

Goal oriented random forest (GORF)

Kevin Elie-Dit-Cosaque, Bérénice-Alexia Jocteur, Véronique Maume-Deschamps, Clémentine Prieur, Pierre Ribereau, Ri Wang

▶ To cite this version:

Kevin Elie-Dit-Cosaque, Bérénice-Alexia Jocteur, Véronique Maume-Deschamps, Clémentine Prieur, Pierre Ribereau, et al.. Goal oriented random forest (GORF). Séminaire Phimeca 2023, Nov 2023, Paris, France. hal-04294065

HAL Id: hal-04294065 https://hal.science/hal-04294065

Submitted on 16 Dec 2023

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.



Distributed under a Creative Commons Attribution 4.0 International License

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References

Goal oriented random forest (GORF).

Véronique Maume-Deschamps joint work with Kevin Elie-Dit-Cosaque (SCOR SE), Bérénice-Alexia Jocteur (Natixis, ICJ), Clémentine Prieur (LJK - UGA), Pierre Ribereau (ICJ - UCBL), Ri Wang (ICJ).

17 novembre 2023





Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References O

Plan



- 2 Alternative loss functions
- On the consistency of RF
- 4 Simulation studies



Introduction ●○○○○○○○	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References ○

Plan

1 Introduction

- 2 Alternative loss functions
- 3 On the consistency of RF
- ④ Simulation studies

5 Conclusion

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
0000000					

Why considering GORF?

Random Forests are mainly designed for Regression or Classification purposes. In these cases, the target is observed. Other aims may be pursued, e.g.:

- conditional quantiles
- conditional average treatment effects (CATE)
- other conditional risk measure such as expectiles
- ...

 \implies Distributional Random Forest 1 vs Goal Oriented Random Forest.

¹ Domagoj Ćevid et al. (2022). In: Journal of Machine Learning Research Qiming Du et al. (2021). In: International Conference on Artificial Intelligence and Statistics

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References O

Recall CART

Classification And Regression Tree². Input variables:

 $\mathbf{X} = (X_1, \dots, X_d)$, Output variable: Y. Tree: constant piecewise predictor, obtained by binary recursive partitioning.

Separate the data from the current node, by looking for the split reducing the most the heterogeneity of Y at the two child nodes.



² Leo Breiman (2001). In: Machine learning



Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
0000000					

Loss function

In the regression context, $Y = m(\mathbf{X}) + \varepsilon$, the goal is to estimate $\mathbb{E}(Y|\mathbf{X}) = m(\mathbf{X})$. Consider a random sample $\mathcal{D}_n = (\mathbf{X}^i, Y^i), i = 1, ..., n$, the heterogeneity of Y is measured by the intra-groups variance, so that we shall maximise:

$$\mathcal{L}_{C}^{n}(j,z) = \frac{1}{\#C} \sum_{i=1}^{n} \left(\mathbf{Y}^{i} - \overline{\mathbf{Y}}_{C} \right)^{2} \mathbf{1}_{\{\mathbf{X}^{i} \in C\}} - \left[\frac{1}{\#C} \sum_{i=1}^{n} \left(\mathbf{Y}^{i} - \overline{\mathbf{Y}}_{C_{L}} \right)^{2} \mathbf{1}_{\{\in C_{L}\}} + \frac{1}{\#C} \sum_{i=1}^{n} \left(\mathbf{Y}_{i} - \overline{\mathbf{Y}}_{C_{R}} \right)^{2} \mathbf{1}_{\{\mathbf{X}_{i} \in C_{R}\}} \right],$$

where *C* is the current cell, ie an hyper-rectangle $\prod_{j=1}^{n} [a_j, b_j]$, #*C* is the number of elements of \mathcal{D}_n for which \mathbf{X}^i , i = 1, ..., n belongs to *C*; $C_L = C \cap \{x_j \le z\}$; $C_R = C \cap \{x_j > z\}$.

Introduction ○○○○●○○○	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References O
_					

Random Forests

Agregate several CART's to reduce the estimation variance \implies Bootstrap aggregating

- Training sample: $\mathcal{D}_n = (\mathbf{X}^i, Y^i), i = 1, \dots, n$
- Θ_ℓ, ℓ = 1,..., k are independent random variables, following Θ = (Θ¹, Θ²)' law: Θ¹ provides the bootstrap indices on D_n and Θ² gives which mtry variables are considered for the splits of each node. Θ_ℓ is assumed to be independent of D_n.
- A_n(x; Θ_ℓ, D_n): the leaf that is obtained when dropping x down the tree.
- $N_n(\mathbf{x}, \Theta_{\ell}, \mathcal{D}_n)$: the number of points which are in $A_n(\mathbf{x}; \Theta_{\ell}, \mathcal{D}_n)$.
- N^b_n(x, Θ_ℓ, D_n): the number of points of the bootstrapped sample, which are in A_n(x; Θ_ℓ, D_n).

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000					

CART estimation

In the regression framework, $\mathbb{E}(Y|\mathbf{X}) = m(\mathbf{X})$ is estimated on the random forest by. Let $B_j(\Theta_\ell, \mathcal{D}_n)$ be the number of times that the observation (\mathbf{X}^j, Y^j) has been drawn from the original dataset for the ℓ -th tree construction. Consider the weights:

$$\omega_{n,i}(\mathbf{x},\Theta) = \frac{1}{k} \sum_{j=1}^{k} \frac{\mathbf{1}_{\mathbf{x}^{i} \in A_{n}(\mathbf{x},\Theta_{j},\mathcal{D}_{n})}}{N_{n}(\mathbf{x},\Theta_{j},\mathcal{D}_{n})},$$

$$\omega_{n,i}^{b}(\mathbf{x},\Theta) = \frac{1}{k} \sum_{\ell=1}^{k} \frac{B_{i}(\Theta_{\ell},\mathcal{D}_{n}) \mathbf{1}_{\mathbf{x}^{i} \in A_{n}(\mathbf{x};\Theta_{\ell},\mathcal{D}_{n})}}{N_{n}^{b}(\mathbf{x};\Theta_{\ell},\mathcal{D}_{n})},$$

and the corresponding estimations of $m(\mathbf{x})$:

$$\hat{m}_{n}^{\mathbf{b}}(\mathbf{x}) = \sum_{i=1}^{n} \omega_{n,i}^{\mathbf{b}}(\mathbf{x}) Y^{i}.$$

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000					

Focus on causality and quantile estimation

Conditional quantile estimation : α ∈]0, 1[, the conditional quantile q_α(Y|X) could be estimated by inverting the estimated conditional distribution function: and the corresponding estimations of F(y|X = x) (introduced in Meinshausen 2006 and a.s. consistency proved in Elie-Dit-Cosaque and Maume-Deschamps 2022):

$$\hat{F}_{n}^{b}(y|\mathbf{X}=\mathbf{x}) = \sum_{i=1}^{n} \omega_{n,i}^{b}(\mathbf{x}) \mathbf{1}_{\{Y^{i} \leq y\}}.$$

• Conditional Average Treatement Effect

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000	00000	00000000	0000000	00	

Focus on causality and quantile estimation

- Conditional quantile estimation
- Conditional Average Treatement Effect:

 $CATE(\mathbf{X}) = \mathbb{E}(Y(1) - Y(0)|\mathbf{X})$ where Y(W) is a target variable (eg some biological quantity) in presence (Y(1)) or absence (Y(0)) of a treatement W. Some causal models write: $Y = \tau(\mathbf{X})W + \gamma(\mathbf{X})$, in this case, $CATE(\mathbf{X}) = \tau(\mathbf{X})$. Modifications on the CART construction has to be done in order to estimate CATE, since for each individual, we observe either $Y^{i}(1)$ or $Y^{i}(0)$.

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
0000000					

Other forests

Generalized Random Forests³ consider θ the target (eg conditional quantiles or CATE and use the following loss function:

$$\frac{\#C_L\#C_R}{\#C^2}\left[\hat{\theta}_{C_L}-\hat{\theta}_{C_R}\right]^2$$

where $\hat{\theta}_{C}$ is an estimator of the target θ on the cell C.

- Distributional Random Forests⁴ use Maximal Mean Discrepancy as loss function or the Wasserstein distance ⁵ for the construction of the split criterion.
- Trees designed for Extreme Value Analysis⁶.

³Susan Athey, Julie Tibshirani, Stefan Wager, et al. (2019). In: *The Annals of Statistics*

⁴ Domagoj Ćevid et al. (2022). In: Journal of Machine Learning Research

⁵ Qiming Du et al. (2021). In: International Conference on Artificial Intelligence and Statistics

⁶ Sébastien Farkas et al. (2021). In: arXiv preprint arXiv:2112.10409

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References O

Plan

1 Introduction

- 2 Alternative loss functions
- 3 On the consistency of RF
- ④ Simulation studies

5 Conclusion

Introduction Alternative	e loss functions On	the consistency of RF	Simulation studies	Conclusion	References
0000000 00000		0000000	0000000	00	

Goal oriented loss function

We propose loss functions specifically designed for

- CATE estimation,
- Conditional quantile estimation,

with a common proof scheme for a.s. consistency.

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References O
UTED	-				

Heterogeneous Treatment Effect Random Forest⁷ Consider the loss function:

$$\mathcal{L}_{C}(j,z) = \frac{\#C_{L}\#C_{R}}{\#C^{2}} \left(\left(\overline{Y}_{C_{L1}} - \overline{Y}_{C_{L0}} \right) - \left(\overline{Y}_{C_{R1}} - \overline{Y}_{C_{R0}} \right) \right)^{2}, \quad (1)$$
where $C_{L1} = \{ \mathbf{X}^{i} \in C_{L}, W^{i} = 1 \}, \quad C_{L0} = \{ \mathbf{X}^{i} \in C_{L}, W^{i} = 0 \},$
 $C_{R1} = \{ \mathbf{X}^{i} \in C_{R}, W^{i} = 1 \}, \quad C_{R0} = \{ \mathbf{X}^{i} \in C_{R}, W^{i} = 0 \},$

⁷ Bérénice-Alexia Jocteur, Véronique Maume-Deschamps, and Pierre Ribereau (2023). In: https://hal.science/hal-04112079

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000	00000	00000000	0000000	00	

HTERF estimation

For the estimation of CATE we use:

$$\widehat{CATE}_{(\mathbf{x})} = \sum_{i:W^{i}=1} \omega_{n,i} (\mathbf{x}, \Theta) Y^{i} - \sum_{i:W^{i}=0} \omega'_{n,i} (\mathbf{x}, \Theta) Y^{i}, \quad (2)$$

where ω (resp. ω') are the weights associated to observations such as $W^i = 1$ (resp. $W^i = 0$).

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
	00000				

Pin-ball loss

Consider the pin-ball function⁸: $\psi_{\alpha}(y, \theta) = (y - \theta)(\alpha - \mathbf{1}_{\{y \le \theta\}})$ Recall that the α -quantile is given by: $q^{\alpha}(Y) = \arg \min_{\theta} \mathbb{E}[\psi_{\alpha}(Y, \theta)].$ Consider the loss function⁹: $\mathcal{L}_{C}(j, z) = \sum_{i=1}^{n} \psi_{\tau}(Y^{i}, \hat{\theta}_{C}) \mathbf{1}_{\mathbf{X}^{i} \in C} - \left[\sum_{i=1}^{n} \psi_{\tau}(Y^{i}, \hat{\theta}_{C_{L}}) \mathbf{1}_{\mathbf{X}^{i} \in C_{L}} + \sum_{i=1}^{n} \psi_{\tau}(Y^{i}, \hat{\theta}_{C_{R}}) \mathbf{1}_{\mathbf{X}^{i} \in C_{R}}\right],$

where $\hat{\theta}_C$ is an estimator of the α -quantile in C. Estimate the conditional distribution function then the quantile as before.

⁸It is also named check function, quantile loss.

⁹ Harish S Bhat, Nitesh Kumar, and Garnet J Vaz (2015). In: 2015 IEEE International Conference on Big Data (Big Data). IEEE

Introduction	Alternative loss functions	On the consistency of RF •00000000	Simulation studies	Conclusion	References ○
Plan					

1 Introduction

- 2 Alternative loss functions
- On the consistency of RF
- ④ Simulation studies

5 Conclusion

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
		0000000			

Consistency of random forests

Results by Scornet, Biau, Vert (2015) in a linear model context:

$$Y = m(X) + \varepsilon$$
 with $\varepsilon \rightsquigarrow \mathcal{N}(0, \sigma^2)$ and $m(X) = \sum_{j=1}^d m_j(X_j)$.

Under various assumptions including tree size wrt n and a forest correlation control, for $\mathbf{X} \rightsquigarrow \mathcal{U}[0,1]^d$,

$$\mathbb{E}[(m_n(\mathbf{X}) - m(\mathbf{X}))^2] \longrightarrow 0$$
, with $m_n = \mathbb{E}_{\Theta}(\hat{m}_n)$.

- The bootstrap is not taken into account in Θ
- No results for $m(\mathbf{x})$
- Results for fully grown trees and for limited grown trees.

Introduction	Alternative loss functions	On the consistency of RF 00●000000	Simulation studies	Conclusion	References ○
Consist	ency of GRF				

Asymptotic normal laws obtained¹⁰ under

- regularity asumptions of the target function,
- constraints on the tree construction

¹⁰Susan Athey, Julie Tibshirani, Stefan Wager, et al. (2019). In: *The Annals of Statistics*

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References O
Consist	ency of GRF				

Asymptotic normal laws obtained¹⁰ under

• regularity asumptions of the target function, which write in the *CATE* estimation case:

 $\mathbf{x} \mapsto \mathbb{E}[Y(u)|\mathbf{X} = \mathbf{x}]$ and $\mathbf{x} \mapsto \mathbb{E}[Y(u)^2|\mathbf{X} = \mathbf{x}]$ are Lipschitz-continuous, $Var[Y(u)|\mathbf{X} = \mathbf{x}] > 0$ and $\mathbb{E}[|Y(u) - \mathbb{E}[Y(u)|\mathbf{X} = \mathbf{x}]|^{2+\delta}|\mathbf{X} = \mathbf{x}] \leq M$ for some constants $\delta, M > 0$ uniformly over all $\mathbf{x} \in [0, 1]^d$.

• constraints on the tree construction

¹⁰ Susan Athey, Julie Tibshirani, Stefan Wager, et al. (2019). In: *The Annals of Statistics*

Introduction	Alternative loss functions	On the consistency of RF 00●000000	Simulation studies	Conclusion	References O
Consist	ency of GRF				

Asymptotic normal laws obtained¹⁰ under

- regularity asumptions of the target function,
- constraints on the tree construction:
 - at every step of the tree building procedure, the probability that the next split is done along the j th feature is bounded below by π/d for some $0 < \pi \le 1$ for all $j = 1, \ldots, d$ random split hypothesis.
 - for some fixed $\gamma,$ each split leaves at least a fraction γ of the available training sample on each side of the split, $\gamma\text{-regularity hypothesis}$
 - for some fixed p, the leaf containing x has at least p observations for each treatment group and the leaf containing x has either less than 2p-1 observations with Wⁱ = 0 or 2p-1 observations with Wⁱ = 1.

¹⁰ Susan Athey, Julie Tibshirani, Stefan Wager, et al. (2019). In: *The Annals of Statistics*

Introduction	Alternative	loss functions	On the consistency of RF	Simulation studies	Conclusion	References
a						

Consistency of GRF

Asymptotic normal laws obtained¹⁰ under

- regularity asumptions of the target function,
- constraints on the tree construction

The bootstrap is not taken into account, the proof is done for honest forests (ie independance of the sample used for the tree construction and the estimation).

¹⁰ Susan Athey, Julie Tibshirani, Stefan Wager, et al. (2019). In: *The Annals of Statistics*

Introduction Alte	ernative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
)00 C	0000000			

General scheme for the a.s. consistency

Conditions

Relations between k (number of trees) and $N_n^b(\mathbf{x}; \Theta, \mathcal{D}_n)$ (number of bootstrap observations in a leaf node):

$$\bullet \ \mathbf{k} = \mathcal{O}\left(\mathbf{n}^{\alpha}\right), \text{ with } \alpha > 0.$$

an asumption on mean and variance of $N_n^b(\mathbf{x}; \Theta, \mathcal{D}_n)$ The variation of the target function θ is small on the trees' leaves: for any \mathbf{x} ,

$$\sup_{\mathbf{z}'\in A_n(\mathbf{x},\Theta_j)} |\theta(\mathbf{z}) - \theta(\mathbf{z}')| \xrightarrow[n\to\infty]{a.s.} 0.$$

 $^{s}f\left(n
ight)=\Omega\left(g\left(n
ight)
ight)\iff \exists k>0, \exists n_{0}>0\midorall n\geqslant n_{0}\quad\left|f\left(n
ight)
ight|\geqslant k\cdot\left|g\left(n
ight)
ight|$

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
		00000000			

Consistency: result

Theorem

Assume the conditions above are verified then

$$|\hat{ heta}_n(\mathbf{x}) - heta(\mathbf{x})| \stackrel{a.s.}{\underset{n \to \infty}{\longrightarrow}} 0$$

Proof for conditional distributions¹¹ and $CATE^{12}$ but the proof scheme applies more generally.

 ¹¹ Kevin Elie-Dit-Cosaque and Véronique Maume-Deschamps (2022). In: Electronic Journal of Statistics
 ¹² Bérénice-Alexia Jocteur, Véronique Maume-Deschamps, and Pierre Ribereau (2023). In: https://hal.science/hal-04112079

24 / 41

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
		000000000			

Remark on the variation hypothesis

In order to get the variation hypothesis we need the continuity of $\boldsymbol{\theta}$ and either

- the random split and γ -regularity hypothesis or,
- the convergence of the empirical loss function to the theoretical one (with some uniformity on the cells C) and the fact that the theoretical loss function is 0 on a cell C of a theoretical tree¹³ implies that θ is zero on C. This last condition is true e.g. in the CATE estimation setting for a large class of functions (including sums, products, dense classes).

¹³introduced in Scornet, Biau, and Vert 2015 and deeply used in Elie-Dit-Cosaque and Maume-Deschamps 2022

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References ○
Theroe	tical trees				

A theoretical tree is grown following the same rules as an empirical tree, except that the theoretical equivalent of the empirical split criterion on a node C is used to choose the best split. E.g. for the CART-tree, the theoretical split criterion is:

$$\begin{split} \mathcal{L}_{C}^{\star}(j,z) = & \operatorname{Var}\left(\left.Y\right| \mathbf{X} \in C\right) \\ & - \mathbb{P}\left(\left.\mathbf{X} \in C_{L}\right| \mathbf{X} \in C\right) \operatorname{Var}\left(\left.Y\right| \mathbf{X} \in C_{L}\right) \\ & - \mathbb{P}\left(\left.\mathbf{X} \in C_{R}\right| \mathbf{X} \in C\right) \operatorname{Var}\left(\left.Y\right| \mathbf{X} \in C_{R}\right) \ . \end{split}$$

Hence, a theoretical tree is obtained thanks to the best consecutive cuts (j^*, z^*) , among $j \in \mathcal{M}_{try}$, $z \in C^j$ optimizing the previous criterion $\mathcal{L}^*_C(\cdot, \cdot)$.

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
		000000000			

Empirical vs theoretical trees

Consider the model $Y = m(\mathbf{X}) + \varepsilon$ with $m(\cdot)$ in the \mathbf{A} -class¹⁴ and ε with ligth tails.

Proposition

For any $h \in \mathbb{N}$ fixed, for any empirical tree with node sizes greater than $C\sqrt{n}(\ln n)^{\beta}$, $\beta > \frac{5}{2}$, consider a node at height h in the theoretical tree (resp. empirical tree) and \mathcal{T}_h the set of theoretical trees of height h, then

let
$$A = \prod_{j=1}^{d} [a_j, b_j]$$
 and $A^n = \prod_{j=1}^{d} [a_j^n, b_j^n]$. We have:

$$\inf_{\mathcal{T}_h} \max_{j=1,...,d} \max\left(|a_j - a_j^n|, |b_j - b_j^n|\right) \longrightarrow 0 \text{ a.s. as } n \to \infty.$$

 14 Functions in this class satisfy that if $\mathcal{L}_{\mathcal{C}}^{\star}(j,z)=0$ for all j,z the m is constant on $\mathcal{C}.$

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000	00000	000000000	0000000		

Empirical vs theoretical trees

Proposition

For any $h \in \mathbb{N}$ fixed, for any empirical tree with node sizes greater than $C\sqrt{n}(\ln n)^{\beta}$, $\beta > \frac{5}{2}$, consider a node at height h in the theoretical tree (resp. empirical tree) and \mathcal{T}_h the set of theoretical trees of height h, then

let
$$A = \prod_{j=1}^{d} [a_j, b_j]$$
 and $A^n = \prod_{j=1}^{d} [a_j^n, b_j^n]$. We have:

$$\inf_{\mathcal{T}_h} \max_{j=1,...,d} \max\left(|a_j - a_j^n|, |b_j - b_j^n|\right) \longrightarrow 0 \text{ a.s. as } n \to \infty.$$

Proved for CART trees¹⁴ and for causal forests¹⁵.

¹⁴ Kevin Elie-Dit-Cosaque and Véronique Maume-Deschamps (2022). In: *Electronic Journal of Statistics*

¹⁵ Bérénice-Alexia Jocteur, Véronique Maume-Deschamps, and Pierre Ribereau (2023). In: *https://hal.science/hal-04112079*

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References ○

The two samples method.

One of the main idea to prove the constistency of random forests is to use an auxiliary sample: let $(\mathbf{X}^{i\diamond}, Y^{i\diamond}, i = 1, ..., n)$ be a second sample, independent from $(\mathbf{X}^i, Y^i, i = 1, ..., n)$ and consider the weights

$$\omega_{n,i}^{\diamond}(\mathbf{x},\Theta) = \frac{1}{k} \sum_{j=1}^{k} \frac{\mathbf{1}_{\mathbf{x}^{i\diamond} \in A_n(\mathbf{x},\Theta_j,\mathcal{D}_n)}}{N_n^{\diamond}(\mathbf{x},\Theta_j,\mathcal{D}_n)}$$

and the corresponding estimator θ_n^\diamond of θ . We prove:

| θ̂_n (x) - θ[◊]_n (x) | ^{a.s.}/_{n→∞} 0, uses a Hoeffding like inequality + Vapnik-Chervonenkis classes¹⁶ (proximity of N[◊] and N^b),
 |θ[◊]_n (x) - θ (x)| ^{a.s.}/_{n→∞} 0, uses Vapnik-Chervonenkis classes again and the variation hypothesis.

¹⁶ V. N. Vapnik and A. Ya. Chervonenkis (1971). In: *Theory of Probability* and its Applications

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References

Plan

1 Introduction

- 2 Alternative loss functions
- 3 On the consistency of RF
- ④ Simulation studies

5 Conclusion

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
			000000		

CATE estimation: a first example

We consider simulated data close to causal frameworks previously studied ¹⁷. $\mathbf{X} \sim U([0,1]^p)$, $W \sim Bern(0.5)$ and $Y = \tau(\mathbf{X})W + \beta\gamma(\mathbf{X})$, p = 10, $\tau(\mathbf{x}) = \sin(x_1)$ and $\gamma(\mathbf{x}) = \cos(2x_2 + 3x_3)$. The scalar β allows to consider the impact of the magnitude of τ relative to γ .

β	GRF	HTERF
5	0.276	0.117
1	0.122	0.012
0.2	0.079	0.004

Table: Mean squared errors of GRF and HTERF methods that estimate heterogeneous treatment effect, with 500 tree forests.

¹⁷ Susan Athey, Julie Tibshirani, Stefan Wager, et al. (2019). In: *The Annals of Statistics*

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000	00000	00000000	000000		

CATE estimation: interpretability

		GRF				HT	ERF	
β	dep.3	dep.5	dep.10	imp.	dep.3	dep.5	dep.10	imp.
5	0.870	0.378	0.150	0.852	1	0.498	0.175	0.985
1	0.874	0.526	0.174	0.866	1	0.995	0.282	1
0.2	0.875	0.627	0.2	0.866	1	1	0.603	1

Table: Frequencies of splitting on X_1 at depths 3, 5 and 10 and importance of X_1 .

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
			0000000		

CATE estimation: a non linear framework

Let $\mathbf{X} \sim U([0,1]^p)$, $W \sim Bern(0.5)$ and $Y = \sin(X_1)(W+2)^3 + \cos(X_2)$, where p = 3. Hence we have CATE that satisfies: $\tau(\mathbf{x}) = 19\sin(x_1)$.

Method	RMSE	importance
GRF	0.321	0.777
HTERF	0.209	1

Table: Root mean squared errors of GRF and HTERF methods that estimate heterogeneous treatment effect, with 500 tree forests. We also consider the importance of X_1 .

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000	00000	00000000	0000000	00	

Conditional quantiles estimation: two examples¹⁸

Example 1: $Y | \mathbf{X} \sim \mathcal{N}(0, 1 + \mathbf{1}(X_1 > -0.5)), p = 10, X_j, j = 1, 2, ..., 10$ are independent draws from uniform distribution $\mathcal{U}(-1, 1)$, $\mathbf{x}_i = (x_i, 0, 0, ..., 0)$, and x_i is taken from regular grid over [-1, 1].

Example 2: $Y = X_1 - X_2 + \varepsilon, X_j, j = 1, 2, ..., 10$ are independent draws from $X_j \sim Exp(1), \varepsilon \sim \mathcal{N}(0, 1),$ $\mathbf{x}_i = (x_{i1}, x_{i2}, 1, 1, ..., 1),$ and (x_{i1}, x_{i2}) are taken from regular grid over $[0, 5]^2$.

¹⁸ Véronique Maume-Deschamps, Clémentine Prieur, and Ri Wang.

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
			0000000		

Conditional quantiles estimation: Example 1

α	n	GRF	Quantile06	DRF	Pin-ball
0.1	500	0.197	0.219	0.107	0.085
	2000	0.038	0.140	0.027	0.028
	4000	0.025	0.110	0.019	0.016
0.5	500	0.015	0.016	0.017	0.035
	2000	0.007	0.007	0.009	0.010
	4000	0.006	0.006	0.010	0.010
0.9	500	0.205	0.221	0.112	0.083
	2000	0.037	0.137	0.025	0.022
	4000	0.026	0.119	0.017	0.014

Table: MSE¹⁹ for Example 1: $Y | \mathbf{X} \sim \mathcal{N}(0, 1 + \mathbf{1}(X_1 > -0.5))$

¹⁹100 repetitions, number of trees B = 200, min.node.size = 20, mtry = p. Other parameters use the default setting.

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
00000000			000000		

Conditional quantiles estimation: Example 2

au	п	GRF	Quantile06	DRF	Pin-ball
0.1	500	5.700	2.679	3.074	1.629
	2000	4.754	1.770	1.949	0.763
	4000	3.959	1.133	1.373	0.474
0.5	500	1.848	1.166	1.713	0.817
	2000	0.761	0.405	0.644	0.383
	4000	0.411	0.243	0.341	0.235
0.9	500	5.869	2.833	3.022	1.812
	2000	4.842	1.717	1.809	0.847
	4000	4.093	1.201	1.332	0.497

Table: MSE²⁰ for Example 2: $Y = X_1 - X_2 + \varepsilon$

²⁰100 repetitions, number of trees B = 200, min.node.size = 20, mtry = p. Other parameters use the default setting.

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References
				•0	

Plan

1 Introduction

- 2 Alternative loss functions
- 3 On the consistency of RF
- ④ Simulation studies



Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion ○●	References O
Conclu	sion				

- Use Goal Oriented Random Forest for specific purposes:
 - better estimation (in general)
 - better interpretability.
- Results developped for *CATE* and conditional quantile estimations but other quantities are reachable.
- General scheme for a.s. consistency results.
- Theoretical random forest = a powerfull tool to better understand the random forests methods.

References I

ā

- Athey, Susan, Julie Tibshirani, Stefan Wager, et al. (2019). "Generalized random forests". In: The Annals of Statistics 47.2, pp. 1148–1178.
- Bhat, Harish S, Nitesh Kumar, and Garnet J Vaz (2015). "Towards scalable quantile regression trees". In: 2015 IEEE International Conference on Big Data (Big Data). IEEE, pp. 53–60.
- Breiman, Leo (2001). "Random forests". In: Machine learning 45.1, pp. 5-32.
- Čevid, Domagoj et al. (2022). "Distributional random forests: Heterogeneity adjustment and multivariate distributional regression". In: Journal of Machine Learning Research 23.333, pp. 1–79.
- Du, Qiming et al. (2021). "Wasserstein random forests and applications in heterogeneous treatment effects". In: International Conference on Artificial Intelligence and Statistics, pp. 1729–1737.
- Elie-Dit-Cosaque, Kevin and Véronique Maume-Deschamps (2022). "Random forest estimation of conditional distribution functions and conditional quantiles". In: *Electronic Journal of Statistics* 16, 6553–6583.
- Farkas, Sébastien et al. (2021). "Generalized Pareto Regression Trees for extreme events analysis". In: arXiv preprint arXiv:2112.10409.
- Jocteur, Bérénice-Alexia, Véronique Maume-Deschamps, and Pierre Ribereau (2023). "Heterogeneous Treatment Effect based Random Forest: HTERF". In: https://hal.science/hal-04112079.
- Maume-Deschamps, Véronique, Clémentine Prieur, and Ri Wang. "Work in progress".
- Meinshausen, Nicolai (2006). "Quantile regression forests". In: Journal of Machine Learning Research 7.Jun, pp. 983–999.



Scornet, Erwan, Gérard Biau, and Jean-Philippe Vert (2015). "Consistency of random forests". In: The Annals of Statistics.

Vapnik, V. N. and A. Ya. Chervonenkis (1971). "On the Uniform Convergence of Relative Frequencies of Events to their Probabilities". In: Theory of Probability and its Applications 16.2, pp. 264–280.

Introduction	Alternative loss functions	On the consistency of RF	Simulation studies	Conclusion	References ●
Merci					

Thanks for your attention. Merci pour votre attention.