DEEP LEARNING AS A TOOL TO INTERPOLATE CLOUDY PIXELS IN SENTINEL-2 TIME SERIES

Baggio Luiz de C. e Silva [1], Wesley Augusto Campanharo[1], Felipe Carvalho de Souza[1] , Karine R. Ferreira[1] e Gilberto Ribeiro de Queiroz[1]

[1]Instituto Nacional de Pesquisas Espaciais, Av. dos Astronautas, 1758, São José dos Campos, SP, Brazil.
(baggio.silva, wesley.campanharo, felipe.carvalho, karine.ferreira, gilberto.queiroz)@inpe.br

ABSTRACT

Optical remote sensors are extremely susceptible to clouds. Clouds and their shadows affect remote sensing image processing methods to automatically identify and classify land use and land cover types. To detect cloud and cloud shadows in remote sensing images, many algorithms have been proposed, such as FMask, Sen2Cor and s2cloudless. In image time series analysis, interpolation techniques are used to produce valid values when pixels are covered by clouds or shadows. This paper evaluates the use of deep learning approaches to interpolate cloudy pixels in Sentinel-2 time series. Twelve different model configurations were evaluated and their differences and limitations were highlighted. The model proved to be very promising in dealing with the limitations of the cloud mask interpolating clouds and cloud shadows.

Key words *– Cloud interpolation, Satellite image time series, Cloud and shadows, Neural network, Multilayer perceptrons.*

1. INTRODUCTION

Passive sensors, such as the Multi-Spectral Instrument (MSI) onboard Sentinel-2, are extremely susceptible to clouds. Even with this highly temporal recurrence, around five days considering both satellites [1], clouds affect the ability to identify land covers, decreasing the accuracy of surface parameters as they interfere in the solar and terrestrial radiation [2, 3].

In Brazil, due to its continental proportion, the cloud cover follows an irregular temporal and spatial distribution. For instance, the Pampa biome presents a constant number of clear imagery regions throughout the year. At the same time, the Amazon suffers from a lack of clear images, more frequently available during the dry season but not too successfully in the Northwest of the biome [4].

Most optical imagery already uses cloud detection algorithms to correct the radiance, indicate cloudy areas, and indirectly help users work properly with remote sensing imagery. In this sense, we can point out the Fmask [5] used in the Landsat collection; Sen2Cor [6] used by the European Space Agency (ESA) in their Sentinel-2 satellite products, and; s2cloudless [7] used by the Sentinel Hub's in its derivation products from known satellites.

However, those techniques can overestimate cloud areas harming temporal analyzes. A study performed by Sanchez et al. (2020) [8] in the Amazon tropical forest shows that using the FMask 4 in Sentinel-2 images results in an overall accuracy greater (90%) than the regular Sen2Cor (79%). While Coluzzi et al. (2018) [3] shows that globally the Sentinel-2 cloud mask underdetect systematically, reaching a maximum difference (70%) in the Amazon basin.

In analyses that use temporal information, such as deforestation, forest degradation, and phenology, the lack of valid data due to cloud periods is crucial, making the interpolation technique required. There are many available methods to interpolate time series, since more simple such as nearest-neighbor interpolation, until stochastic methods, such as machine learning methods [9].

In this sense, this paper aims to verify if the use of deep learning is capable to improve visually and temporally the values of an image without the use of a cloud mask, and adopting all-time series as input.

2. MATERIAL AND METHODS

We use the Sentinel-2 images from the Brazil Data Cube (BDC) project. The project generates multidimensional analysis-ready data cubes. We use `S2-SEN2COR_10_16D_STK-1` product that contains MSI surface reflectance at full spatial resolution (10m) and 16 days Temporal Compositing considering the SCL cloud mask. The temporal composite is generated from the images with less clouds over the 16 days [10, 11].

The Multilayer perceptrons (MLP) was chosen to test a new interpolate approach. This method is a feedforward neural network that indicates that the information flows through layers always forward until it approximates to the determined function and finally reaches the output [12]. The MLP implementation used is from the `scikit-learn MLPRegressor` library [13].

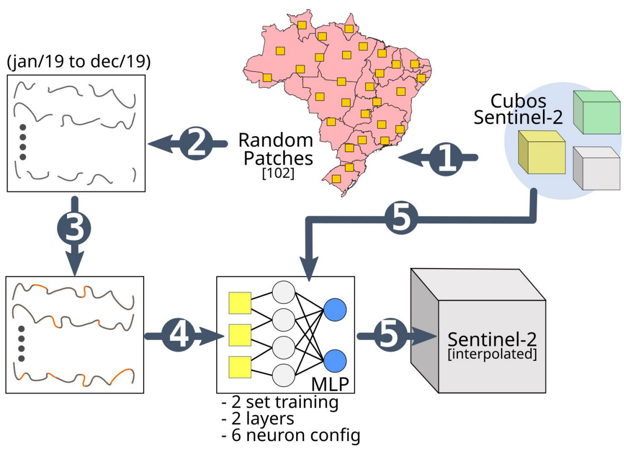

Figure 1: The methodological scheme used in this study, with the main tasks: 1-random points generation; 2-extraction of the time series; 3-linear interpolation; 4-training the MLP model; 5-predicting the cloud-free values for a year of images.

The methodology follows five main steps, as shown in Figure 1. First, we collected 102 random patches across Brazilian territory, selecting areas that encompass each class of clouds following the Sentinel cloud mask (Scene Classification Layer - SCL). Each patch has a 500 x 500 pixels size, encompassing 250,000 time series, which will represent a total of 25,500,000 samples.

Secondly, we extract the time series of one year (01-01-2019 to 12-31-2019) from the Sentinel-2 data cube, including three bands (B04-Red, B08-Nir, B11-Swir). Next, those time series were interpolated using the `interp` function by the `numpy` library [14], based on the classes "No data", "Saturated or defective", "Dark areas", "Cloud shadow", "Cloud medium probability", "Cloud high probability", "Thin cirrus" and "Snow" contained in the SCL mask. Later, those time series were used to train the MLP model, and then the model was applied in different areas with different cloud cover structures and land use and land cover patterns.

The strategy of this paper is to interpolate the time series without using a cloud mask. In this sense, the model input consisted of two-dimensional arrays. The first dimension is the number of samples, and the second is the time resolution of the time series, resulting in an output with the same time dimension as the input.

Two hidden layers were defined with six variations in the number of neurons: [8,4], [16,8], [32,16], [64,32], [128,64], and [256,128], where the first value is the number of neurons in the first layer and the second value is the number of neurons in the second layer.

Another configuration used is related to the total training samples. From the 25,500,000 random time series extracted for the whole Brazilian territory, we set two different networks, one with a step size of 10 pixels and another with a step size of 100 pixels, summarizing 2,550,000 and 255,000 time series, respectively. The activation function used was `relu` with the solver `adam`, the `alpha` was set to equal to 0.001, `batch` size to 1000, `learning rate init` to 0.001, validation fraction of 20%, and the other parameters as default.

We use two methods to validate the deep-learning approach. First, a visual procedure was applied, this method consist in to indicate if the patterns, texture, and features were modified, and all those features were indicated by a specialist.

The second procedure was to check if the physical value (superficial reflectance) changes at the point to invalidate the intrinsic characteristic of the object. We collect random points and compare our approach with linear interpolation. The root mean squared error (RMSE), the mean squared error (MSE), and the mean absolute error (MAE) were used to check the divergences between both approaches.

3. RESULTS

Figure 2 shows the application of the 12 models for one pixel of the NIR band along with the raw time series and the linear interpolated time series based on the cloud mask. The gray highlighted background represents the pixels identified as clouds and needs interpolation.

The gain between the two sets of samples (255,000; 2,550,000) is not too large as compared to the number of neurons in each layer. In this sense, we notice that for the model with fewer neurons, like the 8_4, 16_8, and 32_16, the time series gets more flat, losing all the variation. While layers with more numbers of neurons, such as 64_32, 128_64, and 256_128 are more similar to the linear interpolation.

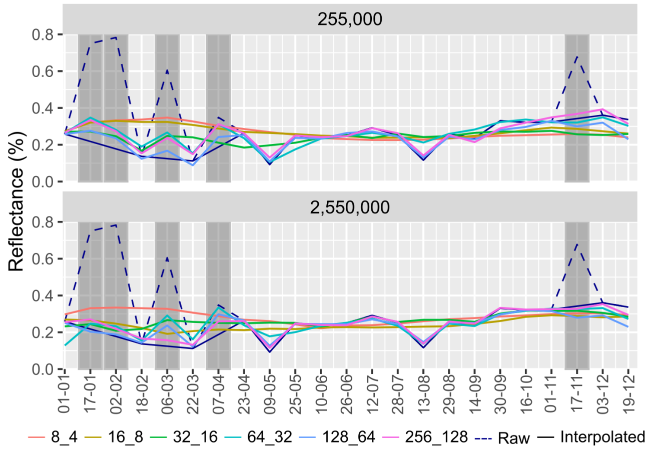

Figure 2: Result for the six models of each training set, comparing it to the linear interpolation and the raw data. The bars represent the cloud class extract from the SCL band.

Table 1 shows the reference values for each model with their respective MSE, RMSE, and MAE values calculated using linear interpolation as a baseline.

Model configuration	S255,000			S2,550,000		
	MSE	RMSE	MAE	MSE	RMSE	MAE
8_4	0.0121	0.1100	0.0866	0.0100	0.1000	0.074
16_8	0.0100	0.1000	0.0758	0.0038	0.0616	0.0526
32_16	0.0049	0.0701	0.0577	0.0480	0.0694	0.0461
64_32	0.0038	0.062	0.0431	0.004	0.0629	0.0426
128_64	0.0014	0.0374	0.0281	0.0022	0.0466	0.0299
256_128	0.0026	0.0510	0.0335	0.0007	0.0258	0.0177

Table 1: Metrics comparing the six models of each training set with linear interpolation.

Figure 3 shows two scenarios of cloud cover selected. In the first, there is the presence of Cirrus clouds and large shadows, indicated by the red and blue polygons. While in the second, there are dense clouds and well-defined shadows. In the baseline section, the raw image and the linear interpolated method result are displaced.

To track the visual efficiency generated by the MLP method, we distribute the images by the number of samples (column) and the number of neurons (lines).

The image reconstructions show a higher variance between the 12 results from the MLP model. Visually the low number of neurons highlighted the features more than the linear interpolations, increasing the contrast and the visual acuity.

However, the cloud's shadow is a persistent artifact in the MLP model, mainly in the model set up with more neurons. From 32_16 the cloud's shadows start to appear as an artifact of the MLP, presents only in the 2,550,000 sample set. After this point, the presence of cloud artifacts was more present, including the insertion of bare soil in the middle of an agricultural area.

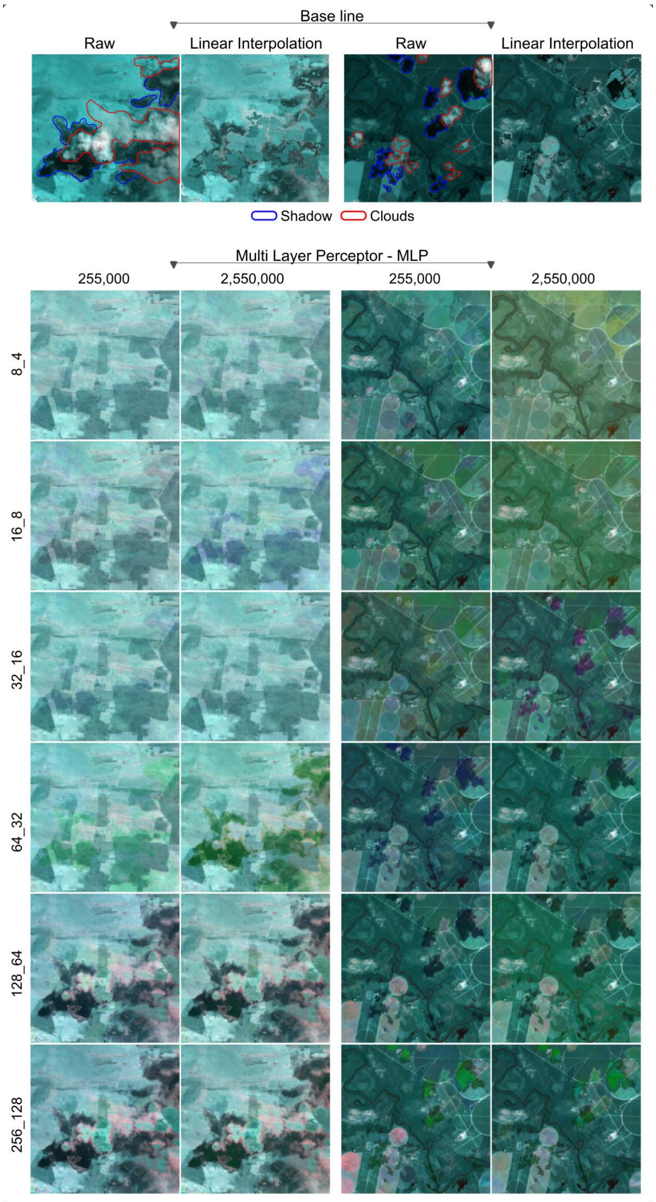

Figure 3: Visual comparison between the original image (Raw), the linear interpolated, and the MLP results. The column shows the results with the different numbers of samples, and the lines show the variation of neurons in each layer. Composite Color Image: Red-B04, Nir-B08 e Swir-B11.

4. DISCUSSION

Table 1 shows a small error, with some configurations performing better quantitatively. Looking at Figure 3, we can observe some characteristics of each model. The models with fewer neurons were smoother and removed most of the temporal breaks. This characteristic is also evident when we reconstruct the image, as in Figure 3. The configurations with fewer neurons removed both the cloud and shadow influence. However, as a result, the temporal variability of use and coverage was lost. At the same time, the other images showed little variability, tending to remove, for example, the crop rotation.

The models with more neurons behaved more like linear interpolation. As a result, where the cloud mask has more errors, the model also "learned" the same errors, making it very difficult to interpolate cloud shadows. Although, it can even interpolate some types of clouds well, with the advantage of preserving some temporal variability in the other images.

The Linear Interpolation (LI) uses the SCL cloud mask to orientate where the algorithm should be interpolated. In this sense, in places where the mask made a misclassification, the pixel still shows the influence of the cloud. In those areas, the imagery presents a very characteristic cloud noise.

As pointed out by some authors [3, 8], the SCL mask underdetect systematically clouds. Thus, even with the best method of filling a gap, the fact of needing the mask as a guide still makes remains noisy in the images.

There are some deep-learning approaches to dealing with clouds. However, they are more interested in indicating where they are and passing the users the process to fill the gap. This is the case of the S2cloudless [7]. However, even this algorithm uses a cloud mask as training data.

Depending on the number of neurons used in each layer, the MLP model resembles linear interpolation. However, in those scenarios, the model shows difficulties in dealing with cloud shadow, not smoothing the valley as visible in the models with fewer neurons.

The size and quality of the samples are very important in machine learning applications [7, 15]. The two sets did not disagree too much visually and physically in our approach.

5. CONCLUSIONS

Optical remote sensors are extremely susceptible to clouds, leaving it up to the user to remove them and deal with possible failures. The use of deep-learning models brings an advantage because, simultaneously, the presence of clouds is removed, and the data are already interpolated and smoothed, two essential steps usually performed when working with time series.

Our study shows that the Multilayer perceptron (MLP) has great potential to perform the basics process leaving the image ready to use in time series as pixel data or imagery composition. However, the choice of the number of neurons must be considered since, when reducing this parameter to remove more cloud artifacts, the final interpolation excludes significant variations of features, such as wetlands, crop rotation, and vegetation growth.

This study was limited to comparing, at the first moment, the MLP model as an alternative to the linear interpolation technique. Further studies need to be developed to check the factual accuracy of the interpolation, the depth of the model more suitable considering all Brazilian territory, and its potential use to rebuild a cloud mask.

In future work, we intend to use visual interpretation, not the SCL cloud mask, to identify clouds and shadows in the random time series used to train the MLP method. Thus, we will be able to evaluate the deep learning approach regardless of the SCL cloud mask errors.

Acknowledgment

We would like to thank the Brazil Data Cube project, which is part of the "Environmental Monitoring of Brazilian

Biomes project", funded by the Amazon Fund through the financial collaboration of the Brazilian Development Bank (BNDES) and the Foundation for Science, Technology and Space Applications (FUNCATE) no. 17.2.0536.1; and to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - Finance Code 001, for the Silva, B.L.C. scholarship.

6. REFERENCES

[1] ESA. Sentinel-2, 2022.

[2] John R. Jensen and José Carlos Neves Epiphanio. *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*. Parêntese, São José dos Campos, SP, 2011.

[3] Rosa Coluzzi, Vito Imbrenda, Maria Lanfredi, and Tiziana Simoniello. A first assessment of the sentinel-2 level 1-c cloud mask product to support informed surface analyses. *Remote Sensing of Environment*, 217:426–443, 2018.

[4] Jean Francois Mas, Carlos Henrique Sopchaki, Francisco Davy Braz Rabelo, Francisca Soares de Araújo, and Jonathan Vidal Solórzano. Análise da disponibilidade de imagens landsat e sentinel para o brasil. *Geografia Ensino & Pesquisa*, 24, dez. 2020.

[5] Shi Qiu, Zhe Zhu, and Binbin He. Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4–8 and sentinel-2 imagery. *Remote Sensing of Environment*, 231:111205, 2019.

[6] Jérôme Louis, Vincent Debaecker, Bringfried Pflug, Magdalena Main-Knorn, Jakub Bieniarz, Uwe Mueller-Wilm, Enrico Cadau, and Ferran Gascon. Sentinel-2 sen2cor: L2a processor for users. In L. Ouwehand, editor, *ESA Living Planet Symposium 2016*, volume SP-740 of *ESA Special Publications*, pages 1–8. Spacebooks Online, August 2016.

[7] SentinelHub. Cloud masks and cloud probabilities, 2022.

[8] Alber Hamersson Sanchez, Michelle Cristina A. Picoli, Gilberto Camara, Pedro R. Andrade, Michel Eustaquio D. Chaves, Sarah Lechler, Anderson R. Soares, Rennan F. B. Marujo, Rolf Ezequiel O. Simões, Karine R. Ferreira, and Gilberto R. Queiroz. Comparison of cloud cover detection algorithms on sentinel–2 images of the amazon tropical forest. *Remote Sensing*, 12(8):1284, Apr 2020.

[9] Mathieu Lepot, Jean-Baptiste Aubin, and François H.L.R. Clemens. Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment. *Water*, 9(10), 2017.

[10] Karine R. Ferreira, Gilberto R. Queiroz, Lubia Vinhas, Rennan F. B. Marujo, Rolf E. O. Simoes, Michelle C. A. Picoli, Gilberto Camara, Ricardo Cartaxo, Vitor C. F. Gomes, Lorena A. Santos, Alber H. Sanchez, Jeferson S. Arcanjo, José Guilherme Fronza, Carlos Alberto Noronha, Raphael W. Costa, Matheus C. Zaglia, Fabiana Zioti, Thales S. Korting, Anderson R. Soares, Michel E. D. Chaves, and Leila M. G. Fonseca. Earth observation data cubes for brazil: Requirements, methodology and products. *Remote Sensing*, 12(24), 2020.

[11] BDC. Brazil data cube - cubo de dados, 2022.

[12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[14] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[15] Aaron E. Maxwell, Timothy A. Warner, and Fang Fang. Implementation of machine-learning classification in remote sensing: an applied review. *International Journal of Remote Sensing*, 39(9):2784–2817, 2018.