Spatial CART Classification Trees

Avner Bar-Hen, Servane Gey, Jean-Michel Poggi

▶ **To cite this version:**

Avner Bar-Hen, Servane Gey, Jean-Michel Poggi. Spatial CART Classification Trees. 2020. hal-01837065v2

HAL Id: hal-01837065

https://hal.science/hal-01837065v2

Preprint submitted on 24 Nov 2020 (v2), last revised 16 Mar 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spatial CART Classification Trees

Avner Bar-Hen* Servane Gey† Jean-Michel Poggi‡

Abstract

We propose to extend CART for bivariate marked point processes to provide a segmentation of the space into homogeneous areas for interaction between marks. While usual CART tree considers marginal distribution of the response variable at each node, the proposed algorithm, SpatCART, takes into account the spatial location of the observations in the splitting criterion. We introduce a dissimilarity index based on Ripley's intertype K-function quantifying the interaction between two populations. This index used for the growing step of the CART strategy, leads to a heterogeneity function consistent with the original CART algorithm. Therefore the new variant is a way to explore spatial data as a bivariate marked point process using binary classification trees. The proposed procedure is implemented in an `R` package, and illustrated on simulated examples. SpatCART is finally applied to a tropical forest example.

Keywords: CART; Bivariate Marked Point Process; Spatial CART; Ripley's intertype K-function

1 Introduction

We propose an extension of CART for bivariate marked spatial point processes, which takes into account the spatial information in the splitting criterion and providing a segmentation of the space into homogeneous areas for interaction between marks.

CART (Classification And Regression Trees) is a statistical method, introduced by Breiman *et al.* [7], and designing tree predictors for both regression and classification. We restrict our attention on the classification case with two populations. Each observation is characterized by some input variables gathered in vector X and a binary label Y which is the output (or response) variable.
The general principle of CART is to recursively partition the input space using binary splits and

*Cnam, Paris, France
†Laboratoire MAP5, UMR 8145, Univ. Paris, France
‡Laboratoire de Mathématiques, Univ. Paris-Saclay, Orsay, France and Univ. Paris, France

then determine an optimal partition for prediction. The classical representation of the model relating Y to X is a tree representing the underlying process of construction of the model. If the explanatory variables are spatial coordinates, we get a spatial decision tree and this induces a tessellation of the space (of the X variables). A cell of this tessellation corresponds to a leaf of the decision tree. For a leaf of this tree, the response variable Y is constant and corresponds to the majority label of the observations belonging to this leaf.

The aim of this article is to adapt CART decision trees within the framework of binary marked point processes by considering that the points inside the cells defined by the partition are a realization of a spatial point process, where the response variable Y can be seen as a mark of the point of position given by X. Following this idea, the proposed variant of the CART algorithm takes into account the spatial dimension of the bivariate marked point processes, and thus makes it possible to provide the user a partition of the plan defining homogeneous areas for interaction between the two marks. Hence, the new variant is a way to explore spatial data as a bivariate marked point process using binary classification trees.

While a classical assumption for CART is to consider a sample of *i.i.d.* observations, our variant takes into account spatial dependence using the Ripley's intertype K-function to suitably redefine the splitting criterion in CART. Instead of being based only on the respective marks' proportions, the splits will be selected according to the strength of interaction between the two populations. When exist, the spatial patterns corresponding to homogeneous interaction sub-windows cannot, in general, be discriminated by traditional CART. Indeed, CART segments space according to the marks' proportions, favoring nodes of individuals of the same mark, without any reference to the spatial positions of the individuals. At the contrary, the new impurity function based on spatial interaction between the marks is adapted to capture homogeneous interaction sub-windows.

Let us mention some akin methods, out of the scope of this paper, since they are related to the regression case. Among the various extensions of CART, Bel *et al.* ([6]) proposed a variant for spatial data by considering the observations as a sample of regionalized variables modeled as a random field with spatial dependence. Various reweighting schemes to equalize the contribution of each surface unit for the choices of splits are proposed. In [14], the authors suggested the use of recursive partitioning scheme like CART for space segmentation for the spatio-temporal hotspot detection problem, where the response variable is continuous. We could also cite some methods able to take into account the spatial effect even if they are not especially designed to cope with spatial data. For example a model-based regression tree with a spatial lag is considered in [23]. Complexifying the approach, by merging the possibility of incorporating covariates and taking into account the spatial effect, we could refer to the GAM regression models framework. For example, the BAMLSS family [22] which allows the parameters of the response distribution to be modeled by explanatory variables. Then ensembles of models like the BART (Bayesian Additive

Regression Trees) [8] method implementing a sum-of-trees model and a regularization prior on the parameters of that model. Finally, other extensions of GAM combined with boosted trees may also be mentioned, see for example [18].

The paper is organized as follows. In Section 2, we recall some basics about CART decision trees in the classification case. In Section 3, we review some conventional ways to quantify interaction between two point processes and we define the heterogeneity function associated with the Ripley's intertype K-function. Then, in Section 4, we propose a spatial variant of CART dealing with point processes and describe how to implement it. Section 5 illustrates its use on simple simulated examples. Section 6 addresses an application to the spatial distribution of two species of tropical forest. Finally, the concluding Section 7 sketches some perspectives.

2 CART method

Let us briefly recall some general background on classical settings about Classification And Regression Trees (CART). The data are considered as an independent sample of the random variables $(X^1, \ldots, X^p, Y)$, where the X^ks are the explanatory variables (supposed to be numerical in this article) and Y is the categorical variable to be explained. In a general framework, Y is a multinomial variable, but we suppose here that Y is a binary variable with two labels.
CART is a rule-based method that generates a binary tree through recursive partitioning. Let us consider a classification tree T, where a class label is assigned to each terminal node (or leaf) of T. Hence T can be viewed as a mapping to assign a value $\widehat{Y}_i = T(X_i^1, \ldots, X_i^p)$ to each observation.

The growing step leading to a deep maximal tree is obtained by recursive partitioning of the training sample by selecting the best split at each node according to some heterogeneity index, such that it is equal to 0 when there is only one class represented in the node to be split, and is maximum when all classes are equally frequent. The two most popular heterogeneity criteria are the Shannon entropy and the Gini index. Among all binary partitions of each set of values of the explanatory variables at a node t, the principle of CART is to split t into two sub-nodes t_L and t_R according to a threshold on one of the variables, such that the reduction of heterogeneity between a node and the two sub-nodes is maximized. Some variables may be used several times while others may not be used at all. The growing procedure is stopped when there is no more admissible splitting. Each leaf is assigned to the most frequent class label of its observations.

Of course, such a maximal tree (denoted by T_{max}) generally overfits the training data and the associated prediction error $R(T_{max})$, with

$$R(T) = \mathbb{P}(T(X^1, \ldots, X^p) \neq Y), \tag{1}$$

is typically large. Since the goal is to build from the available data a tree T whose prediction error is as small as possible, in a second stage the tree T_{max} is pruned to produce a subtree T' whose expected performance is close to the minimum of $R(T')$ over all binary subtrees T' of T_{max}. Since the joint distribution $\mathbb{P}$ of $(X^1, \ldots, X^p, Y)$ is unknown, the pruning is based on a penalized empirical risk to balance optimistic estimates of empirical risk by adding a complexity term $\alpha|T'|$ that penalizes larger subtrees, with $|T'|$ the number of leaves of tree T' and α a positive penalty constant. The pruning step produces a decreasing sequence of K subtrees pruned each one from another, denoted by $T_1 \succ \ldots \succ T_K = \{t_1\}$, and associated with an increasing sequence of K complexities, denoted by $0 = \alpha_1 < \ldots < \alpha_K$.

According to the penalized criterion used in the pruning procedure, the sequences of subtrees pruned from the same maximal tree are quite different. In the most commonly used variant of CART classification trees, the aim is to predict labels $Y \in \{m_1; m_2\}$. Hence the natural penalized criterion $\hat{R}_{pen}(T)$ used in this context is based on the misclassification rate of a tree T:

$$\hat{R}_{pen}(T) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{T(X_i^1, \ldots, X_i^p) \neq Y_i\}} + \alpha|T| \tag{2}$$

where $\mathbb{1}$ is the indicator function, n the total number of observations, α the positive penalty constant, $|T|$ the number of leaves of the tree T, and Y_i the ith random realization of Y.
In the variant of CART class probability trees, the aim is to predict probabilities $p(j|x) = \mathbb{P}(Y = m_j \mid X = x)$ rather than labels $Y \in \{m_1; m_2\}$. Hence this variant prunes the maximal tree through a penalized criterion based on estimation error of the $(p(j|x))_{j=1,2}$ rather than misclassification rate. As it will be seen in subsection 4.2, this kind of variant is useful when dealing with marks' spatial intensity rather than marks' majority vote. The penalized criterion used for class probability trees is derived from the mean square error risk as follows. Let (Z_1, Z_2) be defined by $Z_j = \mathbb{1}_{\{Y=m_j\}}$ and let $p(j|x) = \mathbb{P}(Y = m_j \mid X = x)$. Then $\mathbb{E}(Z_j \mid X = x) = p(j|x)$, $j = 1, 2$. In this context, the criterion is a penalized version of the Gini index. An empirical version of penalized Gini index can be written as [7, sec. 4.6]:

$$\hat{R}_{pen_G}(T) = \frac{1}{n} \sum_{t \in \widetilde{T}} n_t \left(1 - \sum_{j=1}^{2} \hat{p}(j|t)^2 \right) + \alpha|T|, \tag{3}$$

where n_t is the number of observations falling in node t, and $\hat{p}(j|t)$ the proportion of observations with mark j falling in node t.

The maximal tree T_{max} is computed via recursive partitioning, and then pruned with the penalized criterion based on misclassification rate defined in (2) to produce classification trees, or based on Gini index defined in (3) to produce class probability trees. For all $\alpha \geq 0$,

$\text{argmin}_{T \preceq T_{max}} \hat{R}_{pen}(T)$, or $\text{argmin}_{T \preceq T_{max}} \hat{R}_{pen_G}(T)$, belongs to the sequence of nested pruned subtrees $(T_k)_{1 \leqslant k \leqslant K}$. The final step is to choose a convenient tree among the sequence $T_1 \succ \ldots \succ T_K = \{t_1\}$ or equivalently a convenient complexity among the sequence $0 = \alpha_1 < \ldots < \alpha_K$. This choice can be made via cross-validation.

In this article, $p = 2$ since we only consider the spatial coordinates as available information.

3 Quantifying the link between two point patterns

3.1 Basics on point processes

A point process is a random variable that gives the localization of events in a set $W \subset \mathbf{R}^d$. Another way to define a given point process is to consider, for each $B \subset W$, the number of events $\phi(B)$ occurring in B, where ϕ is the distribution of the occurrences of the point process. Since characterization of a spatial repartition is strongly dependent on the scale of observation, the repartition has to be characterized for each scale.

There are classical assumptions about point processes. At first we consider that the probability to observe two points at the same place is null. Up to misrecording, this hypothesis is not very restrictive. Two extra common hypotheses are stationarity and isotropy. Intuitively, it means that the distribution of the occurrences of the point process ϕ is not affected by translation or rotation around the origin, *i.e.* the characteristics of the point process are the same for the whole area under study. A point process that is stationary and isotropic is said to be homogeneous.
A marked point process is a point process such that a random mark is associated with each localization. In this article, we only consider bivariate point processes, *i.e.* the mark is a qualitative random variable with two possible issues. Equivalently, the bivariate point process can be viewed as the realization of two point processes (one per level of the mark).
There are several ways to consider the relationships between two clouds of points, mainly related to three aspects: independence, association and random labelling (see [5] for example). It ends up that relationships between two clouds of points can be described in various ways and therefore many indices can be defined. Each index will give a specific information about these relationships and will greatly depend on the point process that leads to the observed repartition. For bivariate point processes, many tools based on first-order characteristics of the point processes, may be used to quantify departure from independence (see [9] for example).

3.2 Intertype K-function

3.2.1 Definition

Under the assumptions of stationarity and isotropy, the intertype K_{ij}-function is a bivariate extension of Ripley's K-function, proposed in [19], and defined as

$$K_{ij}(r) = \lambda_j^{-1}\mathbb{E}(\text{number of points of type } j \text{ within distance } r \text{ of a randomly chosen point of type } i) \tag{4}$$

where the intensity parameters λ_i and λ_j correspond to the expected numbers of type i and type j points per unit area, respectively.

The idea of intertype K_{ij}-function is to characterize interaction between two point processes, *i.e.* second order characteristic of the bivariate point process. The assumption of stationarity is quite a strong hypothesis and could be relaxed using local estimation of the intensity [3], but this leads to instability of the estimator of the intertype K_{ij}-function.

While Ripley's K-function characterizes the spatial structure of a univariate pattern at various scales, the intertype K_{ij}-function characterizes the spatial structure of a bivariate pattern, and more precisely the spatial relationship between two types of points located in the same study area. The intertype K_{ij}-function is defined so that $\lambda_j K_{ij}(r)$ is the expected number of type j points in a circle of radius r centered on an arbitrary type i point of the pattern. Symmetrically, we can define an intertype K_{ji}-function so that $\lambda_i K_{ji}(r)$ is the expected number of type i points in a circle of radius r centered on an arbitrary type j point.
If the bivariate spatial process is homogeneous, then $K_{ij}(r) = K_{ji}(r)$. Under independence, the intertype K function is $K_{ij}(r) = \pi r^2$, regardless of the individual univariate spatial patterns of the two types of events.

In order to interpret the observed values of $K_{ij}(r)$, we need to compare them to the theoretical values obtained for simple cases of bivariate patterns, and especially to πr^2 that corresponds to a null hypothesis of absence of interaction between the two types of points. Positive values of $K_{ij}(r) - \pi r^2$ indicate attraction between the two processes at distance r while negative values indicate repulsion. However, depending on the context of the study, this appropriate null hypothesis can correspond to at least two different statistical hypotheses (see [10]): independence or random labelling. Note that these two hypotheses are different. On one hand, independence of the marginal processes states that the point process that gives the locations of the first point process is independent of the point process that gives the locations of the second point process. The two processes are therefore distributed independently. On the other hand, random labelling states that, conditionally on the locations of points, their mark is assigned to them at random.

3.2.2 Estimation

The estimator of the intertype K_{ij}-function can be defined by:

$$\hat{K}_{ij}(r) = (\hat{\lambda}_i \hat{\lambda}_j A)^{-1} \sum_{k,l} \mathbb{1}_{\{d_{i_k,j_l} < r\}} \tag{5}$$

where d_{i_k,j_l} is the distance between the kth location of type i point and the lth location of type j point, A is the area of the region of interest and $\hat{\lambda}_i$ and $\hat{\lambda}_j$ are the estimated intensities.
This estimator $\hat{K}_{ij}(r)$ is a function of the distance between points. It can be integrated to provide single valued indices (see [11] for example).
As the theoretical distributions of the estimators are unknown, confidence intervals are commonly estimated through Monte Carlo simulations of a specified null hypothesis. To test independence, a classical method is to keep the patterns of both point processes unchanged, but to randomize their relative position at each Monte Carlo simulation. On the other hand, a classical approach to test random labelling is to simulate realizations of a point process with the same spatial structure as the overall observed pattern (*i.e.* without type distinction), and to randomly attribute marks.

Remark 1. Various edge corrections have been suggested; one common example is the extension of Ripley's estimator, leading to:

$$\hat{K}_{ij}(r) = (\hat{\lambda}_i \hat{\lambda}_j A)^{-1} \sum_{k,l} w(i_k, j_l) \mathbb{1}_{\{d_{i_k,j_l} < r\}}.$$

The coefficient $w(i_k, j_l)$ is the inverse of the proportion of the perimeter of the circle centered at the kth location of type i point with radius d_{i_k,j_l} that lies inside the study area. Ripley ([20]) shows that this corrected estimator is unbiased. When edge corrections are used $\hat{K}_{ij}(r)$ and $\hat{K}_{ji}(r)$ are positively correlated and no more equal.
In the sequel, in order to obtain explicit formulation for variation of heterogeneity, we use a Ripley's intertype function without edge correction as defined in Equation 5.

4 Spatial CART

4.1 Impurity loss based on K_{ij}

The key idea is to take into account the spatial dependence of the data in the CART splitting strategy. It is done by modifying the original impurity loss, which is usually the entropy index. We introduce a dissimilarity index based on Ripley's intertype K-function quantifying the interaction between two populations within a rectangular window of finite area A.

Let us focus on the impurity loss associated with $\hat{K}_{ij}$ as defined in (5). First, we define the impurity of a node t at radius r as $\hat{K}_{ij}^t(r)$, the estimation of the Ripley's intertype K-function restricted to node t:

$$\hat{K}_{ij}^t(r) = \frac{1}{\hat{\lambda}_i^t \hat{\lambda}_j^t A^t} \sum_{i_k, j_l \in t} \mathbb{1}_{\{d_{i_k,j_l} < r\}}, \tag{6}$$

where one has:

A^t	the area of node t
$\hat{\lambda}_*^t$	the estimation of the density of mark $*$ in node $t (* = i, j)$
$d_{i_k j_l}$	the euclidean distance between i-marked individual i_k and j-marked individual j_l

Then, for a split s splitting node t into two child nodes t_L and t_R, the impurity (6) of node t at radius r is decomposed as the sum of the contribution within nodes and between nodes:

$$\begin{aligned}
\hat{K}_{ij}^t(r) &= \frac{1}{\hat{\lambda}_i^t \hat{\lambda}_j^t A^t} \sum_{i_k, j_l \in t} \mathbb{1}_{\{d_{i_k,j_l} < r\}} \\
&= \frac{1}{\hat{\lambda}_i^t \hat{\lambda}_j^t A^t} \sum_{i_k, j_l \in t_R} \mathbb{1}_{\{d_{i_k,j_l} < r\}} + \frac{1}{\hat{\lambda}_i^t \hat{\lambda}_j^t A^t} \sum_{i_k, j_l \in t_L} \mathbb{1}_{\{d_{i_k,j_l} < r\}} \qquad (7) \\
&\quad + \frac{1}{\hat{\lambda}_i^t \hat{\lambda}_j^t A^t} \left[\sum_{i_k \in t_R} \sum_{j_l \in t_L} \mathbb{1}_{\{d_{i_k,j_l} < r\}} + \sum_{i_k \in t_L} \sum_{j_l \in t_R} \mathbb{1}_{\{d_{i_k,j_l} < r\}} \right] \\
&= \frac{A^{t_L}}{A^t} \frac{\hat{\lambda}_i^{t_L} \hat{\lambda}_j^{t_L}}{\hat{\lambda}_i^t \hat{\lambda}_j^t} \hat{K}_{ij}^{t_L}(r) + \frac{A^{t_R}}{A^t} \frac{\hat{\lambda}_i^{t_R} \hat{\lambda}_j^{t_R}}{\hat{\lambda}_i^t \hat{\lambda}_j^t} \hat{K}_{ij}^{t_R}(r) \qquad (8) \\
&\quad + \frac{1}{\hat{\lambda}_i^t \hat{\lambda}_j^t A^t} \left[\sum_{i_k \in t_R} \sum_{j_l \in t_L} \mathbb{1}_{\{d_{i_k,j_l} < r\}} + \sum_{i_k \in t_L} \sum_{j_l \in t_R} \mathbb{1}_{\{d_{i_k,j_l} < r\}} \right]
\end{aligned}$$

This leads to the natural definition of the impurity loss $\Delta I_{ij}(s, t, r)$ as the variation of heterogeneity, i. e. the variation of interaction between marks i and j, coming from the split of node t using s, at radius r:

$$\Delta I_{ij}(s, t, r) := \hat{K}_{ij}^t(r) - \alpha_s \frac{\hat{\lambda}_i^{t_L} \hat{\lambda}_j^{t_L}}{\hat{\lambda}_i^t \hat{\lambda}_j^t} \hat{K}_{ij}^{t_L}(r) - (1 - \alpha_s) \frac{\hat{\lambda}_i^{t_R} \hat{\lambda}_j^{t_R}}{\hat{\lambda}_i^t \hat{\lambda}_j^t} \hat{K}_{ij}^{t_R}(r), \tag{9}$$

where $\alpha_s = \frac{A^{t_L}}{A^t}$. The intensity ratios are similar to marks' proportions within each child node in a spatial context, leading to a classical form for $\Delta I_{ij}(s, t, r)$ as an impurity loss, up to the

area factor α_s. Nonetheless, this area factor is natural when dealing with spatial data since it leads to reweight properly the impurity of the two nodes t_R and t_L.

Finally, this choice of impurity loss is convenient since it can be rewritten as (see (8) and (9)):

$$\Delta I_{ij}(s,t,r) = \frac{1}{\hat{\lambda}_i^t \hat{\lambda}_j^t A^t} \left[\sum_{i_k \in t_R} \sum_{j_l \in t_L} \mathbb{1}_{\{d_{i_k,j_l} < r\}} + \sum_{i_k \in t_L} \sum_{j_l \in t_R} \mathbb{1}_{\{d_{i_k,j_l} < r\}} \right]. \tag{10}$$

$\Delta I_{ij}(s,t,r)$ is positive, which is necessary to define it as impurity loss. In addition, $\Delta I_{ij}(s,t,r)$ is null if and only if the children nodes t_L and t_R are pure at distance r along split s, that is

$$\forall i_k \in t_L, j_l \in t_R \text{ and } \forall i_k \in t_R, j_l \in t_L \quad d_{i_k,j_l} \geqslant r,$$

highlighting splits that do not discriminate marks at all. Hence maximizing $\Delta I_{ij}(s,t,r)$ leads to increasing spatial purity at fixed scale r.

In addition to the positivity of $s \mapsto \Delta I_{ij}(s,t,r)$, which is mandatory, a desirable property is the strict concavity, which would ensure that the best split is unique, and then avoids ties. This is an ingredient of the original CART algorithm but this is not the case here as in most extensions of CART. It should be noted that the growing part of the algorithm still works, that is splitting always purifies nodes even if from an algorithmic point of view, the choice of the split could be arbitrary, without any statistical drawback. However, a possibility, not implemented here, could be to look for a resolution r avoiding ties.

4.2 Description of the algorithm

The usual penalized misclassification error rate (Equation 2) used in CART classification trees is natural to determine areas that are homogeneous in terms of mark. Our focus is to determine areas of the point process that are homogeneous in terms of marks' intensity. To be able to choose a convenient definition for the risk (as defined in Equation 3), we consider the following property of the intensity of the marks: in the case where the marked spatial point process (X, M) is stationary, then the intensity of points with mark $j \in \{1; 2\}$ inside surface $\mathcal{A}$ can be written as

$$\begin{aligned} \Lambda(\mathcal{A}, j) &= \mathbb{P}(X \in \mathcal{A}, M = j) \\ &= \mathbb{P}(M = j \mid X \in \mathcal{A}) \mathbb{P}(X \in \mathcal{A}) \\ &= \mathbb{P}(M = j \mid X \in \mathcal{A}) \lambda A \end{aligned}$$

where $\lambda > 0$ is the intensity of X (without marks), and A is the area of $\mathcal{A}$. Then estimating $\Lambda(\mathcal{A}, j)$ is equivalent to estimate $\mathbb{P}(M = j \mid X \in \mathcal{A})$, since λ can be estimated directly from the

point process. Thus the penalized criterion of Equation 3 used in CART class probability trees is natural to estimate the intensities of the marks: intensity for each mark will then be locally estimated on a tessellation of the plane.

We propose an algorithm using impurity loss ΔI_{ij} defined by (9) to develop the maximal tree T_{max}. The estimator $\hat{K}_{ij}$ of the intertype K_{ij}-function is computed at each node t, the value of $r = r_0$ is fixed as the one for which the estimated intertype K-function is the farthest from the one of random labelling.
Recall that a maximal tree T_{max} is computed via recursive partitioning, and then pruned with the penalized criterion defined in (2) or (3). This produces a decreasing sequence of K subtrees pruned each one from another, denoted by $T_1 \succ \ldots \succ T_K = \{t_1\}$, and associated with a sequence of complexities $0 = \alpha_1 < \ldots < \alpha_K$.

Rephrasing the algorithm in spatial terms, r_0 can be considered as the characteristic scale from the region for which we compute r_0. Then before splitting, we consider for $r = r_0$ the quantity $\Delta I_{ij}(s, t, r)$ and we seek to the best split. After splitting we seek to the best $r \leq r_0$ maximizing $\Delta I_{ij}(\hat{s}, t, r)$. Then, after splitting, the two child are considered in parallel in the same way, recomputing $r_{0,L}$ and $r_{0,R}$. It turns out that $\Delta I_{ij}(s, t, r)$ is decreasing along any branch of the tree. So the reordered sequence of r_i can be used to define a sequence of nested subtrees. Let us remark that, in the algorithm, the value of r_t maximizing the impurity criterion for the best split is set to 0 if t is a leaf, what corresponds to the smallest possible value of r that is the best resolution.

The final step is to choose a convenient tree among the sequence $T_1 \succ \ldots \succ T_K = \{t_1\}$. This choice can be made via cross-validation. Nevertheless, to avoid randomness due to the choice of subsamples, we use two variants of the slope heuristic method proposed by Birgé, Massart in the 2000s (see [2, 4, 13] for more details), and based on the behavior of the number of leaves with respect to the complexity. First, the general idea is that, if a tree is a good predictor, then high energy is necessary to prune it. Hence, the penalty to be chosen in the penalized criterion shall increase a lot before the tree is pruned, leading to a large plateau in the $\alpha_k \mapsto |T_k|$ graph. Second, if there exist plateaus negligible with respect to the largest one, the corresponding trees can be considered as artefacts of the pruning procedure. These trees are then removed from the sequence, and the final tree model is selected through the classical maximal jump in the number of leaves for this new collection of trees. In practice, only jumps occurring before the largest plateau are considered. An artificial example of such a behavior is given in Figure 1, where $\triangle$ represents the tree selected via the modified largest jump method, while $\diamond$ represents the tree selected via the largest plateau method.

The algorithm used to obtain the final tree is detailed in Table 1 for the calculation of the maximal tree, and in Table 2 for the pruning strategy, and the calculation of the nested sequence of

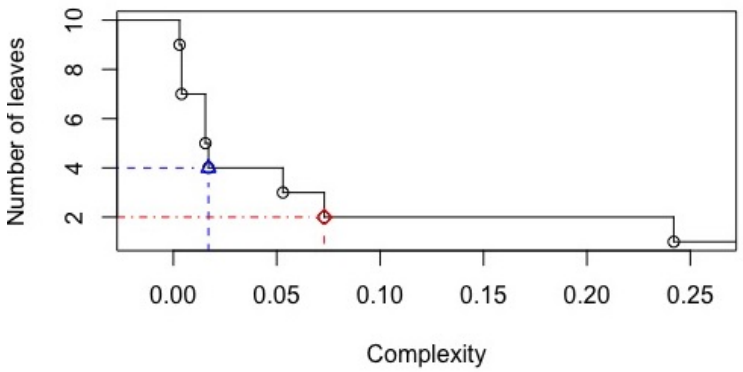

Figure 1: Typical behavior of number of leaves with respect to complexity, $\triangle$ and $\diamond$ represent the two highlighted trees.

pruned subtrees. It should be noted that this pruning step is the same as for original CART, since it is based either on the minimization of penalized misclassification rate or on the minimization of penalized Gini index. Let us emphasize that this sequence of pruned subtrees and the corresponding complexities allow the user to choose the final partition according to the data and the peculiarity of the studied problem. This is of particular interest if the best subtree is the root.

Remark 2. To avoid too strong edge effects, a minimum distance from the boundaries is imposed to candidate splits in the construction of maximal tree.

4.3 The initial resolution r_0

The other parameter of our algorithm is the initial resolution r_0, which is crucial to define the first split (the most important one). One first idea to provide a default value is to take as initial value for r_0 the one corresponding to the maximum of the difference between the estimated K_{ij} and its theoretical value, that is the one for which the dissimilarity with the homogeneous case is maximal. Depending on the criterion configurations, it appears that critical values can be far from this default value and user should be careful about automatic choice of r_0.

	Spatial Classification Trees
Input	Marked point process, scale $r_0 \in]0;1[$, minimal number of observations in node to split `minsplit`,

Maximal tree	`Initialize`,
	node $t = t_1$ the root of the tree containing all observations,
	$n_t = n_{t_1}$ the number of observations in node t,
	$r_t = r_0$ the scale value at node t,
	$\text{argmax}\{\lambda_1^t, \lambda_2^t\}$ the label of node t.
	`While` $n_t >$ `minsplit`,
	`Compute`
	$i_0 = \text{argmax}_{i\in\{1;2\}}\lambda_i^t$, $j_0 = \text{argmin}_{i\in\{1;2\}}\lambda_i^t$,
	$\hat{s} = \text{argmax}_s \Delta I_{i_0 j_0}(s, t, r_t)$,
	`Set`
	$t_L = \{\text{points in } t \mid \text{answer "yes" to } \hat{s}\}$,
	$t_R = \{\text{points in } t \mid \text{answer "no" to } \hat{s}\}$.
	`Recursion`
	$r_t = \text{argmax}_r \Delta I_{i_0 j_0}(\hat{s}, t, r)$,
	`left`: $t = t_L$,
	`right`: $t = t_R$.
	`Output`
	Maximal tree T_{max}.

Pruning	Sequence $(T_k)_{1\leqslant k\leqslant K}$ of subtrees pruned from T_{max}, and complexities $(\alpha_k)_{1\leqslant k\leqslant K}$ (see Table 2).
Selection	`Set` $\hat{k}_1 = \max\left\{k \mid k = \text{argmax}_{0\leqslant j\leqslant K-1}(\alpha_{j+1} - \alpha_j)\right\}$, $\hat{k}_2$ the index of the tree selected via the modified largest jump method,
Output	Trees T_{max}, $T_{\hat{k}_1}$, $T_{\hat{k}_2}$. Sequences $T_1 \succ \ldots \succ T_K = \{t_1\}$ and $0 = \alpha_1 < \ldots < \alpha_K$. Collections r_t and $K_{ij}(r_t)$ for each node t of T_{max}.

Table 1: Spatial Classification Trees (`SpatCART`).

	Pruning Algorithm
Input	Maximal tree T_{max}, number of observations n, risk ρ from misclassification error or Gini index.
Initialization	$k = 1$ and $\alpha_1 = 0$, $T = T_1$ the smallest subtree pruned from T_{max} at complexity α_1, n_f the number of leaves of T,
Compute	for each node t of T n_t the number of observations in t, $\rho(t)$ the local risk of t: $\rho(t) = n^{-1} \sum_{(x_i,m_i)\in t} \mathbb{1}_{\hat{m}_t \neq m_i}$ for misclassification rate, or $\rho(t) = (1 - \sum_{j=1}^{2} \hat{p}(j\|t)^2) n_t/n$ for Gini index $\rho(T_t) = \sum_{\{f \text{ leaf of } T_t\}} \rho(f)$ the risk of the branch T_t issued from t, n_{T_t} the number of leaves of branch T_t.
While	While $n_f > 1$, `compute` $\alpha_{k+1} = \min_{\{t \text{ internal node of } T\}} \frac{\rho(t) - \rho(T_t)}{n_{T_t} - 1}$. `Prune` all branches T_t of T verifying $\rho(T_t) + \alpha_{k+1} n_{T_t} = \rho(t) + \alpha_{k+1}$ `Set` T_{k+1} the pruned subtree obtained in that way. `Set` $T = T_{k+1}$ and $k = k + 1$.
Output	Sequence $(T_k, \alpha_k)_{1 \leqslant k \leqslant K}$.

Table 2: Pruning Algorithm.

5 CART and SpatCART in action: illustration by simulations

An `R` package `spatcart`, and the `R` codes to reproduce experiments of this section, are available on `https://github.com/Servane-Gey/Spatial-classification-trees`. Package `spatcart` may also be directly installed with `R` package `devtools` from the github repository `Servane-Gey/spatcart`. Package `spatcart` requires the following `R` packages to implement the results:

- `spatstat` to deal with point processes, and in particular to compute ΔI_{ij} in the construction of the maximal tree,
- `tree` to deal with tree structures.

Two methods are to be considered: Spatial CART (SpatCART, see Table 1), and CART with

Gini splitting criterion. Class probability trees and classification trees are constructed using either SpatCART or CART. We could expect the following results.
First, since the pruning strategy (see Table 2) is the same for the two methods, based either on the minimization of penalized misclassification rate or on the minimization of penalized Gini index, the major differences between our proposal (Spatial CART) and the existing schemes must be concentrated on the growing step. In addition to the final trees, we also have to look at the comparison of the two maximal trees.
Second, since the new splitting criterion reduces more or less to the classical one in the case of spatial homogeneity, we have to consider some artificial homogeneous examples to illustrate this and, what is more crucial, to define some artificial inhomogeneous examples to capture the interest of our proposition. This second point must be considered because the spatial nature of the data must be taken into account, and we can refer to [17] and [1].

5.1 Illustrative examples

So, the idea here is to check that CART and SpatCART behave as expected on two typical examples: (a) a control example where the Bayes classifier is a classification tree. In this case, the two algorithms could find a partition close to the partition defined by the underlying model. (b) an example where the classification problem is trivial, and where the interaction between marks plays an important role. In this case, SpatCART must produce splits significantly different from those produced by CART. Let us detail the two illustrative simulated examples.

The first example is the `Chess` bivariate Poisson point process on the unit square, with marks simulated from a blue and red chessboard with 9 squares, represented in the left panel of Figure 2. It is an example of spatial homogeneity for which CART and SpatCART could lead to similar results.

The second example is the `Locally repulsive` bivariate Poisson point process on the unit square, with a majority red mark, and a minority blue one. The marked point process is designed by splitting vertically the unit square into two parts (see the right panel of Figure 2): on the left part, the minority blue marked points repulse red ones, with a constant repulsion radius equal to $r = 0.05$; on the right part, blue and red marked points are independently distributed.

In this example, a spatial inhomogeneity is introduced, then CART and SpatCART should lead to different results: CART should miss the difference between the two parts and SpatCART should highlight it.

With these two different scenarios in mind, it remains to analyze the results. But before that, we propose to focus on the kernel of our proposal: the new splitting criterion convenient for detecting a certain kind of spatial heterogeneity and the initial spatial resolution to be selected.

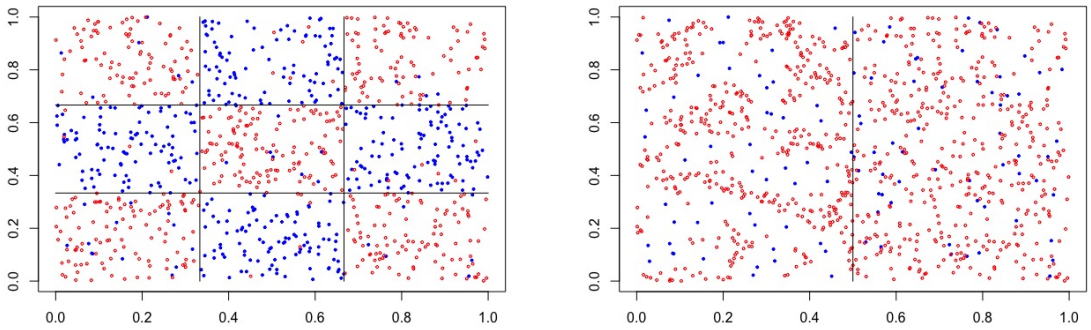

Figure 2: *Left*: `Chess` simulated data set. *Right*: `Locally repulsive` simulated data set.

5.2 Intertype K-function and splitting criterion

The key tool for defining the splitting criterion is the intertype K-function. More precisely the difference between the estimated one and its theoretical value is evaluated in order to select the split. In each cell of the tesselation, the intensity is assumed to be constant, *i.e.* the process is assumed to be homogeneous. Figure 3 presents the `Chess` simulated example: on the left we have the estimated intertype K-function (as a function of r) and its theoretical counterpart and, on the right, the difference between these two functions. In this typical situation, this difference is strictly decreasing and the natural choice is to take r_0 as large as possible.

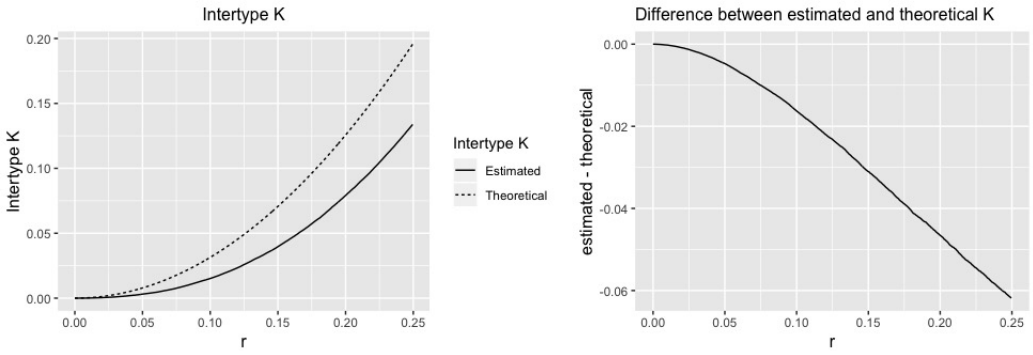

Figure 3: Example of intertype K function and its theoretical value in the homogeneous situation for `Chess` simulated example. *Left* : Intertype K-functions. *Right* : Difference between estimated and theoretical intertype K-functions.

The next object of interest is the impurity function directly connected with the splitting criterion.

Figure 4 illustrates the behavior of the impurity function with respect to first split for `Chess` example, which appears to be clear and expected.

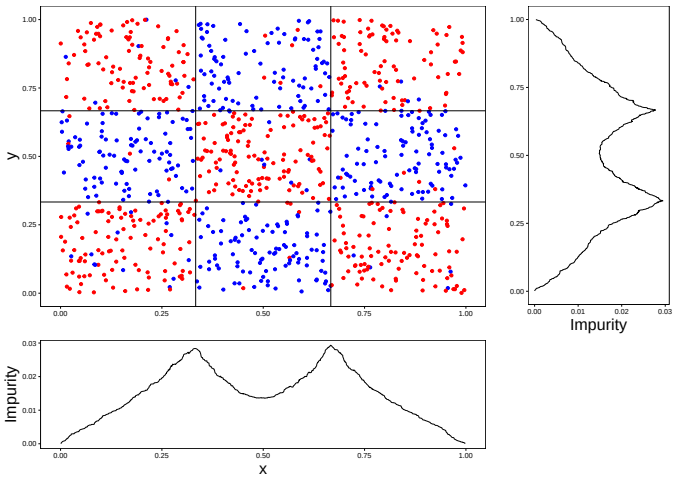

Figure 4: Behavior of impurity with respect to first split for `Chess` example.

5.3 Maximal and optimal partitions

With two different situations depicted by Figure 2, we compare the results of the usual CART strategy (for growing and pruning) and those obtained applying our new proposal.

5.3.1 The homogeneous case

Depending on the run of simulations, the obtained partitions can be roughly the same, or quite different due the ties problem for the split choice, which occurs more frequently in such an example exhibiting high symmetry. Selecting an example for which the maximal trees are similar leads to Figure 5, where SpatCART trees have been obtained from initial resolution $r_0 = 0.25$ equal to the maximal difference between estimated and theoretical intertype K-functions (see Figure 3). Let us remark that the true underlying frontiers are recovered by the CART and SpatCART optimal classification trees, and that the differences appear inside the blocks of the maximal trees.
Examining the corresponding behaviors of the number of leaves as a function of complexity for SpatCART and CART (see Figure 6), this similarity is visible.
So to summarize, as expected, when there is nothing to discover from the spatial viewpoint, SpatCART can behave as the usual CART but in most cases would lead to quite different partitions, due to ties.

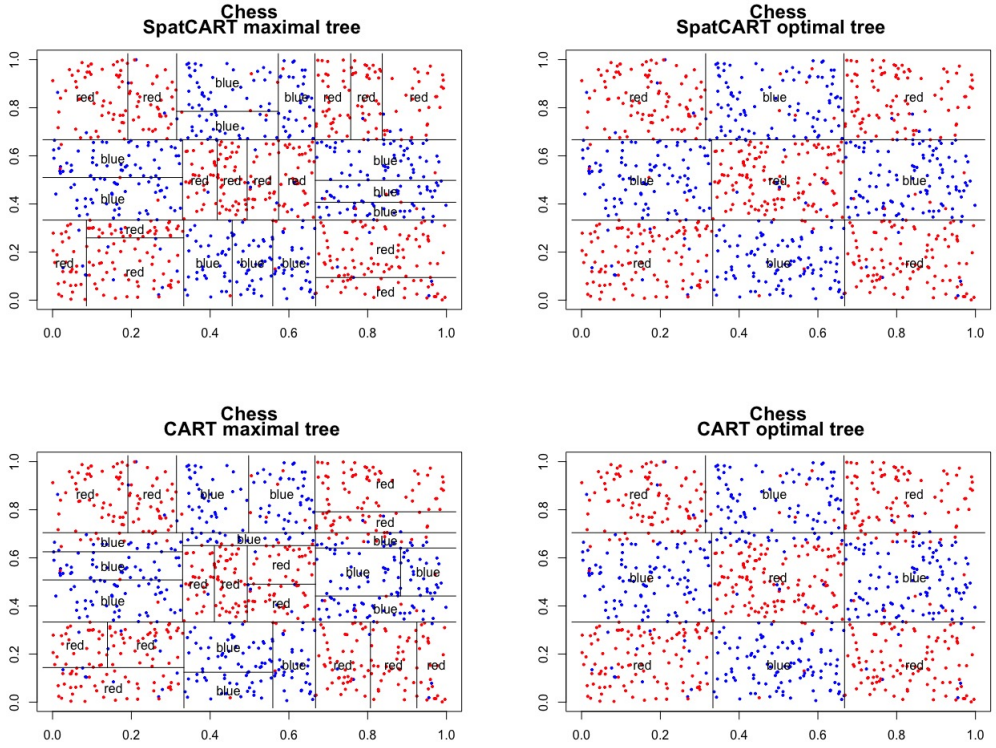

Figure 5: Maximal (*left*) and optimal (*right*) trees for SpatCART (*up*) and CART (*bottom*) on `Chess` data set with initial resolution $r_0 = 0.25$. The pruned subtrees are selected through the largest plateau method (see Figure 6)

5.3.2 The inhomogeneous case

Using roughly the same sequence of plots, let us examine now a more interesting aspect, crucial to illustrate what is the specific contribution of SpatCART.

The first step is to compare the maximal trees. We start with an initial resolution r_0 smaller than the locally repulsion scale $r = 0.05$. Indeed, since the domain consists of a left part where the points of the minority mark push back those of the majority mark to at least 0.05 units, and a right part where the two marks are distributed independently, therefore, to detect this difference, it is necessary to adopt a sufficiently small resolution, which corresponds to values of r_0 lower than the repulsion threshold.

Figure 7 contains the SpatCART (top) and the CART (bottom) trees on `Locally repulsive` data set. The partitions are extremely different. More precisely, since the initial r_0 is small enough, SpatCART does not split the left half of the square, and the right half is split iteratively

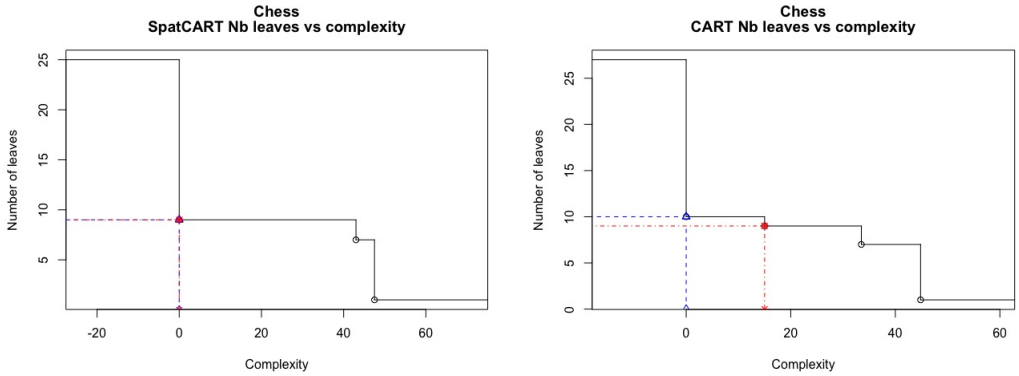

Figure 6: Behavior of number of leaves versus complexity for SpatCART and CART on `Chess` data set. For each case, $\triangle$ represents the tree selected via the modified largest jump method, while $\diamond$ represents the tree selected via the largest plateau method.

according to the direction of the first split. This is expected since the precision of the estimation of the intensities is directly related to the number of points and the intensity of the rarer is too low to provide a reliable estimate. A solution could be to impose a minimum number of points for each mark to allow splitting. We have not implemented it since pruning will remove these artificial splits. Note that if the gain of impurity is the same for both direction, the x direction is chosen. On the other hand, CART splits according to the empirical distribution of the blue points, leading to a very different partition. So the generated maximal partitions are, as expected, different in the inhomogeneous situation.
The corresponding sequences of complexities for class probability trees are given by Figure 8. A second difference appears: SpatCART seems to generate less splits to provide partitions similar to CART in terms of spatial repartition, generating less false alarms (useless spurious splits).
To end this analysis, let us mention that in this case, results using the pruning strategy are convenient: the selection based on misclassification rate leads to root trees. This is logical since the percentage of blue marks is too low to be recovered. Here, the difference is clear: since there is no heterogeneity in the marginal structures of the response distribution, nothing is inferred using CART.

To summarize, SpatCART seems to provide better partitions in terms of marks spatial repartition (see the maximal trees figures) and SpatCART needs less splits to provide partitions similar to CART in terms of spatial repartition, suggesting that it avoids false alarms in a better way.

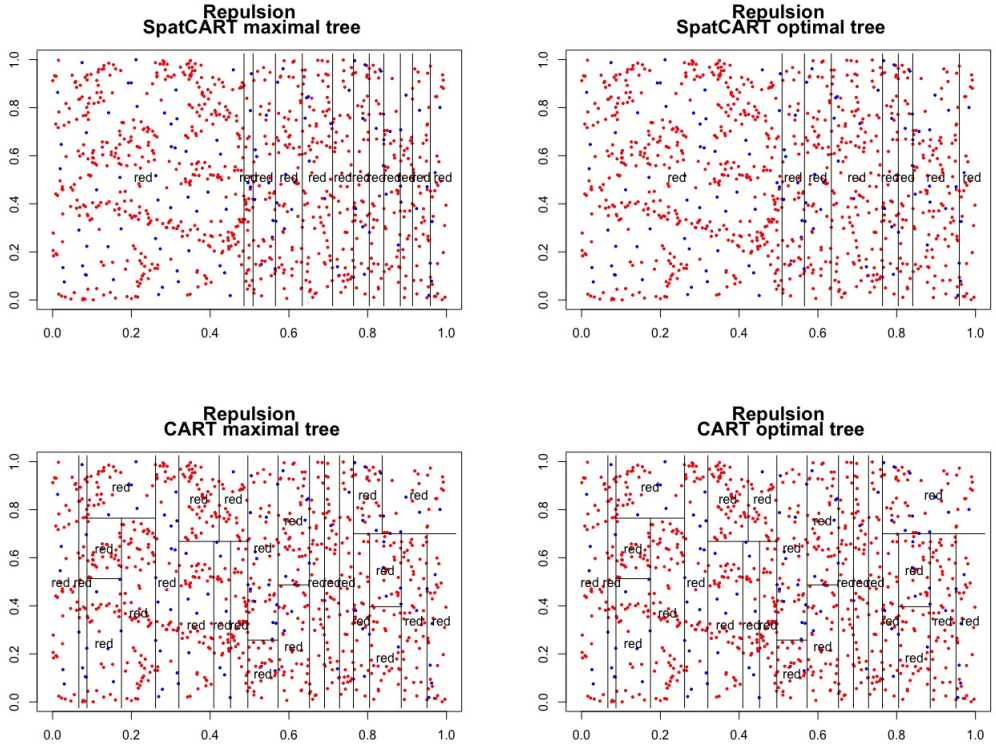

Figure 7: Maximal (*left*) and optimal (*right*) trees for SpatCART (*up*) and CART (*bottom*) on `Locally repulsive` data set, with initial resolution r_0 smaller than the locally repulsion scale $r = 0.05$.

6 Example

6.1 Data

We applied these methods to a tropical rain-forest located at Paracou, 40 km west from Kourou in French Guiana (5°15'N, 52°55'W). It is an experimental site that is devoted to studying the effects of logging damage on stock recovery. Twelve plots (6.25 hectares each) of undisturbed forest were settled in 1984. On each plot, the circumference of every tree with a DBH (diameter at breast height) greater than 10 cm was measured with a precision of 0.5 cm. Its spatial coordinates ($\pm$ 50 cm) and its specie were noted too. Measurements have been carried annually from 1984 to 1995, and once every two years since. A more precise description of the Paracou plots may be found in [16] and the data can be freely downloaded from the website `https://paracou.cirad.fr/`. Experiments on the Paracou data set may be reproduced with the `Paracou` function from `R` package `spatcart`.

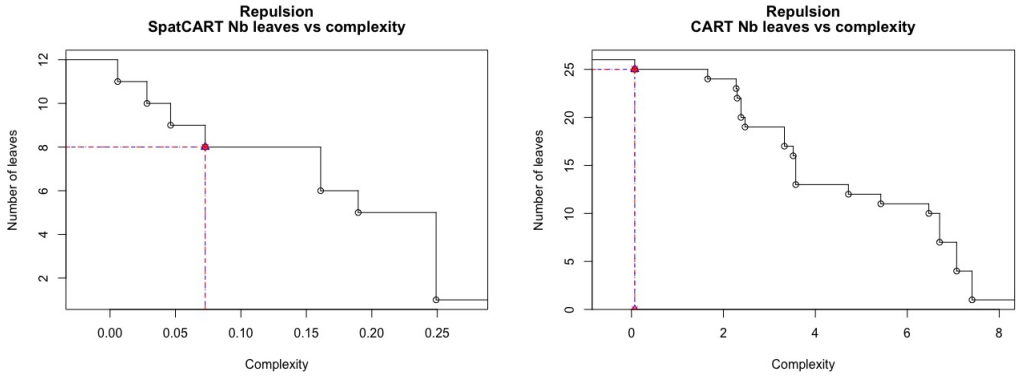

Figure 8: Behavior of number of leaves versus complexity for SpatCART and CART on `Locally repulsive` data set. For each case, △ represents the tree selected via the modified largest jump method, while ◇ represents the tree selected via the largest plateau method.

We focus on two species:

- *Vouacapoua americana* is a hermaphroditic shade tolerant tree specie of mature tropical rainforests and whose distribution spans the eastern part of the Guiana shield. Its local density averages around 10 individuals per hectare for trees with greater than 10 cm DBH, but this varies greatly because of spatial clustering. Individuals are clustered in large patches of a few hectares that are mainly located on hill tops and slopes [21].

- *Oxandra asbeckii* is a specie of the understorey, the largest individuals of which measure less than 15 cm DBH. It is a shade tolerant specie [12]. It shows animal dispersal: the seeds are dispersed after passage through the vertebrate gut. *Oxandra asbeckii* is located on hill tops and slopes.

This explains the spatial distribution for the two species displayed on Figure 11. The two species *Vouacapoua americanaa* and *Oxandra asbeckii* were selected at Paracou because their spatial distribution is linked to the relief: they are both located on hill tops and slopes. Elevation is the environmental factor that drives their spatial distribution and this creates a strong interaction between both repartitions. We focus on one plot represented in Figure 11. The figure also gives the contour lines in order to be able to determine hill tops and slopes of the plot. Seventy trees of *Vouacapoua americana* and eighty trees of *Oxandra asbeckii* were referenced for this plot.
The data consists of seventy lines (one per tree) and four columns: the 3-D coordinates (longitude, latitude and elevation) as well as the specie indication.

6.2 Results and discussion

From the representation of the estimated and theoretical intertype K-functions and their difference evaluated on the Paracou data set (see Figure 9), we can see three regimes: one before the scale value $r = 6$ (blue dashed line on the right of Figure 9), where there is no interaction between species; one between scale values $r = 6$ and $r = 24$ (red dashed line on the right of Figure 9) where species begin to interact; and one after the scale value $r = 24$ where the interaction between species increases rapidly. From an ecological point of view, $r = 24$ can be interpreted as the limit of seeds dispersal (see [21]). Hence, we choose the initial median scale value $r_0 = 15$ to proceed with SpatCART, in order to be sure to catch a sufficiently large scale to capture interaction, and not too large to avoid deeper maximal trees.

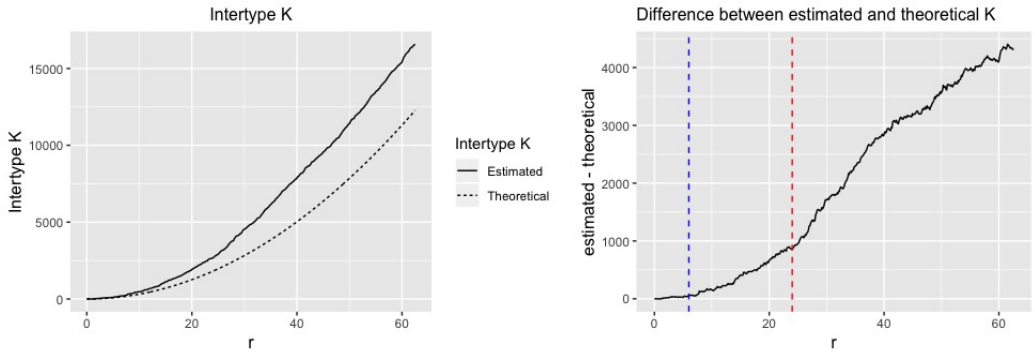

Figure 9: Intertype K function and its theoretical value in the homogeneous situation for the Paracou data set. *Left* : Intertype K-functions. *Right* : Difference between estimated and theoretical intertype K-functions; `blue`: scale $r = 6$, `red`: scale $r = 24$.

To better interpret the regions highlighted by the algorithm in terms of interaction between marks, we color each region of the space (each leaf t) using a monotonic gray colormap based on the value of $\hat{K}_{ij}^{t}(r_t)$ (see equation 6). Thus Figure 10 colouring the pieces of the partition using this interaction-based colormap, highlights the presence of *Oxandra asbeckii* at the hill of left top of the plot as well as the competition between both species for the hill at the bottom of the plot. This representation magnifies the regions of high interaction, showing that SpatCART recovers the spatial structure and that the interaction-based grey colormap is meaningful. In addition, the colormap based on interaction allows to visually appreciate the difference between two adjacent regions. This opens a very natural post-processing by ignoring useless frontiers and provides more interpretable tessellations.

The class probability tree has more leaves and if both resulting partitions are coherent, even exactly the same from the right side of the window space, an interesting difference must be noted. On the right side of the window space, the class probability tree partition provides a

more detailed segmentation (possibly induced by the slope of the hill) but cleaning the two partitions according to the interaction-based colormap leads to highlight the same areas.

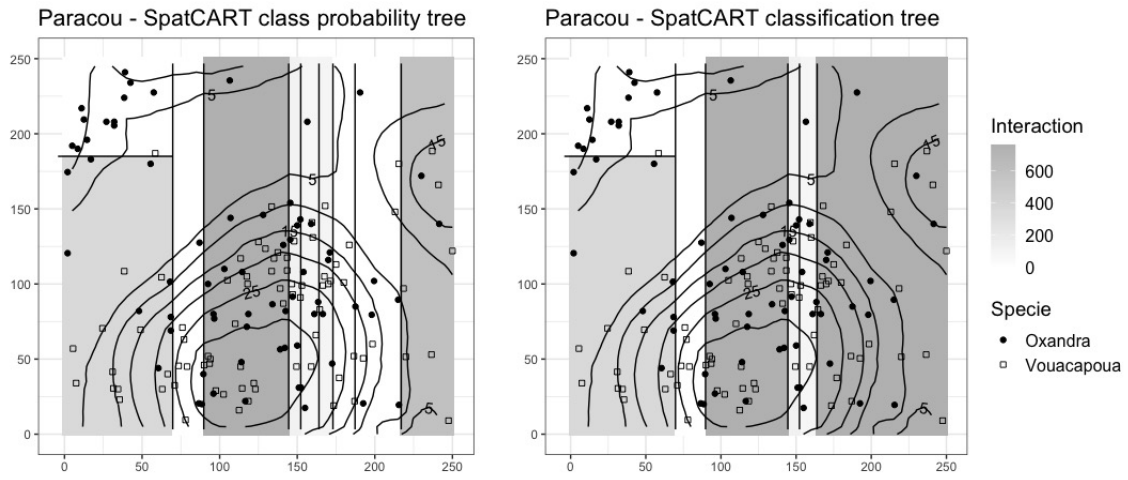

Figure 10: SpatCART class probability (*left*) and classification (*right*) optimal trees on the Paracou data set with initial resolution $r_0 = 15$. The grey colored *Interaction* between marks is the estimation of the intertype K-function restricted to each leaf of the tree as defined by (6).

A contrario, CART results are really poor with only two leaves, for the modified largest jump variant (left panel of Figure 11). Basically it separates the hill at the bottom of the plot from the rest but cannot catch the mixed structure of species along the hill. The spatial structure as well as the ecology of the two species on this plot cannot be inferred from CART results. CART results using the largest plateau variant (see the right panel of Figure 11) are not more informative for the spatial viewpoint, highlighting the regions only according to the most prevalent specie.

From an ecological perspective, few points can be highlighted. At first, elevation is a key variable to describe interactions between these two species. Even if this two species are shade tolerant, competition for light is a major determinant of forest composition. Hills intercept a greater portion of incoming solar radiation and thus, generate contrasting strategies between species to invest a greater portion of their resources in expanding their crowns. The result is not surprising but the method allows to determine more precisely the interaction areas.
On the other hand, below-ground competition for water is also an important driver of tree growth. Slope does not store water and it is a strong disadvantage during drought events. Competition for water among neighbouring trees clearly lead an advantage to *Vouacapoua americana* compared to *Oxandra asbeckii*.

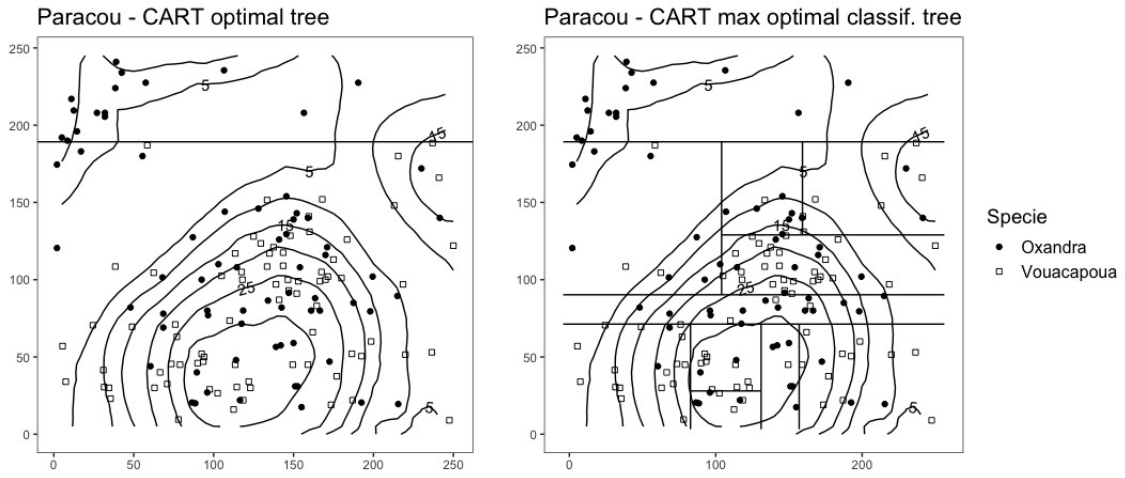

Figure 11: CART common optimal class probability and largest plateau optimal classification tree (*left*), and modified largest jump variant optimal classification tree (*right*) on the Paracou data set.

Finally, we examine the sensitivity of SpatCART trees to the choice of the initial resolution r_0 through the behaviour of the number of leaves of maximal and optimal SpatCART trees (the two variants) when r_0 varies (see Figure 12). To facilitate the comparison with the problem-driven choice of the scale parameter used previously, vertical lines are added at the particular values $r_0 = 6$, $r_0 = 24$ extracted from the behaviour of the estimated intertype K-function (Figure 9), and $r_0 = 15$ used to produce the trees represented in Figure 10.
Figure 12 highlights that SpatCART naturally finds few interaction when r_0 is too small, leading to optimal trees similar to the maximal one. When r_0 becomes larger than 24, the numbers of leaves of the three trees become stable, with similar optimal trees. Let us mention that, for these large initial resolution, SpatCART tends to split always along the same coordinate. Let us remark that between $r_0 = 6$ and $r_0 = 24$, the optimal trees can be very different for a given value of r_0, with a very unstable behaviour, while the maximal one seems to stabilize around $r_0 = 13$. The value $r_0 = 15$ is one of the particular cases where optimal trees are very different. Without giving any procedure to choose r_0, to select a value in the interval where the two variants curves are different could be an indication which seems to be consistent with the ecologically chosen analysis carried out at the beginning of the section.

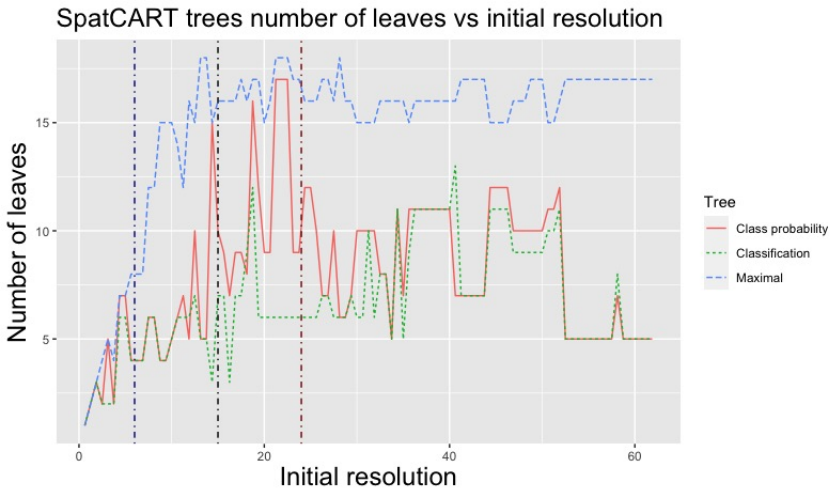

Figure 12: Number of leaves of the SpatCART maximal, optimal class probability and optimal classification trees with respect to initial resolution r_0 on the Paracou data set. The three dotted lines represent respectively the thresholds $r_0 = 6$ (*left*) and $r_0 = 24$ (*right*) proposed in Figure 9, and the value $r_0 = 15$ used to produce trees represented in Figure 10.

7 Conclusion and perspectives

For a bivariate marked spatial point process, we extend CART to segment the space into homogeneous areas for interaction between marks. The original CART constructs homogeneous zones with respect to the distribution of the variable of interest (here, the mark) conditionally to the explanatory variables (here, the position in space). By modifying the splitting criterion in the CART algorithm, using an empirical version of the intertype K-function, we obtain a new procedure, called SpatCART, adapted to the problem at hand. The intertype function itself depends on a parameter r_0 which must be carefully chosen: not to set it once and for all, but rather to start from a rather large value (at the root of the tree) and gradually decrease it as the tree is built by SpatCART. The proposed variant is a way to explore spatial data as a bivariate marked point process using binary classification trees.

This paper is strongly dependent on the binary case, some extensions to handle the multiclass case can be sketched. One possibility is to adopt the strategy used to handle multiclass Support Vector Machines (SVMs) which are intrinsically two-class classifiers (see [15]). A technique widely used in practice is to build one-versus-rest classifiers. Then we could obtain several tessellations and select the final partition by taking into account some criterion maximizing some global measure of heterogeneity between cells, which could be related to the $K_{ij}(r_*)$ for a suitably chosen scale r_*.

Another possible extension is to modify, not only the growing step of the CART algorithm, but also the pruning strategy in order to drive it by point processes properties. This requires to define some additive measure of heterogeneity of the partitions. An intermediate solution could be to use a classical pruning step to simplify and avoid spurious useless splits but the final choice could be to select the final tree among the nested sequence of tree by maximizing the same kind of heterogeneity criterion mentioned before.

Instability is one of the main drawbacks of CART and many classical ways to contain it are available. Ensemble methods like bagging, boosting and random forests (see [15] for example) can then be similarly defined in the spatial case but the aggregation part should be defined. To introduce and experiment a bagging-like scheme could be of interest for a future work. Note that the sensitivity of CART with respect to rotation could also be used as an alternative way to generate several tessellations.
Using such an idea, we could imagine to develop a variant selecting randomly, in case of ties, the splits on each direction to choose, the best candidate per direction and if necessary to randomly select the direction. Such a scheme would lead to equivalent but different partitions which could finally be combined.

Another possible extension is to incorporate covariables. In the specific forestry example, it is possible to think about the introduction of the elevation as a third spatial coordinate. Nevertheless, it is not straightforward to estimate intensity of a point process in higher dimension. Alternatively and more generally, we could imagine to incorporate covariables using a three-steps procedure. First by performing a classical CART using only the additional covariables and then, in each leaf, to apply a SpatCART on spatial variables and finally select the best space tessellation or aggregate them.

`R` package and source codes

An `R` package `spatcart`, and the `R` codes to reproduce experiments of sections 5 and 6, are available on the repository `https://github.com/Servane-Gey/Spatial-classification-trees`. Package `spatcart` may also be directly installed with `R` package `devtools` from the `github` repository `Servane-Gey/spatcart`.

Acknowledgements

The authors thank the Associate Editor and two anonymous referees for their valuable comments which led to a considerable improvement of the presentation.

References

[1] L Anselin and A Getis. Spatial statistical analysis and geographic information systems, *in Perspectives on spatial data analysis, Springer* 35–47, 2010

[2] S Arlot. Minimal Penalty and the Slope Heuristic: A Survey (with discussion), *in Journal de la Société Française de Statistique* 160(3), 1–106, 2019

[3] A Baddeley, J Moller and R Waagepetersen. Non- and semiparametric estimation of interaction in inhomogeneous point patterns, *Statistica Neerlandica* 54, 329-350, 2000

[4] JP Baudry, C Maugis and B Michel. Slope heuristics: overview and implementation, *Statistics and Computing*, 22(2), 455-470, 2012

[5] A Bar-Hen and N Picard. Simulation study of dissimilarity between point process *Computational Statistics*, 21(3-4):487-507, 2006

[6] L Bel, D Allard, JM Laurent, R Cheddadi and A Bar-Hen. CART algorithm for spatial data: application to environmental and ecological data, *Computat. Stat. and Data Anal.*, 53(8):3082-3093, 2009

[7] L Breiman, JH Friedman, RA Olshen and CJ Stone. *Classification and regression trees.* Chapman & Hall, 1984

[8] HA Chipman, E George, JM Laurent and RE McCulloch. BART: Bayesian additive regression trees, *Annals of Applied Statistics*, 4(1):266–298, 2010

[9] N Cressie. *Statistics for Spatial Data*, John Wiley & Sons, New York, 1991

[10] PJ Diggle and AG Chetwynd. Second-order analysis of spatial clustering for inhomogeneous populations, *Biometrics* 47:1155–1163, 1991

[11] PJ Diggle and RK Milne. Bivariate Cox processes: some models for bivariate spatial point patterns, *Journal of the Royal Statistical Society, Series B* 45:11–21, 1983

[12] V Favrichon. Classification des espèces arborées en groupes fonctionnels en vue de la réalisation d'un modèle de dynamique de peuplement en forêt guyanaise. *Rev. Ecol.*, 49, 379–403, 1994

[13] S Gey and E Lebarbier. Using CART to detect Multiple Change Points in the Mean, *Preprint in Statistics and System Biology* 12, HAL 00327146, 2008.

[14] M Loecher and K Ropkins, Model-based boosting in R: a hands-on tutorial using the R package mboost, *Journal of Statistical Software* 63(4):1–18, 2015

[15] T Hastie, R Tibshirani and J Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer, 2009, Second edition.

[16] S Gourlet-Fleury, JM Guehl and O Laroussinie (eds). Ecology and Management of a Neotropical Rainforest: Lessons Drawn from Paracou, a Long-term Experimental Research Site in French Guiana. Paris: Elsevier, 2004

[17] R Haining. Bivariate correlation with spatial data, *Geographical Analysis* 23(3):210–227, 2014

[18] B Hofner, B Mayr, N Robinzonov and M Schmid, Model-based boosting in R: a hands-on tutorial using the R package mboost, *Computational statistics* 29(1-2):3–35, 1991

[19] HW Lotwick and BW Silverman. Methods for analysing spatial processes of several types of points, *Journal of the Royal Statistical Society, Series B*, 44(3):406–413, 1982

[20] BD Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 172-212, 1977.

[21] S Traissac. Dynamique spatiale de Vouacapoua americana (Aublet), arbre de forêt tropicale humide à répartition agrégée. PhD Thesis. Université Claude Bernard-Lyon 1, Lyon, 2003

[22] N Umlauf, N Klein and A Zeileis, BAMLSS: Bayesian additive models for location, scale, and shape (and beyond), *Journal of Computational and Graphical Statistics, 27(3):612–627, 2018*

[23] M Wagner and A Zeileis. Heterogeneity and spatial dependence of regional growth in the EU: A recursive partitioning approach, *German Economic Review*, 20(1):67–82, 2019