connection. In the remaining of the paper, this network is called CalibNet.

The second investigated architecture is presented in figure 2(b) and consists in a 2D batch normalisation, followed by two convolutional layers of 64 units with a kernel size of 3, and an output layer of 4 units with same kernel size followed by hyperbolic tangent activation and a skip connection. We refer to this network ConvCalibNet.

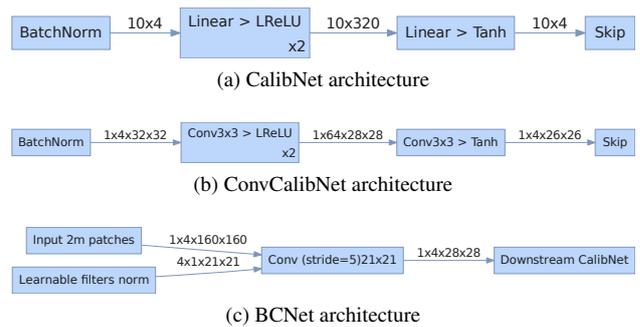


Figure 2. Architectures of the proposed networks. Skip block is a residual connection that adds the input of the network. LReLU is a Leaky ReLU activation function. Learnable filters in BCNet are normalized so that each of them sums up to 1.

The last architecture is shown in figure 2(c). It is similar to the MLP, except for an input module which accepts the full resolution 2m Pleiades image, passes each of its bands through a convolution with a dedicated trainable filter with stride 5, leading to 10m data. For each band, the filter weights are initialised to the Gaussian approximation of the MTF of the corresponding Sentinel-2 bands. To promote the learning of normalised filters, each filter is passed through a 2D soft-max function before the convolution. This network will be called BCNet in the following sections. The rationale for introducing this last architecture is explained in section 3.1.

2.3 Training

Training is performed on a machine with a NVidia Tesla T4 GPU using Pytorch, with batches of 100x32x32 samples for the MLP (for which patches are linearised and samples re-shuffled) and 100 patches for the other two networks. The number of epochs is chosen according to the available number of training samples to perform approximately 5000 iterations. The optimised loss function is the average of the relative error (see equation 1, epsilon being a small quantity to avoid division by zero). The optimiser is Adam with a learning rate of 0.0002 and standard parameters. Training samples are reshuffled for each epoch.

$$L(t, r) = \frac{|t - r|_1}{\epsilon + r} \quad (1)$$

3. RESULTS

3.1 Radiometric Top of Canopy mapping

For the Aude1 dataset, surface reflectances are already very consistent between Pleiades and Sentinel-2, as one can see in the performances chart of all methods in figure 3: visible bands have a RMSE of 0.01, whereas NIR band has a RMSE of 0.07. Nevertheless, we can see on scatter plots (figure 4) that reflectances are over-estimated in Pleiades image for the NIR band, and that for the green band a slight under-estimation of low reflectances and over-estimation of high reflectances exist. This is expected since the spectral bandwidths of Pleiades bands are larger than those of Sentinel-2. Even in this case, where level of radiometric calibration is equivalent and differences are small, the estimation of more consistent values between sensors is important.

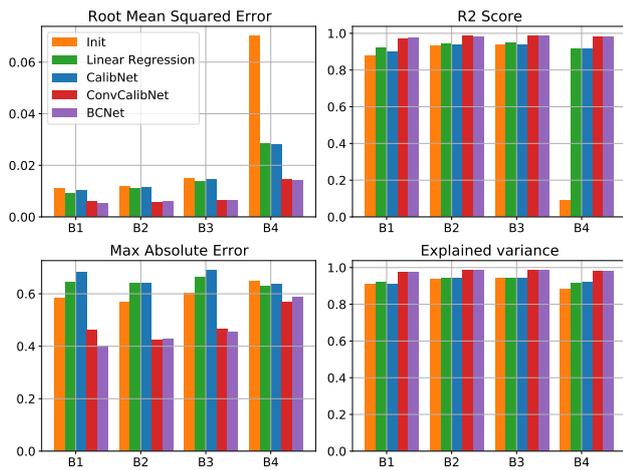


Figure 3. Comparisons of four standard metrics for the initial reflectances and the four proposed methods measured on Aude1 dataset.

Linear Regression alone already enhances greatly the consistency of reflectance values in that case, as one can see in scatter plots (figure 4), and in performances charts (figure 3). It greatly enhances the RMSE of NIR bands and fixes over and underestimations, and brings the NIR band RMSE down to 0.025, at the expense of slightly increasing the maximum error.

During training, CalibNet succeeds in bringing down the relative error loss below the value achieved by linear regression, as shown in figure 5. Training took a little less than 8 minutes in our environment. But the metrics on figure 3 show slightly worse performances with respect to linear regression, and the scatter plots on figure 4 do not show any visible improvement.

Figure 6 shows that ConvCalibNet training succeeds in improving over CalibNet loss, in only 2m30s. Scatter plots on figure 4 show a better, narrower fit, which is also visible in performance metrics (figure 3).

However, if we look at the difference in prediction between linear regression and ConvCalibNet as presented in figure 7, we can see that the prediction error from ConvCalibNet exhibits far less high spatial frequency content. This is due to the convolutional nature of this network, which allows to learn spatial discrepancies such as MTF differences and mis-registration, which is not desired in our case. Indeed, we would like to be able to apply our model at full Pleiades resolution, without altering its spatial content, which is impossible with ConvCalibNet, as shown in figure 8, which shows how our correction affects the input Pleiades image at 2m resolution. We can see that applying CalibNet results in filtering out high spatial frequencies, which is not desired.

The proposed BCNet architecture allows to overcome this problem, by learning a single convolution layer with a single filter for each band before entering the CalibNet MLP architecture. The filter is initialised with a Gaussian kernel mimicking the Sentinel-2 Point Spread Function of the corresponding band, and the kernels are normalised before convolution so as to avoid modifying bands dynamic range at the filter outputs. This first layer concentrates the learning of the spatial discrepancies without altering reflectances, for which the mapping can be learned by the downstream CalibNet architecture. At inference time, the input convolution layer is skipped so as to only apply pixel-wise reflectance mapping.

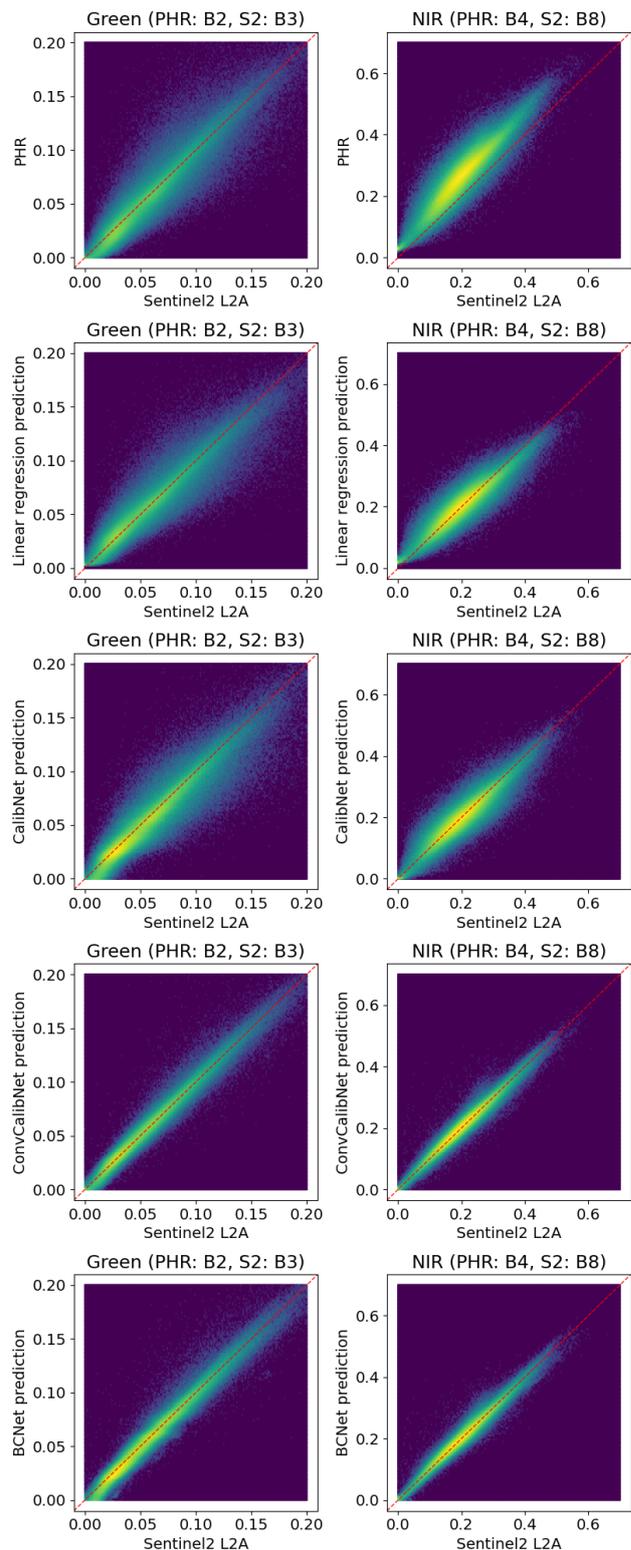


Figure 4. Scatter plot between Pleiades and Sentinel-2 surface reflectance for the Aude1 dataset, for all methods for Green and NIR bands (in this dataset both reflectances are Top of Canopy reflectances)

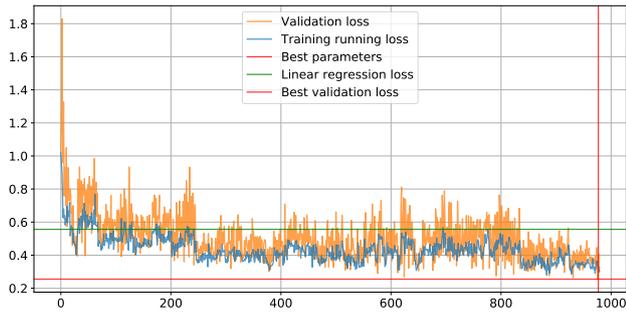


Figure 5. Evolution of training and validation losses during CalibNet network training on dataset Aude1. Green line indicates loss achieved by linear regression. Red lines indicate best loss and iterations.

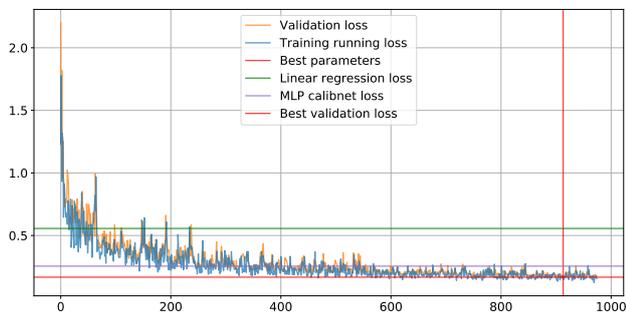


Figure 6. Evolution of training and validation losses during ConvCalibNet network training on datasets Aude1. Green line indicates loss achieved by linear regression, and purple line indicates loss achieved by CalibNet. Red lines indicate best loss and iterations.

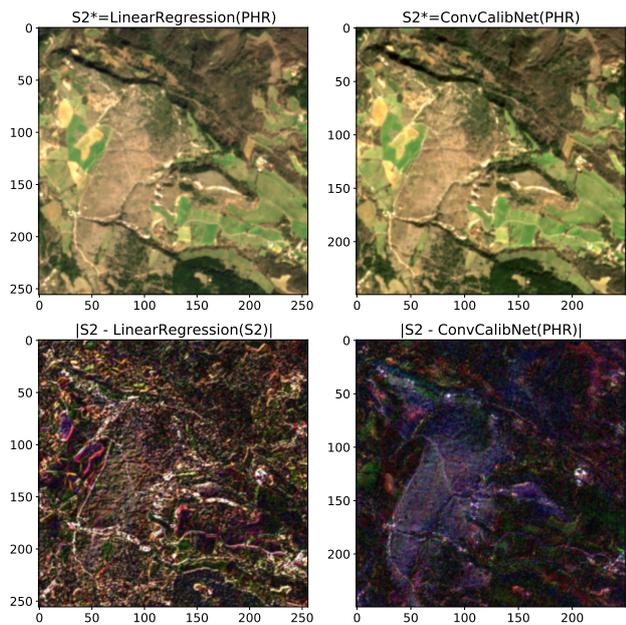


Figure 7. Prediction and prediction error for Linear Regression and ConvCalibNet at 10m resolution.

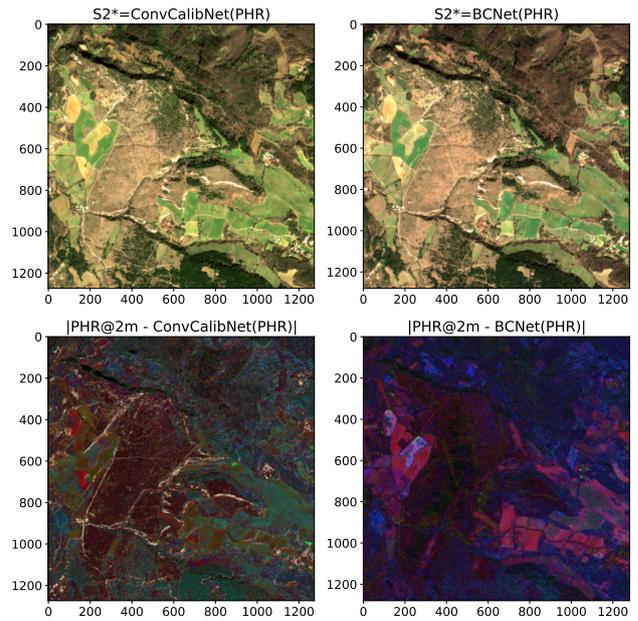


Figure 8. Prediction and alteration of Pleiades image for ConvCalibNet and BCNet at 2m resolution.

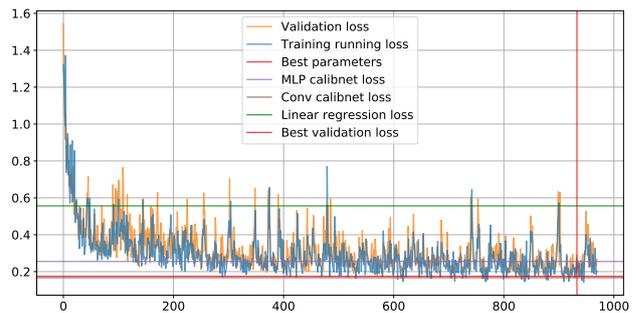


Figure 9. Evolution of training and validation losses during BCNet network training on datasets Aude1. Green line indicates loss achieved by linear regression, purple line indicates loss achieved by CalibNet, and brown line loss achieved by ConvCalibNet (those two last lines are superimposed in the graph). Red lines indicate best loss and iterations.

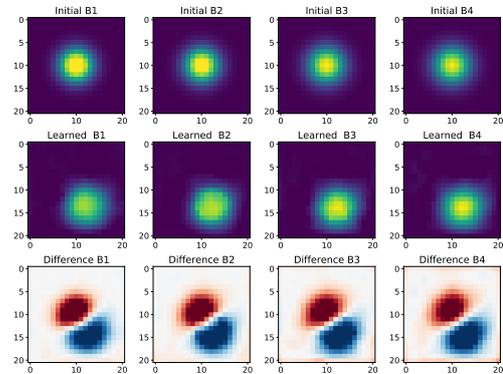


Figure 10. Learned filters in the MTF module of the BCNet architecture, Aude1 case. The module learns both mis-registration and MTF residual, cancelling them for the downstream MLP. Kernels are sampled at 2m resolution.

Figure 9 shows that BCNet reaches the same loss minimum as ConvCalibNet, in 2min36s of training, and those performances are confirmed by looking at scatter plots (figure 4) and metrics (figure 3). Figure 10 shows the filters learning by the first layer of BCNet. We can see that those filters can be interpreted as learning the correct Point Spread Function for each band as well as correcting their spatial registration with the reference data, as we can observe the kernel shifting from their initial centered position. The corrections captured by this layer are not relevant to the radiometric cross-calibration task, and can therefore be skipped during inference.

Indeed, figure 8 shows that by skipping those corrections during inference, spatial details of Pleiades remains unaltered when applying the model at full resolution.

3.2 Radiometric Top of Atmosphere mapping

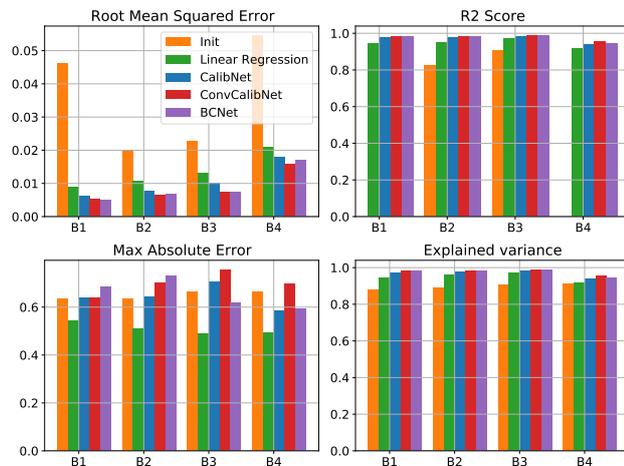


Figure 11. Comparisons of four standard metrics for the initial reflectances and the four proposed methods measured on Aude2 dataset.

In this section, we will investigate the performance of the proposed method on Aude2 dataset, for which Pleiades image contains Top of Atmosphere reflectances.

The initial scatter plots presented in figure 12 obviously show increased discrepancies between the Top of Atmosphere Pleiades reflectances and the Top of Canopy Sentinel-2 reflectances, due to the presence of atmospheric contributions in Pleiades reflectances.

In this conditions, we can see in figure 11 that linear regression is not sufficient to obtain a good mapping, which highlights the non-linearity of atmospheric contributions. The MLP CalibNet architecture allows to capture those non-linearities and to improve RMSE by 15% to 30% depending on the spectral band, which is fairly visible in figure 13. It is important to note that this improvement comes at the expense of increasing the maximum error, which may suggest that the network might be too complex for the task and its supervision. The convolution layer trick brought by BCNet allows to benefit from the same performances without altering the spatial details of the full resolution Pleiades image.

3.3 Dealing with unmasked clouds

Sentinel-2 L2A products distributed by Theia come with handy cloud masking, which allows to only rely on cloud-free pixels during training. Pleiades data on the other hand come with

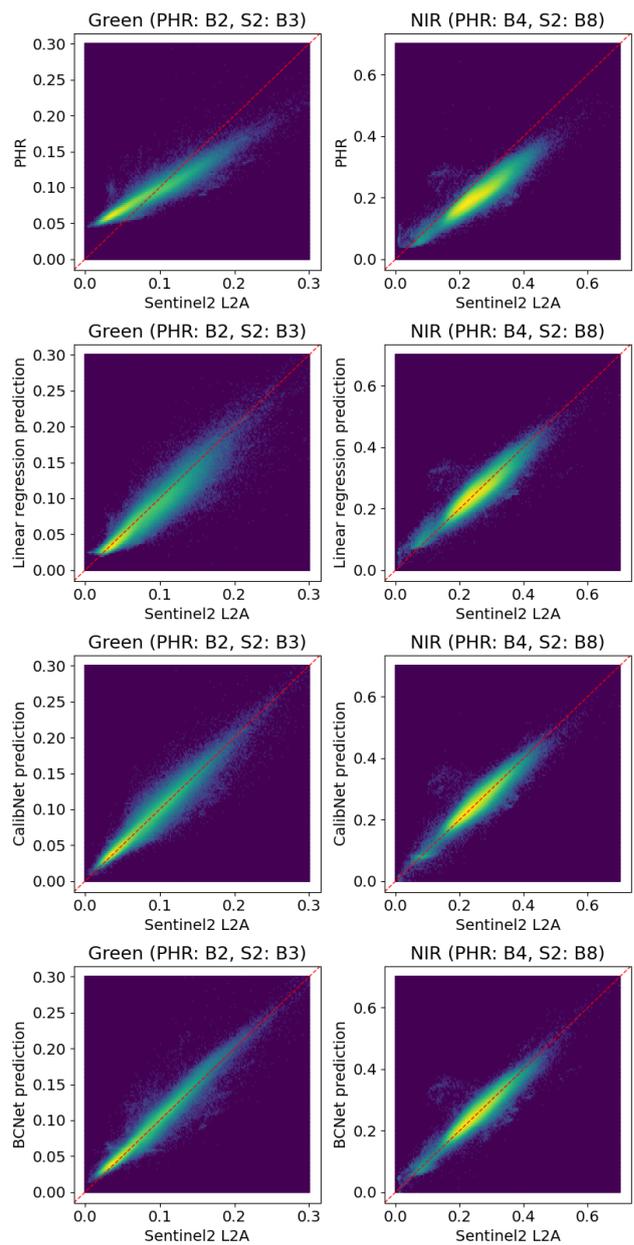


Figure 12. Scatter plot between Pleiades and Sentinel-2 surface reflectance for the Aude2 dataset, for all methods for Green and NIR bands (in this datasets Pleiades image contains Top of Atmosphere reflectances)

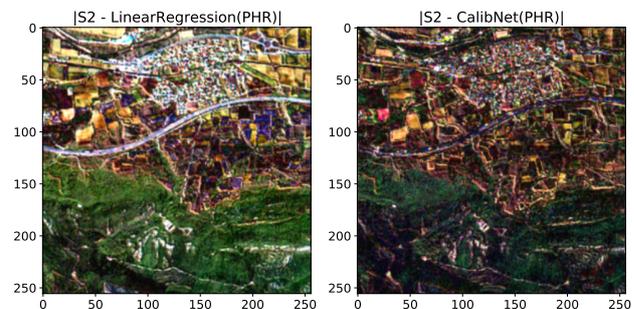


Figure 13. Prediction error of Linear Regression and CalibNet on Aude2 dataset.

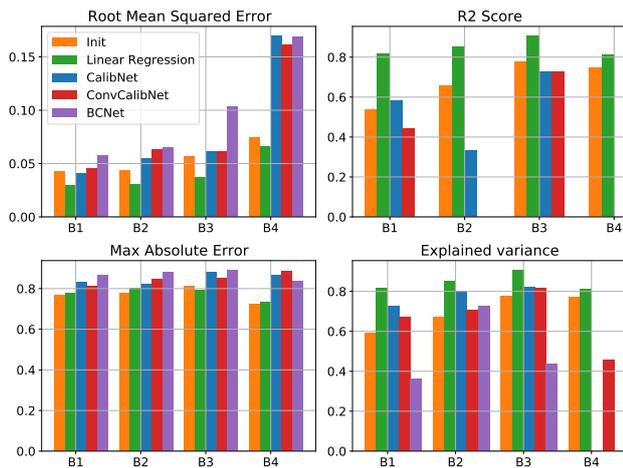


Figure 14. Comparisons of four standard metrics for the initial reflectances and the four proposed methods measured on Bretagne2 dataset, where the Pleiades image is contaminated by unmasked clouds.

a rather coarse mask, that we chose not to use in our experiments. Yet, Pleiades image from the Bretagne2 dataset exhibits high cloud contamination, which prevents all methods to successfully estimate the mapping between Pleiades and Sentinel-2 surface reflectance, as shown in the scatter plots in figure 15. Metrics presented in figure 14 show that none of the methods is able to predict better reflectances than the initial ones.

In order to overcome this problem, we modify all architectures as described in figure 19 to produce a second output which will act as the prediction variance for each band. This will be done by forking the last layer of each architecture, yielding two similar (but different) output layers in parallel. The variance path ends with an Exponential Linear Unit (ELU).

Following (Nix and Weigend, 1994), the loss function is revised as equation 2.

$$L = \frac{(\mu - t)^2}{2 * \sigma^2} + \ln(\sigma), \quad (2)$$

which, up to a constant, corresponds to the negative log-likelihood under a Gaussian error model. Minimising this loss function is equivalent to maximising the log-likelihood of the estimation given the input to the network.

Empirically, this creates a second path for high outlier gradients back-propagation, which can raise the variance output instead of altering the main output. The logarithmic term of the loss ensures that the optimisation does not converge to the trivial infinite variance solution.

Learning with this configuration allows to achieve the same level of performance than cloud-free configurations, as shown in scatter plots (figure 17) and metrics (figure 16).

The new variance output can also be used as a confidence index during inference, as shown in figure 18. This can be used as a cloud mask that is predicted from Pleiades reflectances only, and can be predicted at full resolution. In addition to clouds, this confidence index will denote outlier discrepancies that could not be resolved during learning phase.

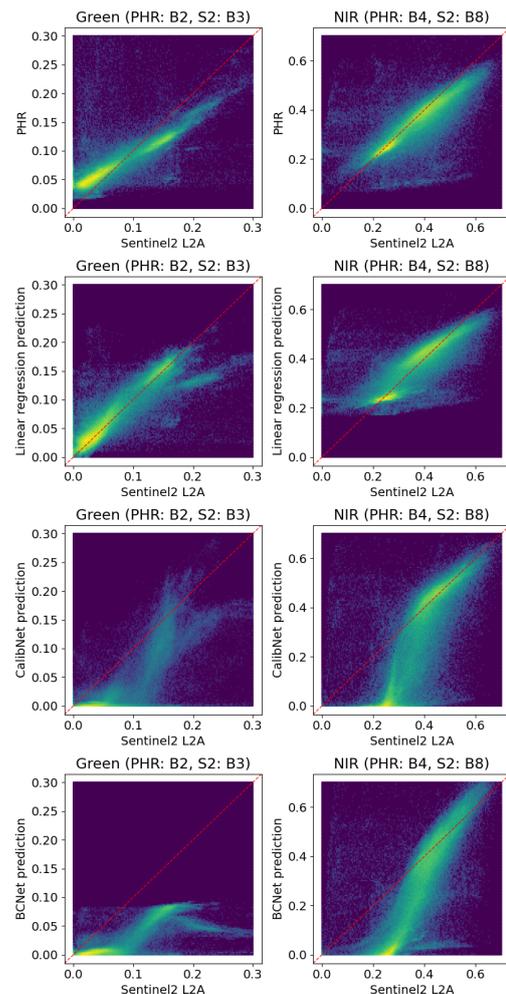


Figure 15. Scatter plot between Pleiades and Sentinel-2 surface reflectance for the Bretagne2 dataset, where the Pleiades image is contaminated by unmasked clouds, for all methods for Green and NIR bands.

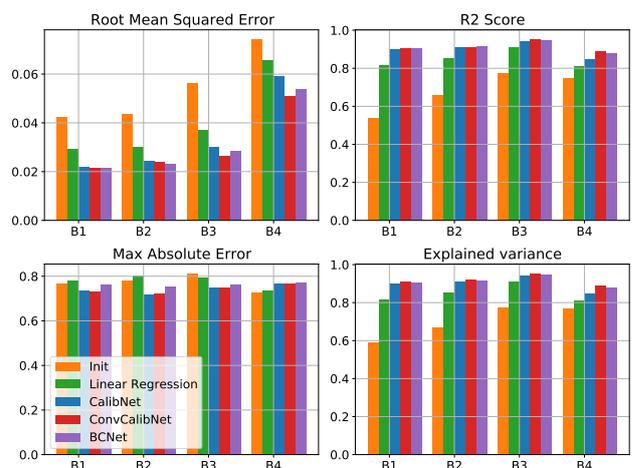


Figure 16. Comparisons of four standard metrics for the initial reflectances and the four proposed methods using variance prediction, measured on Bretagne2 dataset, where the Pleiades image is contaminated by unmasked clouds.

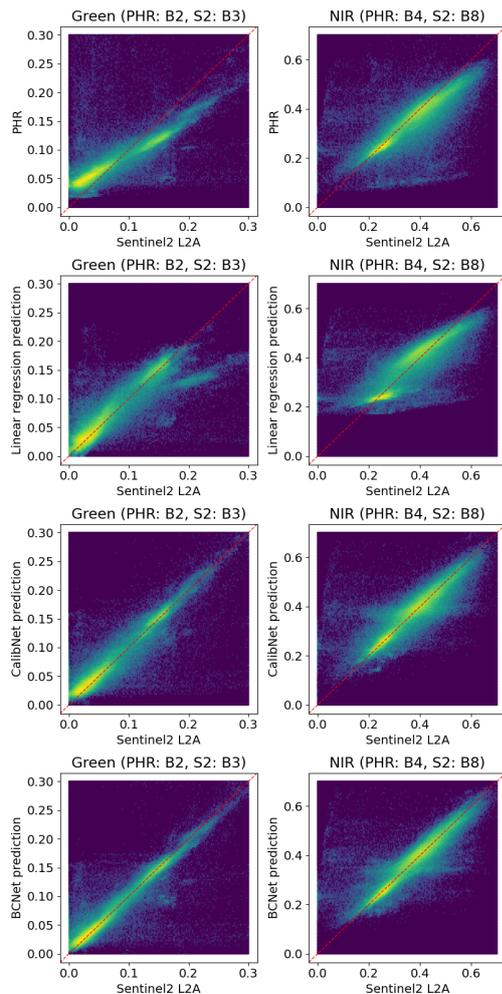


Figure 17. Scatter plot between Pleiades and Sentinel-2 surface reflectance for the Bretagne2 dataset, where the Pleiades image is contaminated by unmasked clouds, for all methods, using variance prediction, for Green and NIR bands.

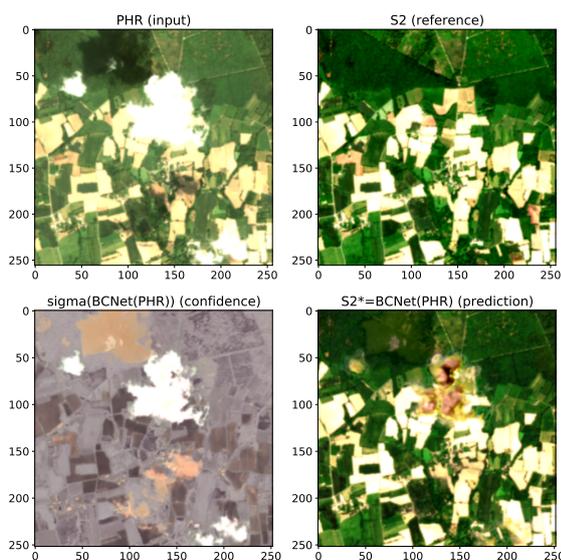


Figure 18. Pleiades input, Sentinel-2 reference, BCNet confidence and prediction for a patch of the Bretagne2 dataset.

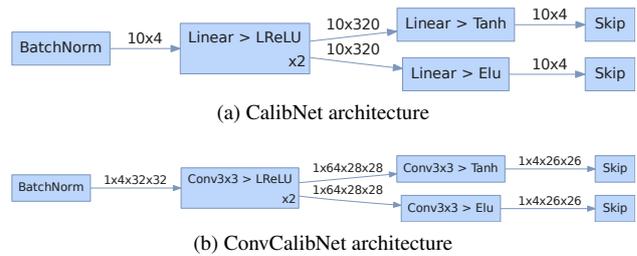


Figure 19. Modifications of the proposed network to include a second output corresponding to the variance of the prediction.

4. CONCLUSIONS

In this paper, we address the problem of obtaining consistent surface reflectance values between Sentinel-2 images and Pleiades images acquired with varying angular conditions and at different times. We saw that physical differences such as BRDF effects or spectral bandwidth, but also differences in radiometric calibration processing chains, can yield significant radiometric differences and impair high level information extraction methods using both sensors.

Instead of trying to model and correct all those differences, we propose to learn harmonised reflectances by means of machine learning techniques, using Sentinel-2 as a reference. In addition to linear regression which has been selected as a baseline method, we compared four different architectures based on Neural Networks. We evaluated those methods on four different datasets over two different sites in France, learning one model for each dataset, since we do not seek for generalisation.

We saw that using a MLP based architecture does not bring much improvement with respect to linear regression if Pleiades products are already calibrated at Top of Canopy level. Both can yield significant improvements especially in the NIR band, where the RMSE error is more than twice smaller. However, when using Top of Atmosphere Pleiades products as input, using a MLP based architecture brings 10% to 25% of improvement to all bands with respect to linear regression. We can assume that the correction to learn is more non-linear in that case, since it contains the removal of the contribution of the atmosphere.

At first sight, using a convolutional architecture seemed to yield even better results but this is mostly due to the learning of spatial discrepancies such as mis-registration or incomplete MTF knowledge, and such a model will alter the high spatial frequency of the Pleiades image when applied at full resolution. To overcome this issue, we proposed a different architecture, with a single input convolution layer as input to the MLP architecture, tailored to learn the spatial discrepancies. This layer can be skipped during inference to fully preserve the high resolution spatial content. Filters learned by this removable layer indeed show MTF and registration residuals.

Last, the process of learning the reflectance mapping can also be altered by unmasked clouds in the input Pleiades image. We therefore propose to add a second output to our networks in order to infer variance of prediction as a proxy for confidence. This new output is trained with a dedicated statistical loss. We shown that this yields stable result in presence of cloud in the input samples, and can also serve as a simple cloud and confidence mask that can be inferred from Pleiades pixels alone, at full resolution.

REFERENCES

Claverie, M., Ju, J., Masek, J. G., Dungan, J. L., Vermote, E. F., Roger, J.-C., Skakun, S. V., Justice, C., 2018. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sensing of Environment*, 219, 145 - 161.

Lonjou, V., Desjardins, C., Hagolle, O., Petrucci, B., Tremas, T., Dejus, M., Makarau, A., Auer, S., 2016. Maccs-atcor joint algorithm (maja). *Remote Sensing of Clouds and the Atmosphere XXI*, 10001, International Society for Optics and Photonics, 1000107.

Luo, Y., Guan, K., Peng, J., Wang, S., Huang, Y., 2020. STAIR 2.0: A Generic and Automatic Algorithm to Fuse Modis, Landsat, and Sentinel-2 to Generate 10 m, Daily, and Cloud-/Gap-Free Surface Reflectance Product. *Remote Sensing*, 12(19), 3209.

Nix, D., Weigend, A., 1994. Estimating the Mean and Variance of the Target Probability Distribution. *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*. <https://doi.org/10.1109/icnn.1994.374138>.