A GENERATIVE METHOD FOR SIMULTANEOUS CLASSIFICATION OF REMOTE SENSING TIME SERIES DATA USING AN ENSEMBLE OF DECISION TREE CLASSIFIERS

Luciano Vieira Dutra [1], Camilo Daleles Rennó [1], Mariane Souza Reis [1], Paolo Gamba [2]

[1]National Institute for Space Research (INPE), São José dos Campos 12227-010, Brazil. luciano.dutra@inpe.br (LVD), camilo.renno@inpe.br (CDR), mariane.reis@inpe.br (MSR)
[2]University of Pavia, Pavia 27100, Italy. paolo.gamba@unipv.it

ABSTRACT

Time series of remote sensing data has become an essential input for land use and land cover (LULC) studies. The current availability of multi-temporal data sets, from different sources and types, demands new classification approaches to explore their full capacity. In this study, we propose a non-parametric version of the Compound Maximum *a posteriori* classifier, based on an ensemble of Decision Tree Classifiers. This classifier was designed to avoid the classification of inconsistent class sequences in time. It was tested in a study area located in Itaituba, Pará state, Brazil, by the classifications of five Landsat images. In our case study, more than 25% of time series would be classified as invalid transitions. The use of the proposed approach substitutes these results with the most probable consistent class trajectory. Improvements in individual accuracies, when compared to post-classification comparison, have also been observed.

Keywords — multi-temporal classification, Landsat, land cover trajectory.

1. INTRODUCTION

Remote sensing (RS) time series is a primary input data for the sequential classification of land use and land cover (LULC) in time [1]. To date, the most common method to classify an image time series into a LULC map time series is the so-called post-classification comparison (PCC). PCC consists in independently classifying each image from each date, and then stacking and comparing the results. However, this method is prone to the mapping of inconsistent transitions/trajectories, i.e sequences of classes in two or more positions in time that could never happen in the real world [2].

A recent study [2] proposed a generative method [3] to classify a sequence of observations (e.g., a pixel time series), simultaneously, into a time sequence of classes (e.g., a LULC trajectory). This method, called Compound Maximum *a posteriori* (CMAP), generates only consistent class sequences. In its first version, CMAP was proposed considering the Gaussian distribution modeling.

The current availability of a vast array of multiple data sources, including categorical and other non-spectral data sets, implies the need for new ways to model the data. To this end, the present study proposes a non-parametric version of CMAP, based on an ensemble of Decision Tree Classifiers (DTCs). Here, we propose a method to use the leaves of DTCs to assign conditional probabilities to the classes involved in the CMAP classification process, as described in Section 2. The advantages of this method are illustrated in the study case presented in Sections 3 and 4. Section 5 brings the conclusions of this study and future perspectives.

2. THEORETICAL CONSIDERATIONS

The initial basis of the presented classification approach is the same as presented in Reis et al. [2]. Consider $s_m = \{\omega_{k_1}^1, \dots, \omega_{k_t}^t, \dots, \omega_{k_T}^T\}$, $s_m \in S = \{s_1, \dots, s_M\}$ the m^{th} temporal class sequence s in a set of M possible class sequences, T is the length of the time sequence and $\omega_{k_t}^t \in \Omega_t = \{\omega_1^t, , \dots, \omega_k^t, \dots, \omega_{K_t}^t\}$, in which $\omega_{k_t}^t$ is the actual class at time position t of s_m. The notation k_t is an indicator of the class ω_k^t in the set Ω_t that holds the K_t possible classes on time t. This development allows for class sets of different sizes/natures in each time t.

A given observation vector $\vec{X} = \{\vec{x}_1, \dots, \vec{x}_t, \dots, \vec{x}_T\}$ contains the T temporal observations that can indicate the class sequence composition. $\vec{x}_t$ can represent, for example, the digital number vector of a specific image pixel at time t. No restrictions on RS data type and size are necessary at this point, including categorical data.

The classification process consists of choosing the sequence with the highest probability $\hat{s}$ of S. A generative method for classification can be formulated as follows:

$$\hat{s} = \arg\max_s \left(P(\vec{X}, s), s \in S\right) \quad (1)$$

or using the definition of conditional probability:

$$\hat{s} = \arg\max_s \left(P(\vec{X}/s) \times P(s), s \in S\right) \quad (2)$$

in which $P(s)$ is the *a priori* probability of a sequence s. Supposing that the observations are time-independent and that each one depends only on the observed object, we have:

$$P(\vec{X}/s) = P(\vec{x}_1/\omega^1_{k_1}) \times \ldots \times P(\vec{x}_t/\omega^t_{k_t}) \times \ldots \times P(\vec{x}_T/\omega^T_{k_T}) \quad (3)$$

The expression (3) is called here a Compound Likelihood (CL). The main point of the expression (2) is modeling the *a priori* probabilities of the sequence $P(s)$. In this research, these probabilities will be formed by a concatenation of transition probabilities, so-called Compound a *priori* (CA) [2].

The use of the equation (3) jointly with expression (2) returns the estimation herein defined as the Compound Maximum *a posteriori* estimation of a class sequence, or CMAP:

$$\hat{s} = \underset{s}{argmax}\left(\prod_{t=1}^{T} P(\vec{x}_t/\omega^t_{k_t}) \times P(s), s \in S\right) \quad (4)$$

2.1. Using a Decision Tree Classifier (DTC) as the base classifier

In Reis et al. [2], $P(\vec{x}_t/\omega^t_{k_t})$ is modeled as a Gaussian distribution. Here, this probability will be modeled using a Decision Tree Classifier (DTC). We propose to use DTC as a feature extractor, by considering the leaves as observational random variables instead of $\vec{x}_t$. Usually, DTC will result in a number of leaves greater or equal the number of classes at hand. For each leaf, it is possible to calculate its conditional probability to each class of interest.

Be $N^t_1, \ldots, N^t_k, \ldots, N^t_K$ the number of reference vectors for class k for a certain timestamp t, $N_t = \sum_k N^t_k$. Usually, either one selects N_t in such way that all N^t_k are proportional to the *a priori* probability of each class at time t, or one chooses equal-sized sets for reference vectors (N^t_k = constant). Using a general technique of bootstrap aggregating, or bagging, a subset with n^t_k samples for each class k is generated to fit a decision tree $\mathcal{D}_t \coloneqq [\mathcal{L}^t_1, \ldots \mathcal{L}^t_j, \ldots, \mathcal{L}^t_{J_t}]$ with J_t leaves. Different algorithms can be used to generate decision trees. Herein, we tested the C4.5 algorithm [4], but other approaches can also be easily implemented.

In many cases, each leaf is associated with a class, but due to the simplification of the tree through pruning or the definition of a minimum number of samples per leaf, samples from different classes may be included in the same leaf. Furthermore, the same class can be represented on different leaves. Considering that n^t_{kj} samples of class ω^t_k fall on leaf $\mathcal{L}^t_j$, the conditional probability of each class can be estimated by:

$$P(\mathcal{L}^t_j/\omega^t_k) \cong \frac{n^t_{kj}}{n^t_k} \quad (5)$$

As mentioned before, the random variable leaf $\mathcal{L}^t_j$ will be used as feature instead of $\vec{x}_t$. Hence, equation (4) can be expressed as follows:

$$\hat{s} = \underset{s}{argmax}\left(\prod_{t=1}^{T} P(\mathcal{L}^t_j/\omega^t_{k_t}) \times P(s), s \in S\right) \quad (6)$$

This approximation enables the calculation of CML based on the DTC. The CA, as mentioned in Reis et. al [2], will be ad-hoc established based on a concatenation of so-called transition matrices.

2.2 Using an ensemble of Trees as a classifier

Now, suppose that D_t different trees can be obtained through resampling the training sets, in general maintaining the same number of training vectors per class.

Such set, called an Ensemble of DTCs (DTCe), is given by $\vec{\mathcal{D}}^t = [\mathcal{D}^t_1, \ldots, \mathcal{D}^t_d, \ldots, \mathcal{D}^t_{D_t}]$, in each timestamp t. Note that it is similar to the traditional Random Forest classifier. However, it considers more relaxed premises, as to be considered a more generalized approach. An input feature $\vec{x}_t$ will be mapped in a random vector of D_t leaves, $\vec{\mathcal{L}}_t = [\mathcal{L}^{t,1}_{j_1}, \ldots, \mathcal{L}^{td}_{j_d}, \ldots, \mathcal{L}^{tD_t}_{j_{D_t}}]$, where j_d indicates which leaf was selected by each tree d, hence equation (6) becomes:

$$\hat{s} = \underset{s}{argmax}\left(\prod_{t=1}^{T} P(\vec{\mathcal{L}}_t/\omega^t_{k_t}) \times P(s), s \in S\right) \quad (7)$$

The leaf vector of conditional probabilities $P(\vec{\mathcal{L}}_t/\omega^t_{k_t})$ would be properly estimated by considering the joint $\vec{\mathcal{D}}^t$ probabilities. This process is not easily treatable, specially noting that the trees of $\vec{\mathcal{D}}^t$ are not mutually exclusive. In this study, a simpler approach is taken by averaging the individual estimates of the leaf conditional probabilities:

$$P(\vec{\mathcal{L}}_t/\omega^t_k) = \frac{\sum_{d=1}^{D_t} P(\mathcal{L}^t_{j_d}|\omega^k_t)}{D_t} \cong \frac{\sum_{d=1}^{D_t} n^t_{kj_d}}{n^t_k D_t} \quad (8)$$

where $n^t_{kj_d}$ is the number of reference vectors for class k which felt in leaf j_d of tree d at time t.

3. EXPERIMENTAL SETUP

3.1. Study area

The chosen study area comprises the municipality of Itaituba, Pará, Brazil (Figure 1), located on the banks of the Tapajós River. Originally, the region was covered by the Amazon rainforest. The main activities in the region include gold mining (mainly in the 1980 - 1990s), agriculture and logging.

This study was based on five image data sets from the Operational Land Imager (OLI) sensor, onboard Landsat 8, from orbit/point 228/63. These images were acquired in the dry season to minimize cloud problems, in dates 2015-08-21 (Figure 1), 2016-08-07, 2017-09-13, 2018-07-04, and 2019-

06-29. We used band 1 to 6, as downloaded from *https://earthexplorer.usgs.gov/*, processing level L1 /Collection 1. Relative calibration of the images was not performed.

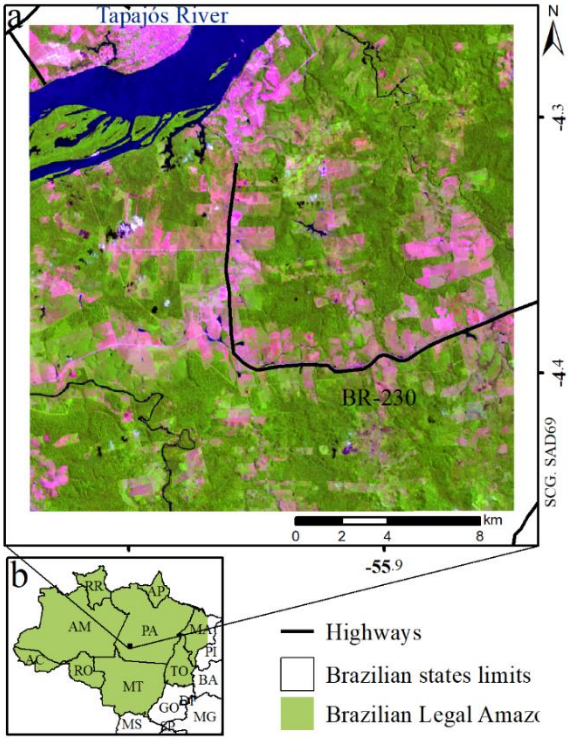


Figure 1. Study area: a) Landsat/OLI image from 2015-08-21, color composition 6(R)5(G)4(B), b) in relation to Amazon.

3.2. Classes and transition matrices

Five LULC classes were considered: a) Primary forest (*forest*); b) Secondary vegetation (*regen*); c) Pasture (*pasture*); d) Bare soil (*baresoil*); and e) Water (*water*). For 2015 and 2018, it was necessary to consider two other classes: Cloud (*cloud*) and Cloud shadow (*shadow*). Samples were collected for each image on each date to characterize the spectral heterogeneity of each target. Table 1 presents the number of collected samples.

Class	Year				
	2015	2016	2017	2018	2019
forest	9113	9564	10487	10737	11061
regen	2246	2329	2182	2121	1576
pasture	1304	1699	2216	1675	1136
baresoil	5840	4725	4259	3747	3247
water	3448	3607	3464	3764	3610
cloud	339	-	-	796	-
shadow	187	-	-	523	-

Table 1. Sample size.

Transition matrices (Table 2) were defined to prevent certain classes being converted to others. For example, if a region was deforested, it could be converted immediately to bare soil or pasture. Only in the following year, this area could start to regenerate.

		time *t+1*						
		forest	regen	pasture	baresoil	water	cloud	shadow
time *t*	forest	1	0	1	1	1	1	1
	regen	0	1	1	1	0	1	1
	pasture	0	1	1	1	0	1	1
	baresoil	0	1	1	1	1	1	1
	water	0	0	0	1	1	1	1
	cloud	1	1	1	1	1	1	1
	shadow	1	1	1	1	1	1	1

Table 2. Transition matrix.

3.3. Methodology

The images were classified using the ensemble of Decision Tree Classifier (DTCe), in two approaches: a) each image was classified independently - trajectories were evaluated considering a PCC approach; b) the images were classified using CMAP-DTCe. The main difference between these approaches is that CMAP constrains the possible transitions between consecutive dates based on the transition matrix defined in Table 2. For both cases, the number of trees has been set to 50 so as not to compromise rating performance. To obtain each tree, 500 samples from each class were randomly selected (with replacement). The parameters that define the classifier C4.5 were the same for all trees: confidence factor = 0.1; Minimum Number of Objects = 5 and Number of Folds = 3.

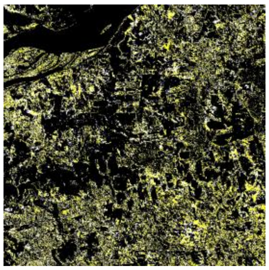
Results were assessed by computing the number of disagreements between the LULC classification results, and also by the traditional Global Kappa index, calculated using a set of references samples not used to train the classifiers. Note that the disagreement between at least one pair of classifications of the same date, in our case, also represents the number of inconsistent trajectories that were avoided with the use of CMAP.

4. RESULTS AND DISCUSSIONS

Figure 2 shows the comparison between trajectory classification results obtained by the independent and CMAP-DTCe. These results include the analysis of disagreement between trajectory classifications (in which 0 means the same results for all images between 2015 and 2019, and 5 means that all LULC classifications differed), and the final LULC classifications for 2015 and 2019, as an example.

The main contribution of CMAP usage is the result robustness by avoiding inconsistent transitions. The use of transition matrices prevented more than 25% of inconsistent LULC trajectories from being generated. The spatial distribution of these occurrences is not homogeneous.

Disagreements between independent and CMAP DTCe trajectories



#	%
0	73.33
1	17.07
2	8.12
3	1.43
4	0.03
5	0.02

LULC classification results

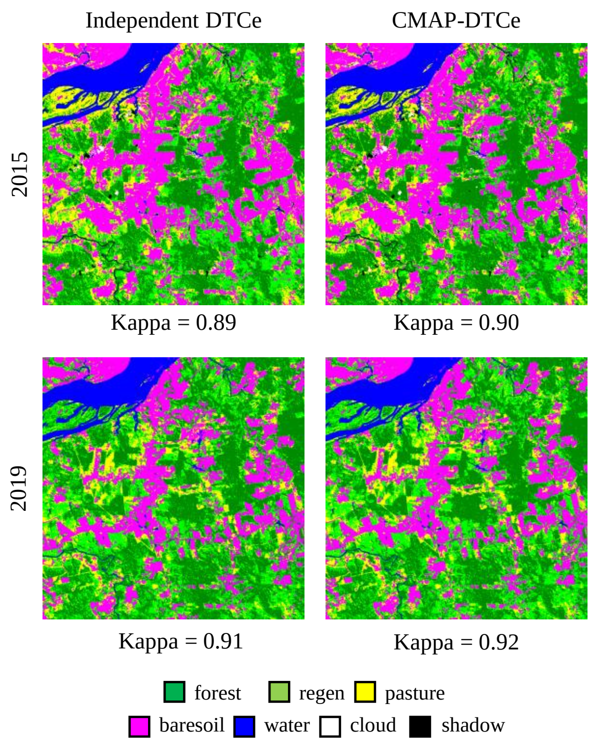


Figure 2. Comparison between Independent and CMAP DTC.

Nonetheless, CMAP usage also leads to some improvement on accuracy values for each date and a decrease in classification noise. However, differences on general accuracy indexes, such as the calculated Kappa values, tend to not be too expressive because of the selection of reference samples: these are usually selected in very stable regions, where inconsistent transitions are less likely to happen. In general, we observed that more pronounced improvements in accuracy tend to happen in: 1) classes more likely to suffer inconsistencies; 2) data sets that these classes are poorly classified with the independent approach; 3) longer trajectory classifications [5] and 4) the proximity of homogeneous field borders.

5. CONCLUSIONS

In the study, we propose a non-parametric version of CMAP, based on an ensemble of Decision Tree Classifiers, the CMAP-DTCe. To this end, we also propose a methodology to derive a discriminant function using DTCe, that can be used to classify a set of RS images. CMAP-DTCe is an efficient methodology for producing inconsistencies-free classifications, with greater impact mainly on classes prone to suffer from invalid transitions.

Future research will continue to improve the methodology in terms of less biased estimations of DTCe probabilities. It will also further verify the effects of different transition matrixes and longer time series for the classification of certain classes of interest, prone to be involved in inconsistent transitions. Also, new adaptations for CMAP are being programmed, by adapting other base classifiers, including contextual ones. In the specific case of DTC based classifiers, it is worth to mention the capability of including non-numeric features in the analysis, which shall be tested shortly.

6. ACKNOWLEDGMENTS

This research was partially funded by the National Council for Scientific and Technological Development, (CNPq) projects “Monitoring Brazilian Biomes by Satellite–Building new capacities”/process 444418/2018-0, grant #350131/2022-8, and PQ grant #309135/2015-0. LVD and CDR also want to thank the University of Pavia, especially Dr Fabio Dell´Acqua, for the warm reception and support during the two months stay in 2019, when most of this research was executed.

7. REFERENCES

[1] M. A. Wulder, N. C. Coops, D. P. Roy, J. C. White, and T. Hermosilla. Land cover 2.0. *International Journal of Remote Sensing*, 39 (12) : 4254–4284, 2018.

[2] M. S. Reis, . V. Dutra, M. I. S. Escada, and S. J. S. Sant’Anna. Avoiding invalid transitions in land cover trajectory classification with a Compound Maximum a Posteriori approach. *IEEE Access*, 8:98787–98799, 2020.

[3] A. Y Ng and M. I Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In Advances in neural information processing systems, pages 841–848, 2002.

[4] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[5] M. S. Reis. Detection and analysis of forest regeneration trajectories in the Lower Tapajós region. 227 p. Thesis (Doctorate in Earth System Science) – National Institute for Space Research (INPE), São José dos Campos, 2022.