Spatial CART Classification Trees

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

[1] Univ. Paris Descartes & LMO, Univ. d'Orsay, Paris Saclay, France
[2] CNAM, Paris, France
[3] MAP5, UMR 8145, Univ. Paris Descartes, France

Univ. de la República, Montevideo, Feb. 2020

Introduction

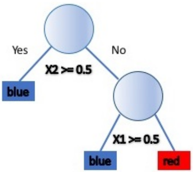
- Classification And Regression Trees, *Breiman et al. (1984)*
- Variants and extensions of the original CART to the spatial domain
  - Ortho-CART Donoho et al. (1997), in image processing, dyadic splits + pruning using the algorithm used for the wavelet packets best basis
  - Dyadic-CART, ideas generalized in Blanchard et al. (2007)
  - Extension to spatial data with kriging type ideas see Bel et al. (2009)
- *Our variant: Spatial CART*
  - For spatial data, *extend CART for bivariate marked point processes*.
  - New splitting criterion in Spatial CART, taking into account the spatial information, to propose a *segmentation of the window into homogeneous areas for interaction between marks*.

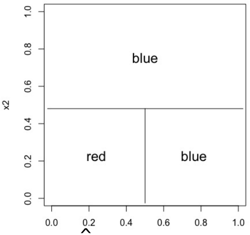
Outline

Classification Trees

- Predict the unknown binary label $Y \in \{0;1\}$ of an observation $X \in \mathbb{R}^p$ via a classifier
$$f : \mathbb{R}^p \rightarrow \{0;1\}$$
- Bayes classifier: minimizer of $f \mapsto \mathrm{P}f := \mathrm{P}\left(Y \neq f(X)\right)$ (with $(X,Y) \sim \mathrm{P}$)

$$f^* = \mathbb{1}_{\eta(x) \geqslant 1/2}, \quad \text{with } \eta(x) = \mathrm{P}(Y = 1 \mid X = x)$$





- $\hat{f}_T = \sum_{t \in \widetilde{T}} \hat{Y}_t \mathbb{1}_t$, $\widetilde{T}$: set of leaves of T, $\hat{Y}_t$: majority vote in the node t

CART Algorithm

Classification And Regression Trees, *Breiman et al. (1984)*

Growing step

- recursive partitioning by maximizing a local decreasing of heterogeneity often based on Gini index or Shannon entropy
- do not split a pure node or a node containing few data
- $\Rightarrow$ maximal tree T_{max}
- T_{max} overfits the data

Pruning step

- Optimal tree: subtree pruned from T_{max}
- Reduce the number of tree candidates: minimize

$$crit_\alpha(T) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\hat{f}_T(X_i)\neq Y_i} + \alpha\frac{|\widetilde{T}|}{n},$$

$|\widetilde{T}|$ = number of leaves of T.

- $\Rightarrow$ sequence $(T_k)_{1\leqslant k\leqslant K}$

CART Algorithm (2)

Classification And Regression Trees, *Breiman et al. (1984)*

Theorem (*Breiman et al. 84*)

For all $\alpha \geqslant 0$, $\operatorname{argmin}_{T \preceq T_{max}} \text{crit}_\alpha(T)$ belongs to the sequence of nested pruned subtrees $(T_k)_{1 \leqslant k \leqslant K}$.

Selection step (Hold Out)

- Data split into a training set $\mathcal{L}$ of size n, and a test set $\mathcal{T}$ of size n_t
- Build $(T_k)_{1 \leqslant k \leqslant K}$ on $\mathcal{L}$ and select

$$\hat{k} = \underset{1 \leqslant k \leqslant K}{\operatorname{argmin}} \frac{1}{n_t} \sum_{(X_i, Y_i) \in \mathcal{T}} \mathbb{1}_{\hat{f}_{T_k}(X_i) \neq Y_i}$$

- $\Rightarrow$ Final CART tree is given by $\hat{f}_{T_{\hat{k}}}$

Outline

Spatial CART Algorithm

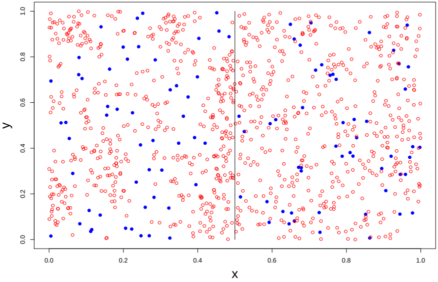
General idea:

- build a tessellation of the window into homogeneous areas for interaction between marks,
- use the spatial information to build a classification tree on the observed points of the bivariate point process.

Spatial CART as a variant of CART:

- Variant in growing: splitting criterion taking into account the spatial characterization of the data, based on the intertype function K_{ij}.
- Variant in pruning: penalized criterion based on least squares criterion to estimate local mark intensities.
- Variant in final selection: optimal tree selected by a variant of the slope heuristic (*Massart et al.*) to keep the spatial information.

Bivariate spatial point process



Observation = realization of (X, M)

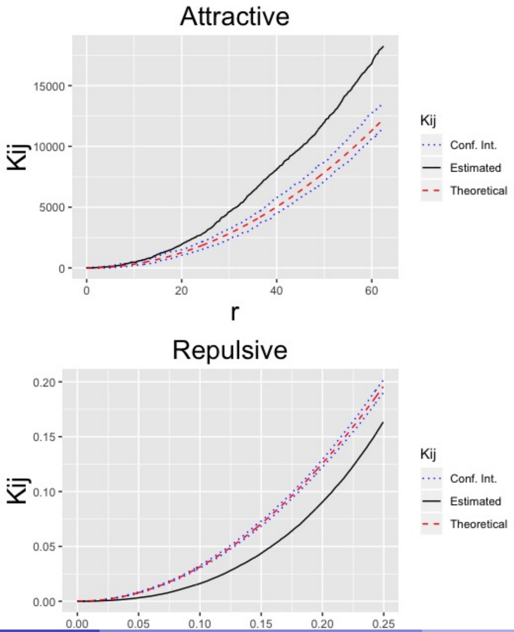
- Left part: blue points repulse red ones (r = 0.05)
- Right part: blue and red points are independently distributed
- Same marginal distribution (color) on left and right side

- **Bivariate spatial point process**: $(X, M) \in W \times \{i; j\} \sim \mathrm{P}$, $W \subset \mathbb{R}^2$.
- **Mark intensity**: for $\star = i, j$, $\lambda_\star$ intensity of the spatial point process $(X \mid M = \star)$.
- **Intertype function**: at scale $r \geqslant 0$,

$$K_{ij}(r) = \lambda_j^{-1}\mathbb{E}\left(N_{ij}(r)\right),$$

where $N_{ij}(r)$ counts the number of type j points at distance at most r of a randomly chosen point of type i.

Interaction Examples



Splitting criterion

At each node t define

- node area: A^t,
- estimates of mark intensities: $(\hat{\lambda}_i^t, \hat{\lambda}_j^t)$,
- estimate of $K_{ij}(r)$: $\hat{K}_{ij}^t(r) = (\hat{\lambda}_i^t \hat{\lambda}_j^t A^t)^{-1} \sum_{\{i_k, j_l \in t\}} \mathbb{1}_{d_{i_k,j_l} < r}$,

  d_{i_k,j_l} Euclidean distance between individuals i_k of mark i and j_l of mark j.

Impurity function: for a node t, a splitting s of t into t_L and t_R, and $r > 0$

$$\Delta I_{ij}(s,t,r) := \hat{K}_{ij}^t(r) - \alpha_s \frac{\hat{\lambda}_i^{t_L} \hat{\lambda}_j^{t_L}}{\hat{\lambda}_i^t \hat{\lambda}_j^t} \hat{K}_{ij}^{t_L}(r) - (1 - \alpha_s) \frac{\hat{\lambda}_i^{t_R} \hat{\lambda}_j^{t_R}}{\hat{\lambda}_i^t \hat{\lambda}_j^t} \hat{K}_{ij}^{t_R}(r) \geqslant 0 \,,$$

with $\alpha_s = A^{t_L}/A^t$ the area proportion of $t_L \subset t$.

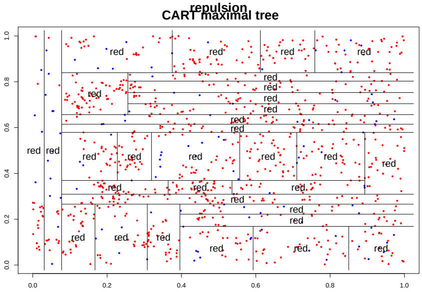
Splitting rule of node t at fixed scale $r > 0$

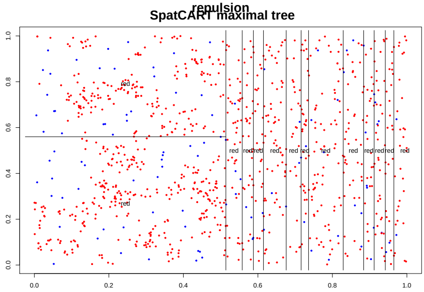
$$\hat{s}(t,r) = \operatorname*{argmax}_s \Delta I_{ij}(s,t,r)$$

Growing maximal tree T_{max}

Input	Bivariate spatial point process, scale r_0,
Initialize	node $t = t_1$ the root of the tree, $r_t = r_0$ the scale value at node t, $\mathrm{argmax}\{\hat{\lambda}_i^t, \hat{\lambda}_j^t\}$ the label of node t.
Recursion	at node t `Compute` $i_0 = \underset{\star \in \{i;j\}}{\mathrm{argmax}}\ \hat{\lambda}_\star^t$, $j_0 = \underset{\star \in \{i;j\}}{\mathrm{argmin}}\ \hat{\lambda}_\star^t$, $\hat{s} = \underset{s}{\mathrm{argmax}}\ \Delta I_{i_0 j_0}(s, t, r_t)$, `Set` $t_L = \{$points in t | answer "yes" to $\hat{s}\}$, $t_R = \{$points in t | answer "no" to $\hat{s}\}$. $r_t = \underset{r}{\mathrm{argmax}}\ \Delta I_{i_0 j_0}(\hat{s}, t, r)$, `left`: $t = t_L$, `right`: $t = t_R$.
Output	Maximal tree T_{max}.

CART *(left)* and Spatial CART *(right)* maximal trees





Spatial CART: initial scale $r_0 < r_{repuls} = 0.05$.

Penalized criterion

Class Probability Trees *Breiman et al. 84*

- If X is locally stationary, estimating local mark intensities amounts to estimating local mark rates.

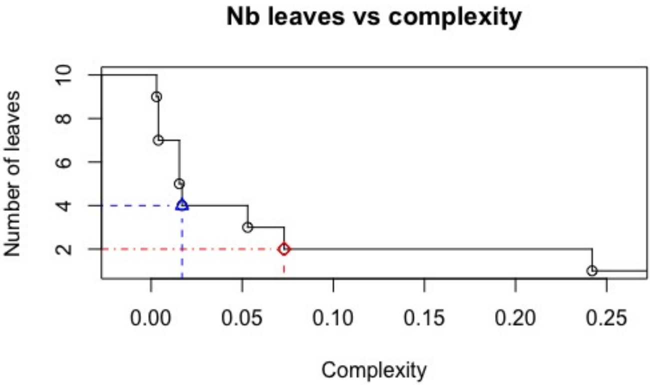
- Use penalized criterion derived from Gini index to prune T_{max}

$$\mathrm{crit}_{\alpha}^{G}(T) \;=\; \frac{1}{n}\sum_{t\in\widetilde{T}} n_t\left(1-\sum_{\star=i,j}\hat{p}(\star|t)^2\right)+\alpha\frac{|\widetilde{T}|}{n},$$

where n = number of observed points; n_t = number of points falling in node t; $\hat{p}(\star|t)$ proportion of points of type $\star$ in node t.

- $\Rightarrow$ sequence of nested pruned subtrees $(T_k)_{1\leqslant k\leqslant K}$.

Final tree selection

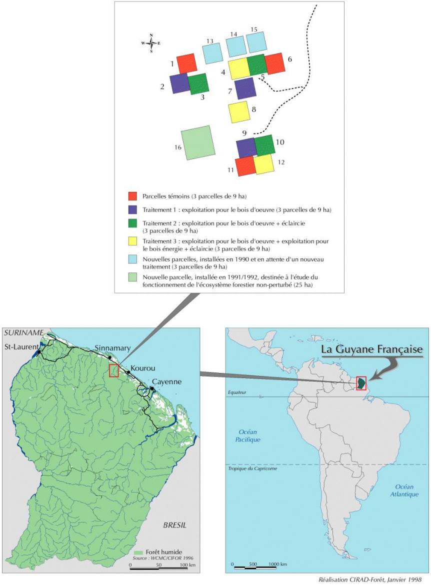


- Identifying the "largest complexity plateau" (red circle) or the modified "largest dimension jump" (blue triangle), more agressive over-penalizing.
- Related to the slope heuristic proposed by Birge, Massart in the 2000s (see Baudry et al. (2012) for a recent survey).

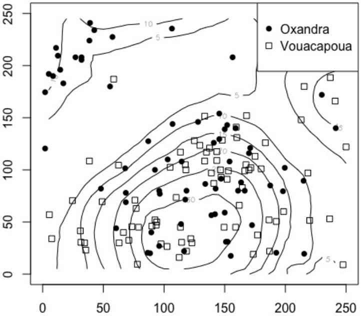
Outline

Data description (Gourlet-Fleury et al. 2004, Traissac 2003)

DISPOSITIF EXPERIMENTAL DE PARACOU



Rain-forest in Paracou: focus on two species



- Two tree species: *Vouacapoua americana* and *Oxandra asbeckii*
- Elevation is the environmental factor that drives their spatial distribution and this creates a strong interaction between both repartitions.
- Competition is *high* for the hill at the bottom of the plot and very *low* at the top left of the plot.

Choice of initial scale r_0

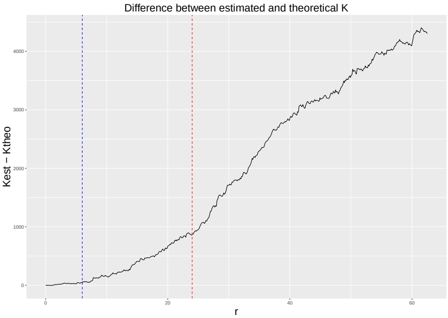
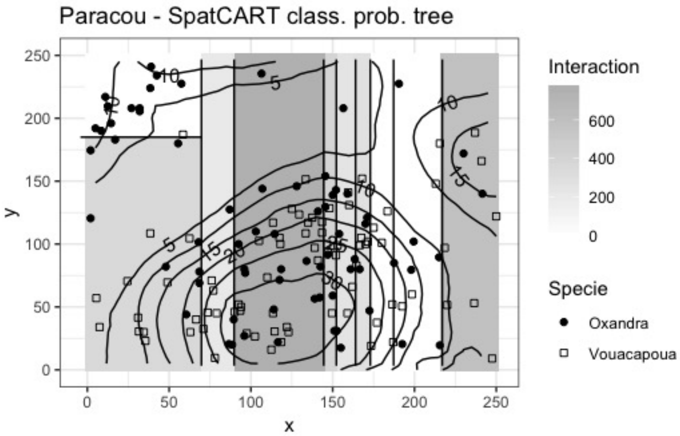


Figure: Difference between estimated and theoretical K_{ij}; blue: $r = 6$, red: $r = 24$.

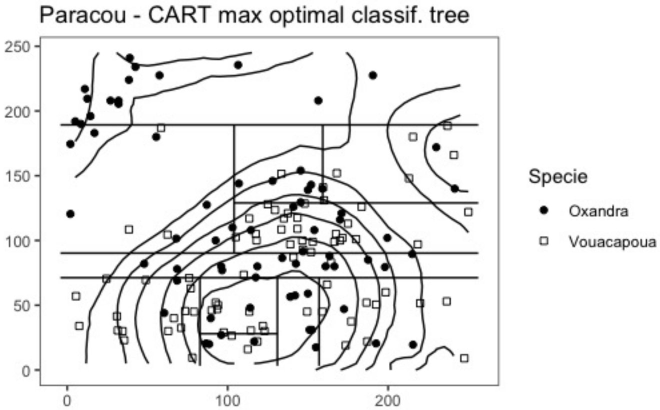
- $r = 6$: species begin to interact
- $r = 24$: the interaction between species increases rapidly.
- Initial median scale value $r = 15$ for SpatCART is sufficiently large to capture interaction, and not too large to avoid deeper maximal trees.

Spatial CART partition



- SpatCART (with $r = 15$) recovers the spatial structure and the interaction-based (on the $K_{ij}(r_t)$ for all the nodes t) colormap is meaningful.

CART partition (the largest plateau variant)



- CART results are not informative from the spatial viewpoint: it highlights the regions according to the specie distribution, not w.r.t. the interaction.
- CART cannot catch the mixed structure of species.

Perspectives

1. Extension to spatial Bagging or spatial Random Forests to cope with instability issue.
2. Use the sensitivity of CART with respect to rotation to generate several tessellations.
3. Extension to multi-marked point processes by combining one-versus-rest classifiers and then obtain several tessellations and select the partition maximizing some global measure of heterogeneity between cells.
4. Incorporate covariables:
   - in the example, elevation could be introduced as a third spatial coordinate.
   - more generally, we could imagine to first perform a classical CART using additional covariables and then, in each leaf, to perform a SpatCART and finally select the best one.

Thank you!