Int J Appl Earth Obs Geoinformation 92 (2020) 102150

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Short communication

# Assessing the optimal preprocessing steps of MODIS time series to map cropping systems in Mato Grosso, Brazil

Patrick Calvano Kuchler[a,b,*], Agnès Bégué[a,c], Margareth Simões[b,d], Raffaele Gaetano[a,c], Damien Arvor[e], Rodrigo P.D. Ferraz[d]

[a] CIRAD, UMR TETIS, 34093 Montpellier, France
[b] UERJ/FEN/DESC/PPGMA, Rio de Janeiro 20559-900, Brazil
[c] TETIS, Univ Montpellier, CIRAD, INRAE, AgroParisTech, CNRS, 34093 Montpellier, France
[d] EMBRAPA Solos, Rio de Janeiro 22460-000, Brazil
[e] CNRS, UMR LETG 6554, Univ Rennes, 35043 Rennes, France

ARTICLE INFO

*Keywords:*
Mato Grosso
Integrated systems
Classification
Phenometrics
Smoothing
Agricultural intensification
TIMESAT

ABSTRACT

The adoption of new cropping practices such as integrated Crop-Livestock systems (iCL) aims at improving the land use sustainability of the agricultural sector in the Brazilian Amazon. The emergence of such integrated systems, based on crop and pasture rotations over and within years, challenges the remote sensing community who needs to implement accurate and efficient methods to process satellite image time series (SITS) in order to come up with a monitoring protocol. These methods generally include a SITS preprocessing step which can be time consuming. The aim of this study is to assess the importance of preprocessing operations such as temporal smoothing and computation of phenological metrics on the mapping of main cropping systems (i.e. pasture, single cropping, double cropping and iCL), with a special emphasis on the iCL class. The study area is located in the state of Mato Grosso, an important producer of agriculture commodities located in the Southern Brazilian Amazon. SITS were composed of a set of 16-day composites of MODIS Vegetation Indices (MOD13Q1 product) covering a one year period between 2014 and 2015. Two widely used classifiers, i.e. Random Forest (RF) and Support Vector Machine (SVM), were tested using five data sets issued from a same SITS but with different preprocessing levels: (i) raw NDVI; (ii) raw NDVI + raw EVI; (iii) smoothed NDVI; (iv) NDVI-derived phenometrics; (v) raw NDVI + phenometrics. Both RF and SVM classification results showed that the "raw NDVI + raw EVI" data set achieved the highest performance (RF OA = 0.96, RF Kappa = 0.94, SVM OA = 0.95, SVM Kappa = 0.93), followed closely by the "raw NDVI" and the "raw NDVI + phenometrics" datasets. The "NDVI-derived phenometrics" alone achieved the lowest accuracies (RF OA = 0.58 and SVM OA = 0.66). Considering that the implementation of preprocessing steps is computationally expensive and does not provide significant gains in terms of classification accuracy, we recommend to use raw vegetation indices for mapping cropping practices in Mato Grosso, including the integrated Crop-Livestock systems.

## 1. Introduction

The implementation of an intensive agricultural development model has led Brazil to play a significant role in the global production for food and energy (Arvor et al., 2012). However, this evolution has also long been criticized for its dramatic environmental impacts, especially regarding deforestation of Amazon rainforests and Brazilian Savannas (Silva and Lima, 2018). In an effort to support a more environmentally responsible agricultural model, the (former) Brazilian government adopted the Low Carbon Agriculture (or ABC for Agricultura de Baixo Carbono) Program to promote the adoption of conservation agricultural practices such as integrated systems, and in particular the integrated Crop-Civestock systems (iCL). Indeed, managing the land via iCL versus specialized systems (continuous cropland or pasture) involves many utility tradeoffs (Gil et al., 2016), in particular the increase of organic matter contained in the soil and an improved biomass production (Carvalho et al., 2014). Monitoring agricultural dynamics related to the generalization of integrated systems thus appears essential to assess the evolution of the Brazilian agricultural sector toward enhanced land use sustainability (Galford et al., 2013).

Whereas remote sensing has long been proven to be effective to map cropping systems (Bégué et al., 2018), most studies led in the Brazilian

* Corresponding author at: UERJ/FEN/DESC/PPGMA, Rio de Janeiro 20559-900, Brazil.
*E-mail address:* geocalvano@gmail.com (P.C. Kuchler).

https://doi.org/10.1016/j.jag.2020.102150
Received 17 October 2019; Received in revised form 21 April 2020; Accepted 30 April 2020
Available online 04 June 2020

Amazon, especially in the state of Mato Grosso, focused on the discrimination of single-cropping (SC) and double-cropping (DC) systems (Galford et al., 2008; Arvor et al., 2011, 2012; Brown et al., 2013; Spera et al., 2014; Zhu et al., 2015; Kastens et al., 2017; Manabe et al., 2018). These studies are based on the analysis of intra-annual spectral series that inform on the average phenology of individual land cover types (Gómez et al., 2016). Because of the large size of the agricultural fields, studies in Mato Grosso make generally use of satellite image time series (SITS) of MODIS vegetation indices at 250 m resolution (Galford et al., 2008; Arvor et al., 2011, 2012; Brown et al., 2013; Spera et al., 2014; Zhu et al., 2015; Kastens et al., 2017; Manabe et al., 2018; Chen et al., 2018; Picoli et al., 2018). These image data are adapted to monitor regional-scale dynamics at a high temporal frequency (1–2 days) that enables to address the high cloud cover rate issue. Unlike the single and double cropping systems, integrated Crop-Livestock systems are based on crop and pasture rotations over and within years (Gil et al., 2016). When the rotation is intra-annual, the radiometric temporal pattern of the iCL systems is close to the one of the DC, challenging the remote sensing community to map these systems in order to build a governmental protocol for monitoring iCL at regional scale (Manabe et al., 2018; Chen et al., 2018).

To map cropping systems from SITS, a wide variety of classification methods have been implemented using different input data sets. For example, some studies used complete time series of one vegetation index (e.g. NDVI or EVI), usually after running a temporal smoothing preprocessing algorithm to remove noise (generally related to atmospheric condition) from raw SITS especially in areas with significant cloud cover as it is the case in the Amazon region (Shao et al., 2016; Hird and McDermid, 2009). Other studies emphasized the potential of phenometrics (e.g. start, end and duration of season) derived from SITS to discriminate land use classes (Chen et al., 2018; Xu et al., 2017; Pan et al., 2015). Nonetheless, Picoli et al. (2018) recently considered that advanced classification algorithms are able to deal with incomplete or noisy data and thus focused on classifying high dimensional models including all the raw available data (EVI and NDVI indices coupled with NIR and MIR bands).

Since the preprocessing of SITS (i.e. temporal smoothing and computing phenological metrics) is a time-consuming and laborious task, it is worth questioning how it affects (positively or negatively) the classification results, especially when using advanced classifiers based on machine learning methods. Hence, the objective of this paper is to assess the importance of preprocessing MODIS time series to accurately classify cropping systems in the Southern Amazon, considering not only single and double-cropping systems but also more complex systems based on crop and livestock integration. To this end, we derived five datasets with different preprocessing levels from a single MODIS time series on a test area in the Brazilian state of Mato Grosso and we assessed their effectiveness to accurately map cropping practices using two state-of-the-art machine learning classifiers.

## 2. Material and methods

### 2.1. Study area

In the southern Brazilian Amazon (Fig. 1), croplands are expanding rapidly, and livestock production systems are highly extensive, turning Mato Grosso State as the main national producer for cattle and soybean. Considering that Mato Grosso is covered by three important biomes, i.e. Pantanal, Cerrado and Amazon, the expansionist agricultural model has often been considered for its impacts on environmental resources, especially rainforests (Arvor et al., 2017). In this study, we focused our analysis on a study area located in the north of Mato Grosso, within the so-called “Amazonian Arc of deforestation” (Fig. 1). In this area historically dedicated to cattle ranching, farmers are benefiting from public policies such as the “Agricultura de Baixo Carbono (ABC) program” to adopt new practices including integrated systems in order to improve land use sustainability. The most common integrated system is the crop and livestock integration (iCL) representing around 89% of the integrated systems in Mato Grosso (Gil et al., 2016). iCL is based either on sequential cropping, i.e. growing crops and pasture consecutively in the same cropping season (intra-annual iCL), or on crop succession, i.e. rotation of consecutive cropping seasons including pasture (inter-annual iCL).

### 2.2. Data

We acquired a complete MODIS time series of vegetation indices for the time period running from August 2014 to July 2015. The two vegetation indices considered were the Normalized Difference Vegetation Index (NDVI) and the Enhanced Vegetation Index (EVI) at 250 m spatial resolution and 16-day temporal resolution from the MOD13Q1 product (Huete et al., 2002). Ground reference data were collected for the 2014–2015 agricultural cycle through farmers’ interviews. This reference data set contains information on four major cropping systems observed in the study area for 77 fields in 5 farms (summing 778 MODIS pixels; see Table 1 for the details). The four systems are: Pasture (P), mainly *Brachiaria* sp.; intra-annual iCL, represented by soybean followed by *Brachiaria* sp.; Single Cropping (SC), represented by a main crop (soybean, maize or rice) followed by a fallow period; Double Cropping (DC), represented by soybean as a main crop and followed by maize or millet. The agriculture production system at Mato Grosso state is characterized by large farms and parcels, generally bigger than 25 ha. The ground data used in this work was collected on fields larger than 12.8 ha, with a mean size of the sampled fields of 80.8 ha. However, in order to minimize the edge effect in mixed pixels, only pixels with more than 50 % contained in the crop plots were considered.

In order to assess the influence of preprocessing steps to discriminate cropping systems, the original MODIS time series was used to derive five data sets corresponding to different preprocessing levels:

- R-NDVI: raw NDVI time series (23 total variables, the per-date NDVI values);
- R-NDVI + R-EVI: combination of raw NDVI and raw EVI time series (46 total variables, 23 NDVI + 23 EVI values);
- S-NDVI: smoothed NDVI time series (23 total variables as R-NDVI);
- P-NDVI: NDVI phenometrics derived from the raw NDVI time series (11 variables);
- R-NDVI + P-NDVI: combination of raw NDVI time series and NDVI phenometrics (34 total variables, 23 NDVI + 11 phenometrics).

Both the S-NDVI and P-NDVI data sets were preprocessed using the TIMESAT 3.2 software (Jönsson and Eklundh, 2004). With regards to the S-NDVI, various methods to smooth time series of MODIS vegetation indices have been tested and compared in the literature (Shao et al., 2016; Atkinson et al., 2012; Hird and McDermid, 2009). All emphasized the importance of choosing the relevant method depending on the user's objectives and nature of noise in the time series. After some tests and based on previous studies carried out in Mato Grosso (Arvor et al., 2008) and Tocantins, a neighboring Brazilian state (Bellón et al., 2017), we decided to use the Savitzky-Golay (SG) (Savitzky and Golay, 1964; Chen et al., 2004; Cao et al., 2018) method. It is worth noting that TIMESAT software allows weighting the smoothing process using a quality band provided with the MOD13Q1 SITS. However, we decided to perform a non-weighted smoothing because the reliability of the MODIS quality band in our study area characterized by high cloud cover rates has been severely criticized (Bellón et al., 2017). We applied the SG algorithm with the following basic parameters: $d = 2$ ($d$ refers to the degree of the filtering polynomial), and $m = 2$ ($m$ is the half-width of the filtering window). The smoothed time series were then analyzed by TIMESAT to extract eleven parameters that make up the P-NDVI data set. The eleven parameters refer to the start, middle, length and end of season, base value, peak value, amplitude, right derivative, left

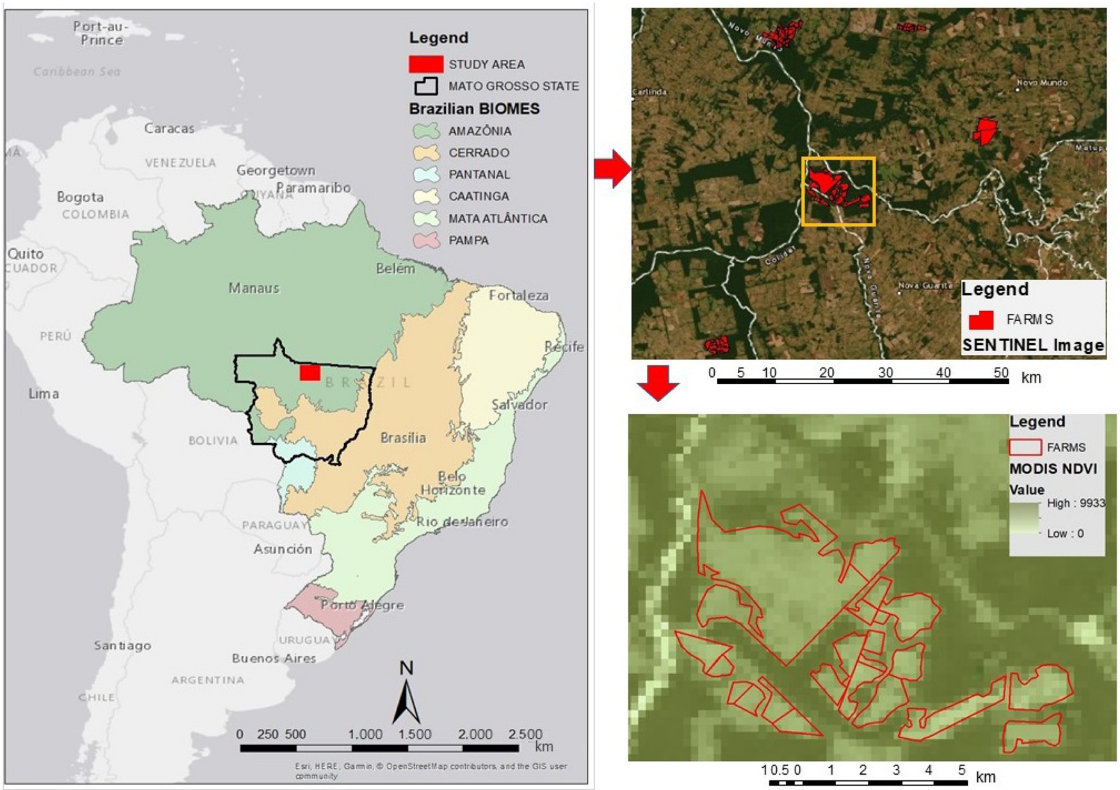


**Fig. 1.** Location of Mato Grosso state and of the study area, location of the sampled farms, and a zoom of a full resolution MODIS NDVI image (09/22/2014) showing sampled fields.

**Table 1**
Accuracies (OA = Overall Accuracy and K = Kappa) achieved for the different classifiers (RF = Random Forest and SVM = Support Vector Machine) and F-scores applied in RF to discriminate cropping practices (P = Pasture, SC = Single Cropping, DC = Double Cropping and iCL = integrated Crop-Livestock system).

| Preprocessing level | RF | | | | | | SVM | |
|---|---|---|---|---|---|---|---|---|
| | F-score | | | | OA | K | OA | K |
| | P | iCL | SC | DC | | | | |
| R-NDVI | 0.95 | 0.92 | 0.94 | 0.91 | 0.94 | 0.91 | 0.91 | 0.88 |
| R-NDVI + R-EVI | 0.97 | 0.95 | 0.97 | 0.96 | 0.96 | 0.94 | 0.95 | 0.93 |
| S-NDVI | 0.90 | 0.90 | 0.86 | 0.87 | 0.84 | 0.84 | 0.83 | 0.77 |
| P-NDVI | 0.58 | 0.64 | 0.41 | 0.58 | 0.58 | 0.42 | 0.66 | 0.57 |
| R-NDVI + P-NDVI | 0.97 | 0.93 | 0.97 | 0.92 | 0.95 | 0.92 | 0.94 | 0.93 |
| Number of samples | 177 | 293 | 107 | 201 | | | | |

derivative, and large and small integrals.

The potential of these five datasets to discriminate cropping systems was then tested using advanced statistical learning classifiers, which are particularly relevant to deal with high dimensional data characterized by strong cross-correlations and redundancies, as well as potentially uninformative components (James et al., 2013). In this regard, two classifiers were tested: Random Forest (RF) (Belgiu and Drăguţ, 2016) with the *randomForest* Package (Liaw, 2018), and Support Vector Machine (SVM) (Mountrakis et al., 2011) with the *caret* Package (Kuhn, 2015) of the R environment (R Core Team, 2019).

Finally, the classification accuracy was evaluated using a 5-fold cross-validation procedure (Wiens et al., 2008) using 80% of samples for training (622 pixels) and 20% for prediction (156 pixels). Each fold generates one confusion matrix. The k-fold cross-validation technique is a common approach that allows (i) using all the data set, which is particularly relevant when the dataset is small, and (ii) being more confident with the algorithm performance and data quality (Dima, 2018). We also decided to perform an analysis of classification accuracies using importance-based variable selection in order to assess the impact of the different preprocessing steps in the redistribution of the discriminant information carried by the time series. Matter of facts, here we are dealing with high-dimensional data characterized by strong cross-correlations and redundancies,as well as potentially uninformative components, which makes the use of advanced statistical learning methods such as RF particularly pertinent (James et al., 2013). To this aim, the mean decrease in accuracy (MDA) index based on Random Forest was calculated to evaluate the importance of each variable in terms of class discrimination (Cutler et al., 2007). The results of MDA indicate the number of important variables optimizing and reducing the number of input variables for each dataset (Lebourgeois et al., 2017). This process was performed using the *rfe* function available in *caret* package (Kuhn, 2015).

## 3. Results

Fig. 2(a) shows the R-NDVI time series of each cropping system, calculated as the mean of the NDVI values over the reference pixels (Table 1). The visual analysis shows that pastures (P) are represented by a long and flat time series while single cropping (SC) sytems display a short cycle with high maximum values. In contrast, the double cropping (DC) and the integrated Crop-Livestock (iCL) systems are both characterized by two phenological cycles apparently similar at first glance. However, the two series differ (i) at the beginning of the season (August–October) with higher NDVI values for iCL, and (ii) at the end of the season with a longer second cycle for iCL than for DC because pasture remain greener long after the harvest of maize. Fig. 2(b) presents an example of S-NDVI time series for a single pixel. Peaks and valleys due to atmospheric noise characterizing the R-NDVI profile are smoothed, thus facilitating the visual interpretation of phenological cycles and the

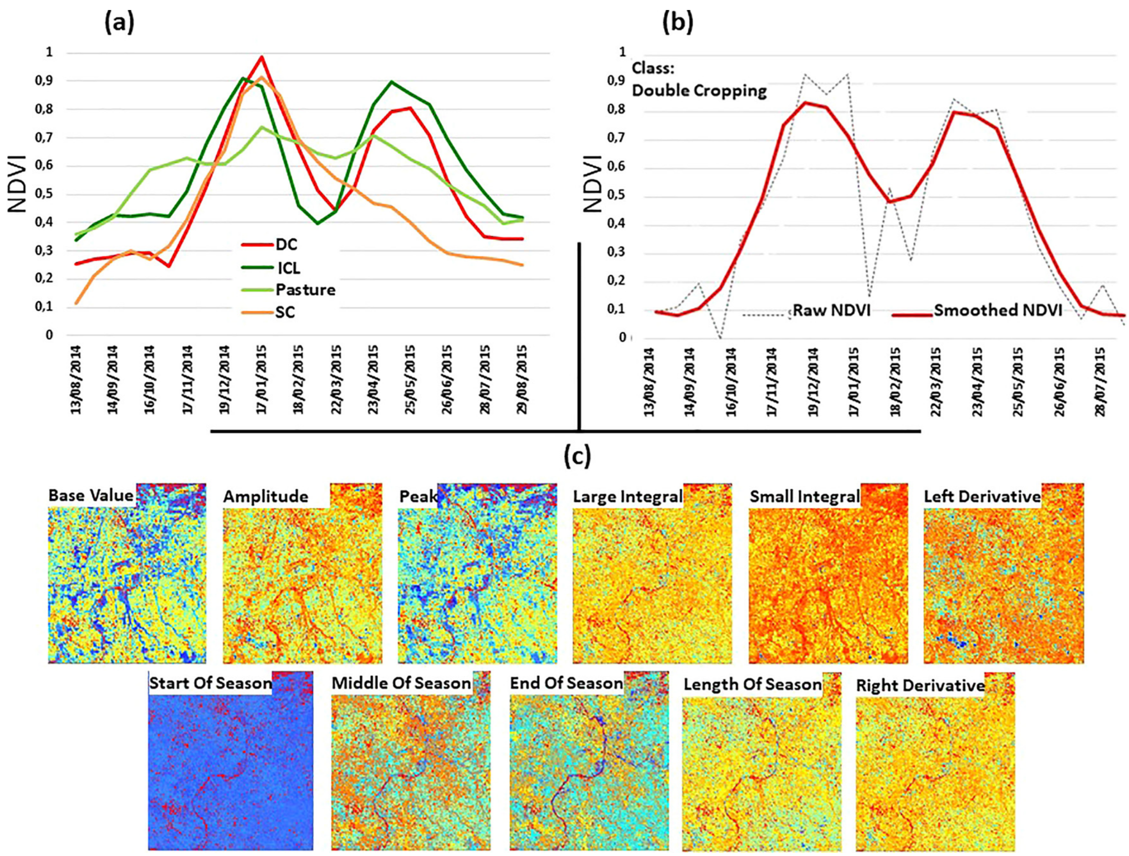


**Fig. 2.** (a) Annual time series of the mean MODIS NDVI calculated for each cropping system class; (b) example of MODIS NDVI time series acquired over a DC pixel, before and after filtering using Savistky-Golay algorithm; (c) Raster data of each 11 phenometrics calculated by TIMESAT.

computation of phenometrics (P-NDVI, Fig. 2(c)).

Fig. 3 shows the evolution of OA according to the number of variables used in the classification model (the variables were included in the model one by one, according to their order of importance based on the MDA measure). As arguable, the best overall accuracies are reached using the datasets with higher dimensionality (R-NDVI + R-EVI and R-NDVI + P-NDVI). For datasets using only NDVI, the one using raw values (R-NDVI) largely outperforms the one using the smoothed time series (S-NDVI). The sole phenological metrics do not seem to provide sufficient discriminative potential for classification, as the

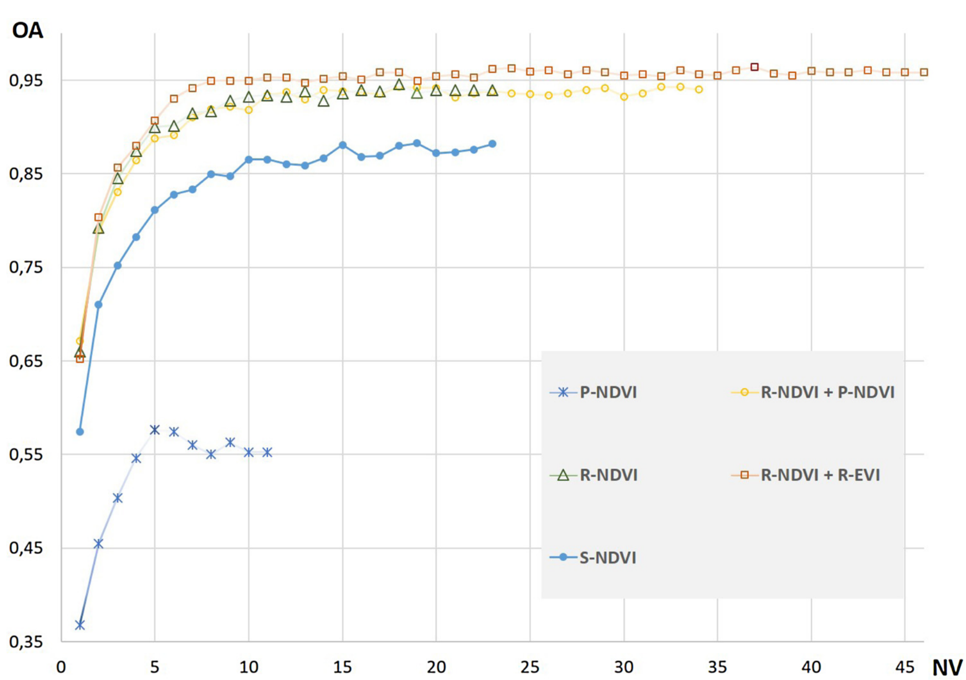


**Fig. 3.** Cross-validated overall accuracies in function of the number of variables.

corresponding dataset achieves the worst performances. Yet, a slight improvement is observed when they are used in combination with the raw NDVI. Summarizing, conclusions drawn in Hastie (2009) are confirmed in this case study since statistical learning methods such as Random Forest, designed to separate noise from relevant information, are more likely to succeed over large sets of “raw” variables than over preprocessed data.

Concerning the importance-based analysis, the R-NDVI + R-EVI dataset is the most efficient as it exhibits the best “OA vs. Number of important variables” combination with an asymptotic high OA value reached with less than 10 variables, while R-NDVI and R-NDVI + P-NDVI reach the asymptotic high OA value with a larger number of variables.

Table 1 provides the F-score for each class of cropping system as well as the overall accuracy (OA) and kappa index (K) produced by RF and SVM classifiers. As there was no significant difference between the results of the RF and SVM classifiers, we introduce only the RF results. The RF classification results showed that the R-NDVI + R-EVI dataset achieved the highest performance (OA = 0.96, K = 0.94 and an average F-score = 0.96), followed closely by the R-NDVI and the R-NDVI + P-NDVI datasets. The P-NDVI alone obtained the lowest accuracies (K = 0.58).

## 4. Discussion and conclusion

In this study, we assessed the importance of preprocessing steps of MODIS VI satellite image time series for mapping the major cropping systems in a northern region of the Brazilian Amazon state of Mato Grosso. Our results show that the temporal smoothing of raw VI time series and the computation of phenometrics tend to decrease the classification accuracy when compared with results obtained by classifying time series of raw vegetation indices. We found that modern statistical models such as Support Vector Machines and Random Forests are robust to noise and able to achieve better results when using raw datasets. This result is in accordance with Chen et al. (2018) who reported a concern in the use of Savitsky-Golay processing, which, on the one hand, reduces atmospheric interference in the time series, but, on the other hand, filters important information, thus hindering the discrimination of classes with confusing time series such as DC and iCL systems. This is even more significant when using a Random Forest classifier, which aims at identifying the most discriminant variables without explicitly handling cross-correlations: when using smoothed NDVI profiles, discriminant information originally concentrated in a single timestamp is spanned over neighboring timestamps as a result of the smoothing process (e.g., a polynomial interpolation for SG), and potentially corrupted by noisy acquisitions, which in our case are particularly recurrent due to the significant cloud coverage. A measure of this effect can be observed in Fig. 3 where the four most important S-NDVI variables are needed to approach the classification accuracy of the first two R-NDVI variables. The flattened discriminative potential of S-NDVI is confirmed by the poorer results obtained using SVM, and obviously impacts the performances of phenometrics, which are directly derived from this source. Nguyen and Henebry (2019) also describe that land surface phenology modeling requires a substantial number of good quality observations over a year, it may be less suitable for areas with persistent cloud cover if only optical data are available to characterize the land surface phenology; thus, Our results also corroborate with the results obtained by Picoli et al. (2018), who used NIR and MIR reflectances in addition to R-NDVI and R-EVI. In our case, the addition of R-EVI was sufficient to increase the accuracy of iCL mapping, confirming that the capacity of Random Forests to perform better in higher-dimensional spaces is suited for this case study.

It is worth noting that we only focused on SVM and RF classifiers because of their large use in remote sensing-based land use classifications. However, additional classification techniques based on a curve similarity approach, e.g. Dynamic Time Warping (DTW) (Mondal and Jeganathan, 2018) and Time-Weighted Dynamic Time Warping (TWDTW) (Maus et al., 2016), should be tested, especially because these methods are more dependent on the temporal smoothing. This delineates a first direction for future work, together with a deeper insight on the different smoothing techniques. Other future works shall concern the testing of the approach over a larger area (whole state of Mato Grosso) for a more definitive assessment of the classification strategy to map integrated systems.

To conclude, our results showed that raw time series of MOD13Q1 product are efficient for mapping cropping systems in regions with large-scale agriculture. This study can be used as a basis for establishing a regional monitoring protocol of the agricultural integrated systems in the Amazon and Cerrado biomes. However, before implementing this image processing chain for an operational service, additional ground data should be collected over the entire state of Mato Grosso, and for different agricultural years (e.g. with different climatic conditions) in order to determine if the present results would be affected by a larger intra-class variability.

## Author contributions

P.C. Kuchler: Conception and design of study, acquisition of data, analysis and/or interpretation of data, drafting the manuscript. A. Bégué: Conception and design of study, analysis and/or interpretation of data, drafting the manuscript, revising the manuscript critically for important intellectual content. M. Simões: Acquisition of data, drafting the manuscript. R. Gaetano: Analysis and/or interpretation of data, revising the manuscript critically for important intellectual content. D. Arvor: Analysis and/or interpretation of data, drafting the manuscript, revising the manuscript critically for important intellectual content. R.P.D. Ferraz: Acquisition of data.

## Conflict of interest

The authors declare that there is no conflict of interest.

## Acknowledgements

The main author received a scholar fellowship from the Capes-Cofecub GeoABC Project(Methodologies and technological innovation for satellite monitoring of low carbon agriculture in support to Brazil's ABC Plan, project No. 845/15). We thank the CIRAD-TETIS Unit for the welcoming stay in Montpellier and methodological support, the H2020-MSCA-RISE-2015 ODYSSEA European project (Project Reference: 691053) for supporting the ground visits, and the teams from EMBRAPA Agrosilvipastoril and Embrapa Labex Europe and the IEA CNRS SCOLTEL project.

## References

Arvor, D., jonathan, M., Meirelles, M.S.P., Dubreuil, V., Lecerf, R., 2008. Comparison of multitemporal MODIS-EVI smoothing algorithms and its contribution to crop monitoring. 2008 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2008, IEEE.

Arvor, D., Jonathan, M., Meirelles, M.S.P., Dubreuil, V., Durieux, L., 2011. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. Int. J. Remote Sens. 32, 7847–7871.

Arvor, D., Meirelles, M., Dubreuil, V., Bégué, A., Shimabukuro, Y.E., 2012. Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. Appl. Geogr. 32, 702–713.

Arvor, D., Tritsch, I., Barcellos, C., Jégou, N., Dubreuil, V., 2017. Land use sustainability on the south-eastern amazon agricultural frontier: recent progress and the challenges ahead. Appl. Geogr. 80, 86–97.

Atkinson, P.M., Jeganathan, C., Dash, J., Atzberger, C., 2012. Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. Remote Sens. Environ. 123, 400–417.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., de Abelleyra, D., Ferraz, R.P.D., Lebourgeois, V., Lelong, C., Sim oes, M., Verón, S.R., 2018. Remote sensing and cropping practices: a review. Remote Sens. 10, 99.

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: a review of applications

and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31.
Bellón, B., Bégué, A., Seen, D.L., de Almeida, C., Sim oes, M., 2017. A remote sensing approach for regional-scale mapping of agricultural land-use systems based on NDVI time series. Remote Sens. 9, 600.
Brown, J.C., Kastens, J.H., Coutinho, A.C., de Castro Victoria, D., Bishop, C.R., 2013. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. Remote Sens. Environ. 130, 39–50.
Cao, R., Chen, Y., Shen, M., Chen, J., Zhou, J., Wang, C., Yang, W., 2018. A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. Remote Sens. Environ. 217, 244–257.
Carvalho, J.L.N., Raucci, G.S., Fraz ao, L.A., Cerri, C.E.P., Bernoux, M., Cerri, C.C., 2014. Crop-pasture rotation: a strategy to reduce soil greenhouse gas emissions in the Brazilian Cerrado. Agric. Ecosyst. Environ. 183, 167–175.
Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. Remote Sens. Environ. 91, 332–344.
Chen, Y., Lu, D., Moran, E., Batistella, M., Dutra, L.V., Sanches, I.D., da Silva, R.F.B., Huang, J., Luiz, A.J.B., de Oliveira, M.A.F., 2018. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. Int. J. Appl. Earth Obs. Geoinf. 69, 133–147.
Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.
Dima, S., 2018. 5 Reasons Why You Should Use Cross-Validation in Your Data Science Projects. (Accessed 17 April 2020).
Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogramm. Remote Sens. 116, 55–72.
Galford, G.L., Mustard, J.F., Melillo, J., Gendrin, A., Cerri, C.C., Cerri, C.E., 2008. Wavelet analysis of MODIS time series to detect expansion and intensification of row-crop agriculture in Brazil. Remote Sens. Environ. 112, 576–587.
Galford, G.L., Soares-Filho, B., Cerri, C.E.P., 2013. Prospects for land-use sustainability on the agricultural frontier of the Brazilian Amazon. Philos. Trans. R. Soc. B: Biol. Sci. 368, 20120171.
Gil, J., Garrett, R., Berger, T., 2016. Determinants of crop-livestock integration in Brazil: evidence from the household and regional levels. Land Use Policy 59, 557–568.
Hastie, T., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics, second edition. Springer.
Hird, J.N., McDermid, G.J., 2009. Noise reduction of NDVI time series: an empirical comparison of selected techniques. Remote Sens. Environ. 113, 248–258.
Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.
Jönsson, P., Eklundh, L., 2004. TIMESAT – a program for analyzing time-series of satellite sensor data. Comput. Geosci. 30, 833–845.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R. Springer.
Kastens, J.H., Brown, J.C., Coutinho, A.C., Bishop, C.R., Esquerdo, J.C.D.M., 2017. Soy moratorium impacts on soybean and deforestation dynamics in Mato Grosso, Brazil. PLoS One 12, e0176168.
Kuhn, M., 2015. Caret: Classification and Regression Training, R Package Version 5.15-044.
Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., Bégué, A., 2017. A combined random forest and OBIA classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated sentinel-2 time series, VHRS and DEM). Remote Sens. 9, 259.
Liaw, A., 2018. randomForest: Breiman and Cutler's Random Forests for Classification and Regression. R Package Version 4.6.14.
Manabe, V., Melo, M., Rocha, J., 2018. Framework for mapping integrated crop-livestock systems in Mato Grosso, Brazil. Remote Sens. 10, 1322.
Maus, V., Camara, G., Cartaxo, R., Sanchez, A., Ramos, F.M., de Queiroz, G.R., 2016. A time-weighted dynamic time warping method for land-use and land-cover mapping. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 3729–3739.
Mondal, S., Jeganathan, C., 2018. Mountain agriculture extraction from time-series MODIS NDVI using dynamic time warping technique. Int. J. Remote Sens. 39, 3679–3704.
Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259.
Nguyen, L.H., Henebry, G.M., 2019. Characterizing land use/land cover using multi-sensor time series from the perspective of land surface phenology. Remote Sens. 11, 1677.
Pan, Z., Huang, J., Zhou, Q., Wang, L., Cheng, Y., Zhang, H., Blackburn, G.A., Yan, J., Liu, J., 2015. Mapping crop phenology using NDVI time-series derived from HJ-1 a/b data. Int. J. Appl. Earth Obs. Geoinf. 34, 188–197.
Picoli, M.C.A., Camara, G., Sanches, I., Sim oes, R., Carvalho, A., Maciel, A., Coutinho, A., Esquerdo, J., Antunes, J., Begotti, R.A., Arvor, D., Almeida, C., 2018. Big earth observation time series analysis for monitoring Brazilian agriculture. ISPRS J. Photogramm. Remote Sens. 145, 328–339.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Austria, Vienna.
Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.
Shao, Y., Lunetta, R.S., Wheeler, B., Iiames, J.S., Campbell, J.B., 2016. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. Remote Sens. Environ. 174, 258–265.
Silva, C., Lima, M., 2018. Soy moratorium in Mato Grosso: deforestation undermines the agreement. Land Use Policy 71, 540–542.
Spera, S.A., Cohn, A.S., VanWey, L.K., Mustard, J.F., Rudorff, B.F., Risso, J., Adami, M., 2014. Recent cropping frequency, expansion, and abandonment in Mato Grosso, Brazil had selective land characteristics. Environ. Res. Lett. 9, 64010.
Wiens, T.S., Dale, B.C., Boyce, M.S., Kershaw, G.P., 2008. Three way k-fold cross-validation of resource selection functions. Ecol. Model. 212, 244–255.
Xu, X., Conrad, C., Doktor, D., 2017. Optimising phenological metrics extraction for different crop types in Germany using the moderate resolution imaging spectrometer (MODIS). Remote Sens. 9, 254.
Zhu, C., Lu, D., Victoria, D., Dutra, L., 2015. Mapping fractional cropland distribution in Mato Grosso, Brazil using time series MODIS enhanced vegetation index and Landsat thematic mapper data. Remote Sens. 8, 22.