

# STMETRICS: A PYTHON PACKAGE FOR SATELLITE IMAGE TIME-SERIES FEATURE EXTRACTION

Anderson R. Soares<sup>1\*</sup>, Hugo N. Bendini<sup>1 †</sup>, Daiane V. Vaz<sup>1</sup>, Tatiana D. T. Uehara<sup>1, ‡</sup>, Alana K. Neves<sup>1 §</sup>, Sarah Lechler<sup>2</sup>, Thales S. Körting<sup>1 ¶</sup>, Leila M. G. Fonseca<sup>1</sup>

General Coordination of Earth Observation - OBT<sup>1</sup>  
Brazil's National Institute for Space Research (INPE)  
Av. dos Astronautas, 1758, São José dos Campos - SP, Brazil  
(anderson.soares,hugo.bendini,daiane.vaz,tatiana.uehara,alana.neves, thales.korting,leila.fonseca)@inpe.br  
Institute for Geoinformatics, University of Muenster<sup>2</sup>  
s.lechler@uni-muenster.de

## ABSTRACT

Producing reliable land use and land cover maps to support the deployment and operation of public policies is a necessity, especially when environmental management and economic development are considered. To increase the accuracy of these maps, satellite image time-series have been used, as they allow the understanding of land cover dynamics through the time. This paper presents the *stmetrics*, a python package that provides the extraction of state-of-the-art time-series features. These features can be used for remote sensing time-series image classification and analysis. *stmetrics* aims to be an easy-to-use package. The package is available under the GNU GPL software license, and the full source code is available for download at: [github.com/andersonreisoares/stmetrics](https://github.com/andersonreisoares/stmetrics).

**Index Terms**— time-series, python, multi-temporal features, remote sensing

## 1. INTRODUCTION

Satellite image time-series (SITS) became an important tool to produce and regularly update accurate temporal land-cover maps [1]. SITS allow the analysis of the dynamic of land cover through the time component. They offer a great potential for understanding the land cover behavior.

\*Thank 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 (Brazil Data Cube project)

<sup>†</sup>The project “Development of systems to prevent forest fires and monitor vegetation cover in the Brazilian Cerrado” (World Bank Project #P143185) – Forest Investment Program (FIP).

<sup>‡</sup>This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

<sup>§</sup>Thank the National Council for Scientific and Technological Development (CNPq).

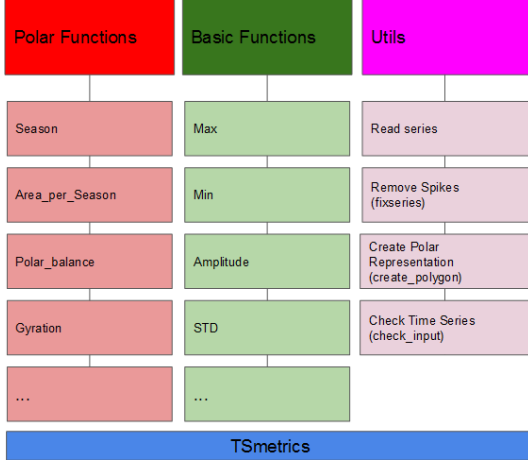
<sup>¶</sup>Grant # 2017/24086-2, São Paulo Research Foundation (FAPESP).

As time-series can be related to land cover patterns, the extraction of features from time-series to perform Land Use and Land Cover (LULC) classification became a common approach [2], especially in the remote sensing field. In order to handle the enormous amount of data available in these datasets, data mining has been employed as a supporting tool. According to [3], data mining can be used to discover correlations, patterns and trends in the dataset through techniques such as pattern recognition, mathematics and statistics.

Several types of features can be combined to assist classification models, such as basic metrics that are derived from statistical measures like mean, standard deviation and minimum and maximum values of the time-series [4]. Besides that, [5] proposed the use polar metrics, that are derived from a time wheel legend, which projects the values of a time-series to angles in the interval  $[0, 2\pi]$ . According to the authors, it provides a new visualization scheme that can help to describe the pattern represented in the time-series. Another important group of metrics are the phenological metrics or *phenometrics*. These metrics aim to be directly related to the agriculture calendar, as the assumption is that every crop type under investigation has an unique phenology [6]. Usually, robust methods are implemented to extract this type of features. They perform local mathematical models fitting to time-series, aiming to handle noisy data and non-periodic time-series with different types of changes [7].

The use of phenological features extracted from image time-series, such as minimum and maximum values and the respective times of their occurrence, can be important assets for developing agriculture mapping methods. [8, 9] showed that phenological metrics can be derived at a 30-meter spatial resolution, and a high temporal resolution, between 5 and 8 days, enables mapping of agricultural classes with good accuracies ( $> 80\%$ ).

In this paper, we present a new python package that has been developed to extract basics and polar features from



**Fig. 1.** General overview of the package structure.

SITS. The paper is structured as follows: first, it is demonstrated how the package is organised and how the data is structured in section 2; in section 3, we present some results and applications for SITS classification and, at section 4, conclusion and remarks are presented.

## 2. THE PACKAGE

The *smetrics* package relies on other commonly used open source libraries: Numerical and Scientific Python (NumPy and SciPy) [10, 11], Matplotlib [12], Shapely [13] and Descartes[14]. A general overview on the architecture and execution pipeline of the package is presented at Fig. 1.

The package has currently two modules for metrics extraction. The first one computes basic metrics, that are derived from the time-series using common statistical approaches. With this module, 7 metrics are computed: mean, maximum, minimum, standard deviation, accumulated sum, amplitude and the maximum value of the first slope. They are referred on the package as *basic metrics*.

The second module implements the metrics proposed by [5]. The authors adapted the time wheel legend proposed by [15], by plotting each cycle of the profile, and by projecting values to angles in the interval  $[0, 2\pi]$ . Let a cycle be a function  $f(x, y, T)$  where  $(x, y)$  is the spatial position of a point, and  $T$  is a time interval  $t_1; \dots; t_N$ , where  $N$  is the number of observations in such a cycle. The cycle can be visualized by a set of values  $v_i \in V$ , where  $v_i$  is a possible value of  $f(x, y)$  in  $t_i$ . Let its polar representation be defined by a function  $g(V)\{A, O\}$  ( $A$  corresponds to the abscissa axis in the Cartesian coordinates, and  $O$  to the ordinate axis) where:

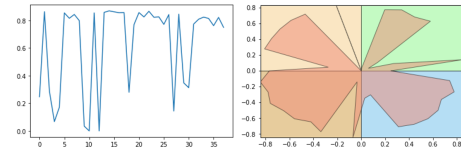
$$a_i = v_i \cos\left(\frac{2\pi i}{N}\right) \in A, i = 1, \dots, N. \quad (1)$$

and

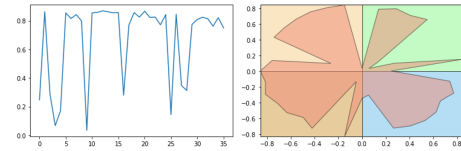
$$o_i = v_i \sin\left(\frac{2\pi i}{N}\right) \in O, i = 1, \dots, N. \quad (2)$$

In both equations,  $\frac{2\pi i}{N}$  is an arbitrary angle that depends on the acquisition date and  $v_i$  is the corresponding time-series value. Table 2 describes the multi-temporal features (basic and polar) available in the package.

There is a filtering approach to remove spikes from the time-series. A spike, in the package context, is a point in the time-series that is preceded and succeeded by points that have 0 as value. On Fig. 2, at the orange quadrant is possible to see the differences caused by the filtering approach. This filtering approach is especially necessary due to the polar transformation since this could lead to invalid polygons. This function is used previously to the basic and polar modules. Currently, the package does not perform smoothing or fitting functions on the time-series, nor have methods to it.



(a) Time-series and polar representation without filtering.



(b) Time-series and polar representation with filtering.

**Fig. 2.** Example of time-series before and after filtering.

Another important operation is the replacement of all negative values on a time-series by their modules. This operation is also necessary due to polar transformation, as complex polygons may be produced during the transformation if the points have negative values. Differently from the filtering approach, this operation is only performed previously to the polar metrics.

## 3. APPLICATIONS

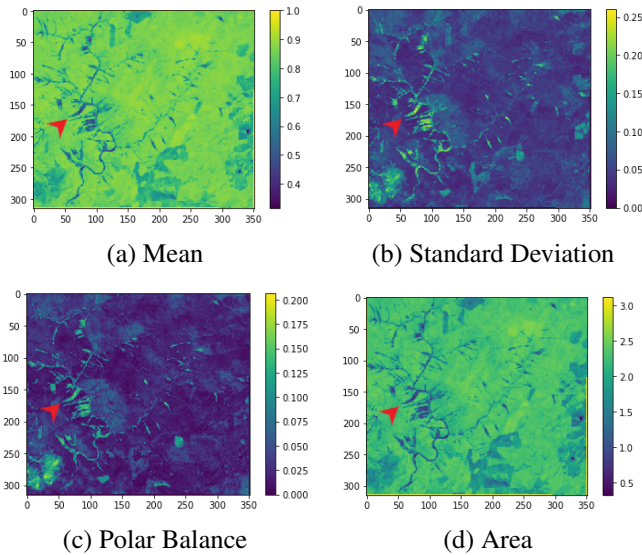
To present the applicability of the package, we present two study cases, the identification of a landslide risk area at Rio Grande do Sul state in Brazil and the classification of Land-Use and Land-Cover (LULC) in the region of Vale do Rio Doce, located in the state of Minas Gerais, Brazil.

For the first study case, the landslide identification on the south region of Brazil, a time-series composed by 26 Normalized Difference Vegetation Index (NDVI) images from November 2015 to July 2019 was used. Polar and basic metrics were extracted, some examples are illustrated in Figure 3. The selected area is mainly composed of dense forest, pasture and landslide scars along a dry river. In the basic metric *mean*, the landslide scars are easy identified by their low val-

**Table 1.** Temporal features currently available on *stmetrics*.

Name	Type	Description	Range
Amplitude	Basic	The difference between maximum and minimum values of a time-series.	[0,1]
Mean	Basic	Average value of the time-series.	[0,1]
Std	Basic	Standard deviation of the time-series.	$\geq 0$
Maximum	Basic	Maximum value of a time-series, sensitive to false highs and noise.	[0,1]
Minimum	Basic	Minimum value of a time-series.	[0,1]
Accumulated Sum	Basic	The sum of values over a time-series.	$\geq 0$
First Slope	Basic	Indicates when the TS presents some abrupt change.	[-1,1]
Eccentricity	Polar	Values close to 0 if the shape is a circle and 1 if the shape is similar to a line.	$\geq 0$
Gyration radius	Polar	Average distance between each point inside the shape and the shape's centroid.	$\geq 0$
Polar balance	Polar	The standard deviation of the areas per season, considering the 4 seasons.	$\geq 0$
Area	Polar	Area of the closed shape. A higher value indicates a series with high values.	$\geq 0$
Angle	Polar	The main angle of the closed shape created by the polar visualization.	$\geq 0$
Area per Season	Polar	Partial area of the closed shape in each quadrant of the polar representation.	$\geq 0$

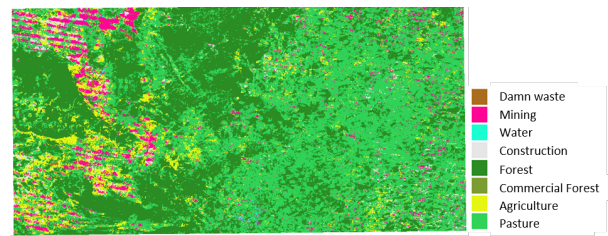
ues (dark blue), once these are not vegetated areas. The landslide scars correspond to the strip format features along the dry river path. On the other hand, once the event caused the overturn of the dense forest, reducing drastically the NDVI values, the *standard deviation* metric allows the identification of the scars by their high values, represented in yellow. Moreover, regarding the polar metrics, the landslide scars present high values for the metric *polar balance* and low for *area*.



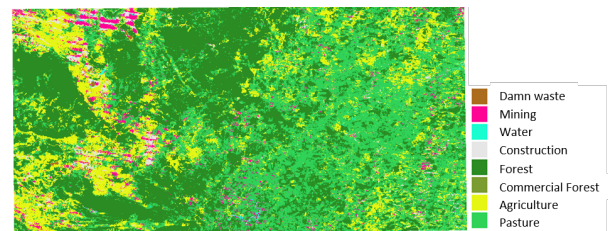
**Fig. 3.** Examples of the spatialization of basic and polar features extracted from all time-series in the first study case. The red arrow points to the landslide scars.

For the second study case, 49 NDVI images to create the time-series were used. They were acquired from Landsat-7 and Landsat-8 satellites, from May 2013 to November 2015. Two experiments using the Random Forest algorithm were performed. The first one considered only the pixels time-

series as features, while the second used not only the pixels time-series, but also basic and polar metrics extracted from them. The analyzed classes include agriculture, constructed area, forest, mining, mining waste dam, grassland, planted forest and water. The dataset used in both classifications was composed of 378 pixel, obtained using high-resolution images from Google Earth, as suggested by [16]. The dataset was equally divided for training and testing purposes respecting the proportion 50/50. The classifications, Fig. 4, using the time-series achieved an overall accuracy (OA) of 75.66%, while the experiment that used also the basic and polar metrics achieved an OA of 91.53%. Therefore, the use of the metrics resulted in a 15% increase in OA.



(a) Classifications obtained using only time-series.



(b) Classifications obtained using time-series with time metrics.

**Fig. 4.** Classifications obtained using the metrics extracted with *stmetrics*.

#### 4. CONCLUSIONS

This paper introduced the *stmetrics* package, available on GitHub and developed for the extraction of basic and polar features from SITS. Previously, the extraction of the polar features was available only via C++ implementation on GeoDMA software [5]. Thus, the development of this package in python language provides a simple way to use these metrics, specially if the users do not have strong programming skills.

The use of the package and the available metrics improved the performance of the identification of landslide risk areas as well as the classification of LULC. However, any other task that uses time series approaches (e.g. land use and spatial dynamics modeling) can test the extraction of basic and polar features through *stmetrics* package to improve their results.

The best classification result derives from the use of the time-series values along with the metric features. Despite the great improvement on the performance, it is important to point out that the best results will be achieved by using a combination of time-series values and the time-series features that are obtained by time metrics. For the next version of the package, phenological metrics to derive the timing of the agriculture calendar will be added, as well as enhancements to improve the package overall performance.

#### 5. REFERENCES

- [1] François Petitjean, Camille Kurtz, Nicolas Passat, and Pierre Gançarski, "Spatio-temporal reasoning for the classification of satellite image time series," *Pattern Recognition Letters*, vol. 33, no. 13, pp. 1805–1815, 2012.
- [2] Gillian L Galford, John F Mustard, Jerry Melillo, Aline Gendrin, Carlos C Cerri, and Carlos EP Cerri, "Wavelet analysis of modis time series to detect expansion and intensification of row-crop agriculture in brazil," *Remote sensing of environment*, vol. 112, no. 2, pp. 576–587, 2008.
- [3] Daniel T Larose and Chantal D Larose, *Discovering knowledge in data: an introduction to data mining*, John Wiley & Sons, 2014.
- [4] Alana Neves, Hugo Bendini, Thales Korting, and Leila Fonseca, "Combining time series features and data mining to detect land cover patterns: a case study in northern mato grosso state, brazil," in *GeoInfo*, 2015.
- [5] Thales Körting, Leila Fonseca, and Gilberto Camara, "Geodma—geographic data mining analyst," *Computers Geosciences*, vol. 57, pp. 133 – 145, 2013.
- [6] Yaozhong Pan, Le Li, Jinshui Zhang, Shunlin Liang, Xiufang Zhu, and Damien Sulla-Menashe, "Winter wheat area estimation from modis-evi time series data using the crop proportion phenology index," *Remote Sensing of Environment*, vol. 119, pp. 232 – 242, 2012.
- [7] Yanfei Wei, Xinhua Tong, Gang Chen, Deqiang Liu, and Zhenfeng Han, "Remote detection of large-area crop types: The role of plant phenology and topography," *Agriculture*, vol. 9, no. 7, pp. 150, Jul 2019.
- [8] H. Bendini, I. D. Sanches, T. S. Körting, L. M. G. Fonseca, A. J. B. Luiz, and A. R. Formaggio, "Using Landsat 8 Image Time Series for Crop Mapping in a Region of Cerrado, Brazil," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41B8, pp. 845–850, Jun 2016.
- [9] Hugo Bendini, Leila Fonseca, Marcel Schwieder, Thales Körting, Philippe Rufin, Ieda, Pedro J. Leitão, and Patrick Hostert, "Detailed agricultural land classification in the brazilian cerrado based on phenological information from dense satellite image time series," *International Journal of Applied Earth Observation and Geoinformation*, vol. 82, pp. 101872, Oct. 2019.
- [10] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: A structure for efficient numerical computation," *Computing in Science Engineering*, vol. 13, no. 2, pp. 22–30, March 2011.
- [11] Eric Jones, Travis Oliphant, and Pearu Peterson, "Scipy: Open source scientific tools for python," 01 2001.
- [12] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science Engineering*, vol. 9, no. 3, pp. 90–95, May 2007.
- [13] Sean Gillies et al., "Shapely: manipulation and analysis of geometric objects," 2007-.
- [14] Sean Gillies, Katy Barnhart, and Sven Willner, "Descartes," 2019.
- [15] R. Edsall, Alan MacEachren, and Donna Peuquet, "Assessing the effectiveness of temporal legends in environmental visualization," in *Proceedings GIS-LIS '97*, 1997, pp. 677–685.
- [16] Pontus Olofsson, Giles M Foody, Martin Herold, Stephen V Stehman, Curtis E Woodcock, and Michael A Wulder, "Good practices for estimating area and assessing accuracy of land change," *Remote Sensing of Environment*, vol. 148, pp. 42–57, 2014.