

**Scornet, Erwan**

**Tuning parameters in random forests.** (English) Zbl 1427.68273  
ESAIM, Proc. Surv. 60, 144-162 (2017).

Summary: *L. Breiman's* [Mach. Learn. 45, No. 1, 5–32 (2001; [Zbl 1007.68152](#))] random forests are a very popular class of learning algorithms often able to produce good predictions even in high-dimensional frameworks, with no need to accurately tune its inner parameters. Unfortunately, there are no theoretical findings to support the default values used for these parameters in Breiman's algorithm. The aim of this paper is therefore to present recent theoretical results providing some insights on the role and the tuning of these parameters.

**MSC:**

**68T05** Learning and adaptive systems in artificial intelligence  
**62H30** Classification and discrimination; cluster analysis (statistical aspects)

Cited in 1 Document

**Software:**

[hgam](#); [SuperLearner](#)

**Full Text:** [DOI](#)

**References:**

- [1] S. Arlot and R. Genuer. Analysis of purely random forests bias. arXiv:1407.3939, 2014. · [Zbl 1402.62131](#)
- [2] S. Bernard, L. Heutte, and S. Adam. Forest-RK: A new random forest induction method. In D.-S. Huang, D.C. Wunsch II, D.S. Levine, and K.-H. Jo, editors, *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, pages 430-437, Berlin, 2008. Springer.
- [3] G. Biau. Analysis of a random forests model. *Journal of Machine Learning Research*, 13:1063-1095, 2012. · [Zbl 1283.62127](#)
- [4] G. Biau and L. Devroye. Cellular tree classifiers. *Electronic Journal of Statistics*, 7:1875-1912, 2013. · [Zbl 1293.62067](#)
- [5] G. Biau, L. Devroye, and G. Lugosi. Consistency of random forests and other averaging classifiers. *Journal of Machine Learning Research*, 9:2015-2033, 2008. · [Zbl 1225.62081](#)
- [6] G. Biau and E. Scornet. A random forest guided tour. *Test*, 25:197-227, 2016. · [Zbl 1402.62133](#)
- [7] A.-L. Boulesteix, S. Janitza, J. Kruppa, and I. R. König. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2:493-507, 2012.
- [8] L. Breiman. Random forests. *Machine Learning*, 45:5-32, 2001. · [Zbl 1007.68152](#)
- [9] L. Breiman. *Consistency for a simple model of random forests*. Technical Report 670, University of California, Berkeley, 2004.
- [10] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees*. Chapman & Hall/CRC, Boca Raton, 1984. · [Zbl 0541.62042](#)
- [11] A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision*, 7:81-227, 2011. · [Zbl 1243.68235](#)
- [12] D.R. Cutler, T.C. Edwards Jr, K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler. Random forests for classification in ecology. *Ecology*, 88:2783-2792, 2007.
- [13] R. Díaz-Uriarte and S. Alvarez de Andrés. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7:1-13, 2006.
- [14] R. Duroux and E. Scornet. Impact of subsampling and pruning on random forests. arXiv:1603.04261, 2016. · [Zbl 1409.62072](#)
- [15] R. Genuer. Variance reduction in purely random forests. *Journal of Nonparametric Statistics*, 24:543-562, 2012. · [Zbl 1254.62050](#)
- [16] R. Genuer, J. Poggi, and C. Tuleau-Malot. Variable selection using random forests. *Pattern Recognition Letters*, 31:2225-2236, 2010.
- [17] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York, 2002.

- [18] T. Hastie and R. Tibshirani. Generalized additive models. *\textit{Statistical Science}*, 1:297-310, 1986. · [Zbl 0645.62068](#)
- [19] H. Ishwaran. The effect of splitting on random forests. *\textit{Machine Learning}*, pages 1-44, 2013.
- [20] L. Meier, S. Van de Geer, and P. Bühlmann. High-dimensional additive modeling. *\textit{The Annals of Statistics}*, 37:3779-3821, 2009. · [Zbl 1360.62186](#)
- [21] N. Meinshausen. Quantile regression forests. *\textit{Journal of Machine Learning Research}*, 7:983-999, 2006. ESAIM: PROCEEDINGS AND SURVEYS159 · [Zbl 1222.68262](#)
- [22] L. Mentch and G. Hooker. Ensemble trees and clts: Statistical inference for supervised learning. arXiv:1404.6473, 2014.
- [23] A.M. Prasad, L.R. Iverson, and A. Liaw. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *\textit{Ecosystems}*, 9:181-199, 2006.
- [24] E. Scornet. On the asymptotics of random forests. *\textit{Journal of Multivariate Analysis}*, 146:72-83, 2016. · [Zbl 1337.62063](#)
- [25] E. Scornet, G. Biau, and J.-P. Vert. Consistency of random forests. *\textit{The Annals of Statistics}*, 43:1716-1741, 2015. · [Zbl 1317.62028](#)
- [26] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *\textit{IEEE Conference on Computer Vision}\textit{and Pattern Recognition}*, pages 1297-1304, 2011.
- [27] C.J. Stone. Additive regression and other nonparametric models. *\textit{The Annals of Statistics}*, pages 689-705, 1985. · [Zbl 0605.62065](#)
- [28] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, and B.P. Feuston. Random forest: A classification and regression tool for compound classification and QSAR modeling. *\textit{Journal of Chemical}\textit{Information and Computer Sciences}*, 43:1947-1958, 2003.
- [29] M. van der Laan, E.C. Polley, and A.E. Hubbard. Super learner. *\textit{Statistical Applications in Genetics and}\textit{Molecular Biology}*, 6, 2007. · [Zbl 1166.62387](#)
- [30] S. Wager. Asymptotic theory for random forests. arXiv:1405.0352, 2014.
- [31] S. Wager, T. Hastie, and B. Efron. Standard errors for bagged predictors and random forests. arXiv:1311.4555, 2013.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.