Proceedings XX GEOINFO, November 11-13, 2019, São José dos Campos, SP, Brazil. p 135-146

Exploiting parallelism to generate meta-features for land use and land cover classification with remote sensing time series

Sávio S. T. de Oliveira[2], Marcelo de C. Cardoso[2], Elivelton Bueno[2], Vagner J. S. Rodrigues[2], Wellington S. Martins[1]

[1]Instituto de Informática - Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Quadra D, Câmpus Samambaia
131 - CEP 74001-970 - Goiânia - GO - Brasil

[2]goGeo
Av. 136, 638 - St. Marista, Goiânia - GO, 74180-040

{savio.teles, marcelo.cardoso, elivelton.bueno, vagner}@gogeo.io, wellington@inf.ufg.br

Abstract. *The automatic classification of remote sensing time series has become essential to identify the rapid and frequent changes that the earth's surface has been undergoing. This work investigates the accuracy of land use and land cover classification with remote sensing time series when distance based meta-features are added to existing features of some classifiers. The distance based meta-features presented are generated by comparing all time series of the region being studied to every time series patterns previously calculated for that region. This is a very costly operation that was made viable through the use of parallel processing. Although expensive, this operation is advantageous because the meta-features generated can be later used as input to any classifier. The experimental work conducted showed promising results when using the distance based meta-feature strategy. The proposed strategy was able to increase from 78% to 93,8% the classification accuracy of the KNN algorithm, and from 92,3% to 93,8% the accuracy of a state-of-art SVM-based algorithm proposed recently. These results indicate that distance-based meta-features allow revealing unknown data characteristics, potentially increasing classification accuracy.*

1. Introduction

Never before in the current era has the Earth's surface changed so fast. While urban and agricultural areas greedily expand into the surrounding natural space, whole forest ecosystems are diminishing at an alarming speed. To identify those changes and highlight their dynamics, the automatic classification of remote sensing time series has become essential [Bégué et al. 2018].

The Time-Weighted Dynamic Time Warping (TWDTW)[Maus et al. 2016] algorithm with temporal weights, was successfully used to identify these changes in the Earth and classify land cover classes including single and double cropping systems [Maus et al. 2019]. However, when there is a high variability of data in each temporal pattern, the overlap of time series patterns increases, leading to confusion in the classification [DADI 2019]. The method proposed by [Picoli et al. 2018] minimizes this impact by using all spectral bands and vegetation indices from time series as input variables for machine learning algorithms.

Our work extends the solution proposed by [Picoli et al. 2018] adding meta-features created from distance measures calculated by TWDTW, for each pattern class. We obtained a higher overall accuracy than the TWDTW and the method proposed by [Picoli et al. 2018]. The TWDTW algorithm has a high computational cost, with time complexity of $O(n^2)$. Therefore, the method, introduced in this paper, explores parallel architectures to create the meta-features efficiently with the Spatial Parallel TWDTW (SP-TWDTW) algorithm [Oliveira et al. 2018], proposed in a previous work. The SP-TWDTW execution time was 246 times faster than the TWDTW in R, and 11 times faster than the TWDTW implementation in C++.

This paper is organized as follows. Section 2 discusses the solutions for classification using remote sensing time series. Section 3 describes the TWDTW algorithm used as the basis for this work. The new method proposed in this paper is presented in Section 4. Section 5 validates the method and discusses the accuracy of it. Finally, Section 6 presents some conclusions and future work.

2. Land cover and Land Use Classification using Remote Sensing Time Series

Automatic classification methods using remote sensing time series have been used for land cover and land use mapping. DTW is one of the most well-known methods in this field. Some previous work like [Petitjean et al. 2012, Petitjean and Weber 2014, Maus et al. 2016] proposed non parallel methods using DTW to analyze time series of satellite images using a maximum time delay to avoid time distortions based on the date of the satellite images.

Some methods process each image independently and compare the results for different time instances [Gómez et al. 2011, Lu et al. 2016]. The technique presented in [Costa et al. 2017] builds a time series of each pixel and process them independently. Some papers perform the time series classification through spatial interpolation [Li and Heap 2014], which is the process of using points with known values to estimate values of other unknown points. Several automated approaches were developed for land use classification, including single or multi-stage supervised classification [Abou EL-Magd and Tanton 2003], decision tree [Simonneaux et al. 2008], and supervised learning models such as Random Forest or Support Vector Machines [Löw et al. 2012, Lebourgeois et al. 2017].

Deep Learning (DL) algorithms have seen a massive rise in popularity for remote-sensing image classification over the past few years. A study was made to conduct a comprehensive review of more than 200 publications in this field, most of which were published during the last two years [Ma et al. 2019]. For time series classification, the Recurrent Neural Networks (RNN) has been traditionally applied, analyzing the observations of each pixel over time. Despite the potential of RNNs, it has been pointed out that there are open challenges in the classification using DL algorithms.

The method proposed by [Picoli et al. 2018] builds high-dimensional spaces using all values of the time series from vegetation indices NDVI and EVI, and the spectral bands NIR and MIR, coupled with advanced statistical learning methods. Three classifiers were evaluated: Support Vector Machine (SVM), Random Forest (RF), and Linear Discriminant Analysis (LDA). The SVM classifier has been show to have better discriminating power than RF and LDA in a case study covering the state of Mato Grosso, Brazil, in an area of 5,300 km^2. As an example, the Table 1 shows the time series from NDVI, EVI, NIR and MIR with two observations each, which are mapped as input variables for machine learning methods.

Table 1. Input variables for machine learning methods using NDVI, EVI, NIR and MIR time series with two observations each [Picoli et al. 2018].

pixel	mir.1	mir.2	nir.1	nir.2	ndvi.1	ndvi.2	evi.1	evi.2
1	0.5	0.1	0.7	0.1	0.11	0.11	0.1	0.01
2	0.8	0.85	0.2	0.1	0.3	0.5	0.4	0.52
3	0.6	0	0.8	0.3	0.3	0.3	0.6	0.23

3. Time-Weighted Dynamic Time Warping (TWDTW)

The TWDTW [Maus et al. 2016] is a variation of the DTW algorithm that is sensitive to seasonal changes of natural and cultivated vegetation types. It considers inter-anual climatic and seasonal variability. The TWDTW method computes the cost matrix $\Psi_{n,m}$ given the pattern $U = (u_1, ..., u_n)$ and time series $V = (v_1, ..., v_m)$. The elements $\psi_{i,j}$ of $\Psi_{n,m}$ are computed by adding the temporal cost ω, becoming $\psi_{i,j}$ = $|u_i - v_j| + \omega_{i,j}$, where $u_i \in U \ \forall \ i = 1, ..., n$ and $v_j \in V \ \forall \ j = 1, ..., m$. To calculate the time cost, the logistic model is used with a midpoint β and a bias α presented in Equation 1.

$$\omega_{i,j} = \frac{1}{1 + e^{-\alpha(g(t_i,t_j)-\beta)}}, \tag{1}$$

in which $g(t_i, t_j)$ is the elapsed time in days between dates t_i for the patterns U and t_j in the time series V. From the cost matrix Ψ an accumulated cost matrix is calculated, named D by using a recursive sum of the minimum distances, as shown in equation 2

$$d_{i,j} = \psi_{i,j} + min\{d_{i-1,j}, d_{i-1,j-1}, d_{i,j-1}\}, \tag{2}$$

which is subject to the following conditions:

$$d_{ij} = \begin{cases} \psi_{i,j} & i = 1, j = 1 \\ \sum_{k=1}^{i} \psi_{k,j} & 1 < i \leq n, j = 1 \\ \sum_{k=1}^{j} \psi_{i,k} & i = 1, 1 < j \leq m \end{cases} \tag{3}$$

The kth lowest cost path in D produces an alignment between the pattern and a subsequence of V with associated distance δ_k, in which a_k is the first element and b_k the last element of k. Each minimum point in the last row of the cost matrix is accumulated, i.e. $d_{n,j} \ \forall \ j = 1, ..., m$, produces an alignment, with $b_k = argmin_k(d_{n,j}), k = 1, ..., K$ and $\delta_k = d_{n,b_k}$, in which K is the minimum number of points in the last row of D.

A reverse algorithm, equation 4, maps the path $P_k = (p_1, ..., p_L)$ along the kth "valley" to the lowest cost in D. The algorithm starts in $p_{l=L} = (i = n, j = b_k)$ and ends with $i = 1$, i.e. $p_{l=1} = (i = 1, j = a_k)$, in which L denotes the last point of alignment. The path P_k contains the elements that have been matched between the series.

$$p_{l-1} = \begin{cases} (i, a_k = j) & \text{se } i = 1 \\ (i-1, j) & \text{se } j = 1 \\ argmin(d_{i-1,j}, d_{i-1,j-1}, d_{i,j-1}) & \text{otherwise} \end{cases} \tag{4}$$

The best accuracy results using TWDTW distance measures have been reached using the k-Nearest-Neighbours (kNN) algorithm with $K = 1$ (1NN) [Maus et al. 2016, Maus 2016, Wegner Maus et al. 2019]. The kNN is a non-parametric classification method that does not require training data to generate the model. Given a set of n training examples, upon receiving a new instance to predict, the kNN classifier will identify k nearest neighboring training examples of the new instance and then assign the class label hold by the majority of neighbors to the new instance. Remote sensing time series classification with 1NN using TWDTW distance measures follows the steps:

1. Initializes $K = 1$.
2. Calculate the distance measure between each time series V and each pattern U with TWDTW.
3. Sort the distances in ascending order based on its values.
4. Assign to V the class from pattern U with the shortest distance to V.

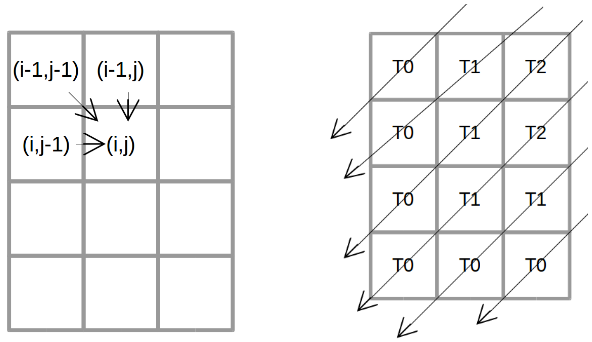
4. Creating Meta-Features for Land Use and Land Cover Classification using Remote Sensing Time Series exploiting Parallel Processing

The TWDTW is a pattern-matching algorithm based on dynamic programming with time complexity $O(n^2)$. This section presents the solution proposed to create meta-features as input for machine learning methods to increases the land use and land cover classification accuracy. These meta-features are created exploiting parallelism with the algorithm proposed in [Oliveira et al. 2018], denoted by SP-TWDTW.

In general, the SP-TWDTW algorithm parallelizes the construction of the accumulated cost matrix. The accumulated cost matrix D is computed from the cost matrix Ψ using the recursive sum of the minimum distances, as shown in equation 2. The construction of D can not be trivially paralleled since the computation of each element (i, j) of the matrix depends on the previously elements $(i-1, j)$, $(i, j-1)$ and $(i-1, j-1)$. This dependency can be seen in Figure 1(a). The idea behind the SP-TWDTW algorithm is presented in Figure 1(b). Each diagonal is computed in parallel, with each thread being responsible for a diagonal cell. Since the elements are not dependent on each other within the diagonal, the calculation of the accumulated cost does not lead to an inconsistent matrix.

4.1. Creating Meta-Features for Machine Learning Algorithms using the SP-TWDTW Distance Measures

In remote sensing time series processing, each pattern belongs to a certain class. The pattern is created using a set of time series sample of the given class. The best accuracy results using TWDTW similarity measures in STSR classification have been obtained using the KNN approach with K = 1 (1NN) [Maus et al. 2016], described in Section 3. However, for regions where data samples cannot capture the correct pattern for each class, TWDTW usually does not have high accuracy on classification [Maus et al. 2016]. Due



(a) The computation of each element in D depends on the values of previous elements.

(b) SP-TWDTW: Parallel processing of D

Figure 1. Computation of the accumulated cost matrix D

to the variability of time series samples, creating a single time series pattern for each class is a difficult task. The solution proposed by [Picoli et al. 2018], described in Section 2, reduces this impact by using all spectral bands and vegetation indices from time series as features for machine learning algorithms, creating high dimensional spaces.

Our work proposes a new method for land use and land cover mapping that creates meta-features from distance measures calculated by SP-TWDTW, in addition to spectral bands and vegetation indices [Picoli et al. 2018], as input variables for machine learning algorithms. Formally, given a time series V, a set of patterns $Q = \{U_1, U_2, ..., U_n\}$, the meta-features array is created using the function *F-TWDTW(V)* following the Equation 5:

$$\textit{F-TWDTW}(V) = (\textit{SP-TWDTW}(V, U_1), ..., \textit{SP-TWDTW}(V, U_n)) \tag{5}$$

where, n is the number of patterns in Q. Since each pattern is related to one class, so n also represents the number of classes, to which each V can be assigned. The meta-features created by *F-TWDTW(V)* is added to the array of features defined in [Picoli et al. 2018], generating the input variables for the machine learning algorithms.

The creation of meta-features with SP-TWDTW for machine learning methods can be exemplified as follows. Suppose there are three classes: “Forest”, “Pasture” and “Cerrado”. In this case, there is one pattern for each class and Q = {U_1, U_2, U_3}. For each time series V (related to one pixel), the distance measure between V and each $U \in Q$ is calculated with SP-TWDTW. If the pattern U_2 (“Pasture”) has the shortest distance to V, so the 1NN method, using the SP-TWDTW similarity measure, would assign the class “Pasture” to V. If we use the SP-TWDTW distance measures as input to machine learning methods, in this example we would have three new meta-features generated from the distances measures between V and each $U \in Q$. These meta-features can then be used as input variables for the machine learning algorithms. The Table 2 shows an example of a training dataset that merges the meta-features created by SP-TWDTW (ptwdtw.forest, ptwdtw.pasture, ptwdtw.cerrado) with the features from [Picoli et al. 2018] (mir, nir, ndvi)[1].

[1]The method proposed by [Picoli et al. 2018] also uses the time series from EVI vegetation index as input variable for machine learning methods.

Table 2. Example of a training dataset using the SP-TWDTW distance measures as meta-features.

pixel	mir.1	mir.2	nir.1	nir.2	ndvi.1	ndvi.2	ptwdtw.forest	ptwdtw.pasture	ptwdtw.cerrado
1	0.5	0.1	0.8	0.7	0.1	0.15	11.3	15.9	5.0
2	0.8	0.85	0.7	0.2	0.1	0.4	3.7	5.0	5.0
3	0.9	0.5	0.8	0.3	0.3	0.3	1.2	10.8	4.1

Although both 1NN and F-TWDTW use SP-TWDTW distance measures, 1NN will use a fixed number of neighbors, in this case just one, to decide the class that V belongs to. On the other hand, machine learning methods use complex functions to improve the classification accuracy using the training data set. For example, these methods can reduce the weight of a training sample that would be the nearest neighbor alone. The classes "Corn" and "Millet" [Picoli et al. 2018], have similar physical characteristics and can be spectrally confused. F-TWDTW can learn to identify these cases and reduce the weight of these samples on the training dataset, which is not possible with 1NN. F-TWDTW thus uses a more complex classification hypothesis than is used by 1NN and therefore F-TWDTW is expected to work better than 1NN, in general.

The machine learning methods are also able to learn the importance of the features from the training dataset and merge the strengths of 1NN with SP-TWDTW in the classification. In this paper, we chose the supervised learning method Support-Vector Machines (SVM) [Cortes and Vapnik 1995], which was shown by [Picoli et al. 2018], as producing the best accuracy using spectral bands and vegetation indices as features. However, the meta-features, proposed in this work, can also be used as input variables to other machine learning algorithms.

5. Experiments and Results

This section aims to evaluate the classification accuracy using the meta-features created by the SP-TWDTW algorithm running on parallel architectures. The classification accuracy was evaluated using an AMD FX-8320E processor machine with 3.2 GHz clock and 24 GB of RAM. The distance measures were calculated by SP-TWDTW on GPU using the NVIDIA GeForce GTX 1050 Ti card with 4 GB GDDR5 of available memory and 768 CUDA cores with a clock of 1392 MHz.

SP-TWDTW execution time was 246 times faster than the TWDTW in R programming language [Oliveira et al. 2018]. The TWDTW was implemented in C++ to be able to compare fairly with SP-TWDTW since the C++ language allows for better performance than R. The SP-TWDTW GPU implementation proved to be a promising solution for processing high temporal resolution data, with a speedup of 10 times over the CPU TWDTW implementation and almost 11 times faster than it for high spatial resolution data.

We used the MOD13Q1 product from National Aeronautics and Space Administration from 2001 to 2016, provided every 16 days at 250-meter spatial resolution in the sinusoidal projection. The dataset used has 2115 time series, in an area of 903,357 km^2 in the state of Mato Grosso, Brazil. Nine ground cover classes were defined, described in Table 3, along with the number of samples per class.

The solution introduced in this work is compared with the method proposed

Table 3. Number of ground samples per class used as training data for machine learning algorithms.

Class	Count
Forest	138
Cerrado	400
Pasture	370
Soy-Fallow	88
Fallow-Cotton	34
Soy-Cotton	399
Soy-Corn	398
Soy-Millet	235
Soja-Sunflower	53

in [Picoli et al. 2018], here denominated SVM-Picoli. The SVM-Picoli method selects the time series observations of vegetation indices, NDVI and EVI, as well as the NIR and MIR spectral bands as input variables for the SVM algorithm. Our work introduces a new method that creates meta-features using the the function F-TWDTW from SP-TWDTW distance measures for each land use and land cover class, adding it to the features vector of SVM-Picoli method. In our experimental scenario, nine features are added to data training, and we denoted our method as SVM-TWDTW in the experiments. The KNN algorithm using the SP-TWDTW similarity measures as input distances is denoted as KNN-TWDTW in the experiments.

A generalized additive model (GAM; Hastie and Tibshirani 1986; Wood 2011) was used by [Picoli et al. 2018] to generate the smoothed temporal patterns. It is flexible for non-parametric fits, with less rigorous assumptions on the relationship between response and predictor. This potentially provides a better fit to satellite data than purely parametric models, due to the data's inter-annual and intra-annual variability. The patterns generated for each one of the nine classes can be seen in Figure 2.

For KNN-TWDTW, we created 100 different data partitions, each one with 90% of the samples for training and 10% for validation. The parameters for the classification method were: the logistic weight function with steepness -0.1 and midpoint 50 for TWDTW; the frequency of the temporal patterns to 8 days, and the GAM smoothing formula to $formula = y\ s(x)$, where function s sets up a spline model, with x the time and y a satellite band.

To estimate the accuracy of the SVM method (SVM-TWDTW and SVM-Picoli), the k-fold cross-validation method was used with $k = 10$ [Kohavi et al. 1995], with 100 different data partitions, each one with 90% as training data and 10% for validation. The setup of both SVM classifiers, SVM-Picoli and SVM-TWDTW, is similar to allow evaluating the impact of adding the SP-TWDTW distance measures as meta-features. We used the SVM algorithm from e1071 [2] library implemented in R with version 1.7-2 and the following parameters: kernel = "radial", degree = 3, coef0 = 0, cost = 10, tolerance = 0.001, epsilon = 0.1, cross = 0, scale = FALSE, cachesize = 1000.

[2]https://www.rdocumentation.org/packages/e1071/versions/1.7-2/topics/svm

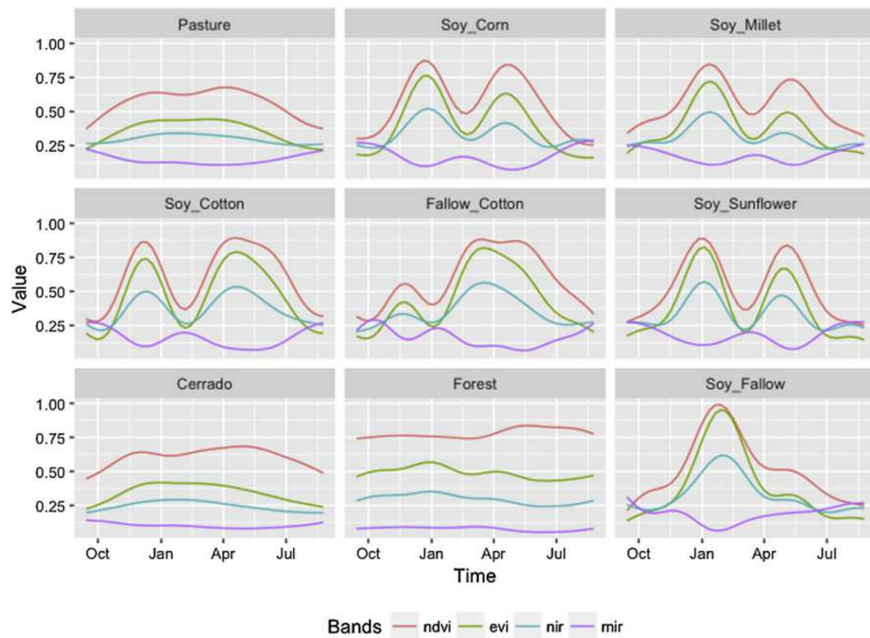


Figure 2. Estimated temporal patterns of NDVI, EVI, NIR and MIR bands for the selected land cover classes using the GAM model [Picoli et al. 2018].

5.1. Results and Discussion

Table 4 shows the confusion matrix with the number of samples classified in each class, the overall accuracy, and the producer and user accuracies of the KNN-TWDTW method. It had a low overall accuracy (OA) of 78% explained by the patterns that have similar temporal behavior. In Figure 2, we can see that the land cover classes Cerrado, Pasture and Forest have similar time series patterns. KNN-TWDTW had good accuracy in identifying forest patterns, with PA of 100%, but it misclassified the Cerrado and Pasture samples, which caused a low UA for the Forest class. This occurs because, for each time series V, the KNN-TWDTW method calculates the distance measure for each pattern U and assigns V to the class with the lower distance measure. But when the patterns are similar, the distance between V and each U is close and leads to misclassification. The same behavior occurs for the Soybean-Corn, Soybean-Millet and Soybean-Sunflower patterns.

To reduce the impact of the similarity of patterns on classification, the work [Picoli et al. 2018] proposes a method that uses the NDVI and EVI vegetation indices, as well as the RED and MIR bands as input variables for SVM. The result of the classification using this method is presented in the confusion matrix of the Table 5, in which the classifier obtained overall accuracy of 92.3%, approximately 12.3% higher than KNN-TWDTW. The matrix shows that there was confusion between the Soy-Corn and Soy-Millet classes, as with KNN-TWDTW, due to similar characteristics of the Soy-Corn and Soy-Millet patterns, but with a lower error rate. Since corn and millet have similar physical characteristics, they can be spectrally confused [Picoli et al. 2018]. Both are grasses, with lanceolate leaves; the height of corn can reach up to 3.5 m, whereas millet varies between 1.5 and 3 m, and can reach more than 5m [Picoli et al. 2018].

Table 4. Confusion matrix using KNN-TWDTW with the Producer's Accuracy (PA) and User's accuracy (UA) values for each class, in addition to the overall accuracy (OA).

		1	2	3	4	5	6	7	8	9	UA (%)
1	**Pasture**	332	3	5	0	0	0	8	0	0	95,4
2	**Soy-Corn**	6	257	32	37	1	14	0	0	0	74,1
3	**Soy-Millet**	1	19	155	2	0	1	0	0	3	85,6
4	**Soy-Cotton**	1	10	2	314	4	0	0	0	0	94,9
5	**Fallow-Cotton**	1	3	2	32	29	0	0	0	0	43,3
6	**Soy-Sunflower**	0	100	29	12	0	37	0	0	0	20,8
7	**Cerrado**	11	0	0	0	0	0	303	0	0	96,5
8	**Forest**	12	0	0	0	0	0	89	138	0	57,7
9	**Soy-Fallow**	6	6	10	2	0	1	0	0	85	77,3
	PA (%)	89,7	64,6	65,9	78,7	85,3	69,8	75,7	100	96,6	**OA:78%**

Table 5. Confusion matrix using method proposed by [Picoli et al. 2018] (SVM-Picoli) with the Producer's Accuracy (PA) and User's accuracy (UA) values for each class, in addition to the overall accuracy (OA).

		1	2	3	4	5	6	7	8	9	UA (%)
1	**Pasture**	355	4	8	0	1	0	2	2	0	95,4
2	**Soy-Corn**	4	350	33	24	0	7	0	0	0	83,7
3	**Soy-Millet**	1	25	191	1	0	3	0	0	0	86,4
4	**Soy-Cotton**	0	13	2	370	5	4	0	0	0	93,9
5	**Fallow-Cotton**	0	1	0	3	28	0	0	0	0	87,5
6	**Soy-Sunflower**	0	4	1	1	0	39	0	0	0	86,6
7	**Cerrado**	9	1	0	0	0	0	396	1	0	97,3
8	**Forest**	1	0	0	0	0	0	2	135	0	97,8
9	**Soy-Fallow**	0	0	0	0	0	0	0	0	88	100
	PA (%)	95,9	87,9	81,2	92,7	82,3	73,5	99	97,8	100	**OA:92,3%**

Table 6 shows the confusion matrix using the SVM-TWDTW method introduced in our work. The SP-TWDTW similarity measures had a positive impact on accuracy, increasing the overall accuracy of SVM-Picoli by 1.5%. This positive impact with an overall accuracy of 93.8% is explained because using the SP-TWDTW similarity measures along with NDVI, EVI, NIR and MIR characteristics, the SVM-TWDTW can combine SVM-Picoli and KNN-TWDTW methods. KNN-TWDTW was able to correctly classify 125 samples that the SVM-Picoli method misclassified, and this number decreases to 91 when compared to the SVM-TWDTW method. An example can be seen by comparing the PA from the class Forest between KNN-TWDTW (Table 4) and SVM-Picoli (5), where KNN-TWDTW has an accuracy of 100% against SVM-Picoli 97.8%. Table 6 shows that SVM-TWDTW had the same accuracy of 100%, thus also better than SVM-Picoli.

The combination of SVM and TWDTW similarity measures also introduces scenarios that generate misclassifications. For example, the PA of Soy-Fallow using the KNN-TWDTW (96.6%) was lower than PA with SVM-Picoli method (100%) because KNN-TWDTW shows some confusion between the Soy-Millet and Soy-Fallow patterns. The SVM-TWDTW method suffered this effect and also had a PA of exactly 96.6%. In

Table 6. Confusion matrix using the method proposed in this paper (SVM-TWDTW) that creates meta-features from SP-TWDTW distance measures. The table shows the Producer's Accuracy (PA) and User's accuracy (UA) values for each class, in addition to the overall accuracy (OA).

		1	2	3	4	5	6	7	8	9	UA (%)
1	Pasture	359	2	4	1	0	0	3	0	0	97,3
2	Soy-Corn	1	358	29	22	1	4	0	0	0	86,3
3	Soy-Millet	4	27	200	3	0	5	0	0	3	82,7
4	Soy-Cotton	1	11	2	373	3	0	0	0	0	95,6
5	Fallow-Cotton	0	0	0	0	30	0	0	0	0	100
6	Soy-Sunflower	0	0	0	0	0	44	0	0	0	100
7	Cerrado	5	0	0	0	0	0	397	0	0	98,7
8	Forest	0	0	0	0	0	0	0	138	0	100
9	Soy-Fallow	0	0	0	0	0	0	0	0	85	100
	PA (%)	97	89,9	85,1	93,4	88,2	83	99,2	100	96,6	**OA:93,8%**

the classification of Soy-Millet samples, SVM-TWDTW had lower accuracy than KNN-TWDTW and SVM-Picoli, with UA equals to 82.7% against 86.4% of SVM-Picoli and 85.6% of KNN-TWDTW. SVM-TWDTW had some difficulties to distinguish the classes Soy-Millet and Soy-Corn than the other methods, which led to mislabeling of some Soy-Corn samples as Soy-Millet. However, SVM-TWDTW had a higher PA than SVM-Picoli (85.1% against 81.2%) and KNN-TWDTW (85.1% against 65.9%) being able to identify correctly a higher number of Soy-Millet samples.

In all other scenarios, Table 6 shows that the user and producer accuracies of the SVM-TWDTW method were better than those of SVM-Picoli and KNN-TWDTW. SVM-Picoli misclassified 145 samples correctly labeled with SVM-TWDTW, while the opposite scenario occurred 113 times. The results show that the SVM-TWDTW method has brought advances in land use mapping accuracy with the use of TWDTW similarity measures as SVM characteristics. So far, the highest overall accuracy in mapping Mato Grosso land use had been achieved with the SVM-Picoli [Picoli et al. 2018] method.

6. Conclusion

This work investigated the accuracy of land use and land cover classification using meta-features derived from SP-TWDTW distance measures, in addition to the spectral band and vegetation indices time series used by [Picoli et al. 2018]. Our method creates meta-features from SP-TWDTW distance measures for each pattern class. The benefit of these meta-features is that they can be used in conjunction with other existing features as input to other classifiers.

SP-TWDTW distance measures were created to be used as input to the KNN classifier, with K = 1. But, when the temporal signatures of each pattern are similar, this leads to some possible confusion. The method introduced in this paper, using the SVM classifier, increases from 78% to 93,8% the classification accuracy compared to 1NN. We also compared our proposal with [Picoli et al. 2018] and there was an increase in accuracy from 92,3% to 93,8% in the land use and land cover classification. The use of SP-TWDTW distance measures allowed revealing previously unseen characteristics of the data, increasing the classification accuracy. These meta-features can also be used as input variables to other classification methods, such as Deep Learning and Random Forest.

Although the study in [Picoli et al. 2018] has evaluated some classifiers, such as Neural Networks and Random Forests, there are other machine learning algorithms not evaluated and that can benefit from these meta-features, such as Deep Learning. Therefore, in future work, it is interesting to evaluate the impact of the accuracy of the land use and land cover classification using other methods besides SVM. Also, it is important to evaluate our proposal in other geographic regions and data from other satellites that may pose challenges not yet revealed in the experimental scenarios of this paper.

Acknowledgement

This work was funded by Companhia Paranaense de Energia (Copel), P&D ANEEL-Copel Distribuição project number 2866-04842017.

References

Abou EL-Magd, I. and Tanton, T. (2003). Improvements in land use mapping for irrigated agriculture from satellite sensor data using a multi-stage maximum likelihood classification. *International Journal of Remote Sensing*, 24(21):4197–4206.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., PD Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., and R Verón, S. (2018). Remote sensing and cropping practices: A review. *Remote Sensing*, 10(1):99.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Costa, W. S., Fonseca, L. M., Körting, T. S., SIMÕES, M., Bendini, H. N., and Souza, R. C. (2017). Segmentation of optical remote sensing images for detecting homogeneous regions in space and time. In *Embrapa Solos-Artigo em anais de congresso (ALICE)*. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 18., 2017, Salvador. Proceedings... Salvador: Unifacs, 2017. p 40-51.

DADI, M. M. (2019). Assessing the transferability of random forest and time-weighted dynamic time warping for agriculture mapping.

Gómez, C., White, J. C., and Wulder, M. A. (2011). Characterizing the state and processes of change in a dynamic forest environment using hierarchical spatio-temporal segmentation. *Remote Sensing of Environment*, 115(7):1665–1679.

Kohavi, R. et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada.

Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., and 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 Sensing*, 9(3):259.

Li, J. and Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53:173–189.

Löw, F., Schorcht, G., Michel, U., Dech, S., and Conrad, C. (2012). Per-field crop classification in irrigated agricultural regions in middle asia using random forest and support vector machine ensemble. In *Earth Resources and Environmental Remote Sensing/GIS Applications III*, volume 8538, page 85380R. International Society for Optics and Photonics.

Lu, M., Chen, J., Tang, H., Rao, Y., Yang, P., and Wu, W. (2016). Land cover change detection by integrating object-based data blending model of landsat and modis. *Remote Sensing of Environment*, 184:374–386.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166 – 177.

Maus, V. (2016). *Land Use and Land Cover Monitoring Using Remote Sensing Image Time Series*. PhD thesis, PhD thesis, Instituto Nacional de Pesquisas Espaciais, Sao José dos Campos.

Maus, V., Câmara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., and de Queiroz, G. R. (2016). A time-weighted dynamic time warping method for land-use and land-cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(8):3729–3739.

Maus, V., Câmara, G., Appel, M., and Pebesma, E. (2019). dtwsat: Time-weighted dynamic time warping for satellite image time series analysis in r. *Journal of Statistical Software, Articles*, 88(5):1–31.

Oliveira, S. S., Pascoal, L. M. L., Ferreira, L., de Castro Cardoso, M., Bueno, E. F., Vagner, J., and Martins, W. S. (2018). Sp-twdtw: A new parallel algorithm for spatio-temporal analysis of remote sensing images. In *GEOINFO*, pages 46–57.

Petitjean, F., Inglada, J., and Gançarski, P. (2012). Satellite image time series analysis under time warping. *IEEE Transactions on Geoscience and Remote Sensing*, 50(8):3081–3095.

Petitjean, F. and Weber, J. (2014). Efficient satellite image time series analysis under time warping. *Ieee geoscience and remote sensing letters*, 11(6):1143–1147.

Picoli, M. C. A., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., Coutinho, A., Esquerdo, J., Antunes, J., Begotti, R. A., et al. (2018). Big earth observation time series analysis for monitoring brazilian agriculture. *ISPRS journal of photogrammetry and remote sensing*, 145:328–339.

Simonneaux, V., Duchemin, B., Helson, D., Er-Raki, S., Olioso, A., and Chehbouni, A. (2008). The use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central morocco. *International Journal of Remote Sensing*, 29(1):95–116.

Wegner Maus, V., Câmara, G., Appel, M., and Pebesma, E. (2019). dtwsat: Time-weighted dynamic time warping for satellite image time series analysis in r. *Journal of Statistical Software*, 88(5):1–31.