How can we explain Random Forests in a spatial framework?

How can we explain Random Forests in a spatial framework?

Natalia Golini[a], Luca Patelli[b], and Xavier Barber[c]

[a]University of Torino, Department of Economics and Statistics Cognetti de Martiis, Lungo Dora Siena, 100A, Torino; natalia.golini@unito.it
[b]University of Pavia, Department of Economics and Management, Via San Felice al Monastero, 5, Pavia; luca.patelli01@universitadipavia.it
[c]Universidad Miguel Hernández de Elche, Centro de Investigación Operativa, Avenida de la Universidad, Elche; xbarber@umh.es

Abstract

Random Forest (RF) is a Machine Learning algorithm, very popular in environmental applications thanks to its flexibility and predictive performances. Even if its working mechanism is simple and intelligible, RF is considered a *black box* model since it prevents grasping how predictors are combined to generate the response variable predictions. This lack of interpretability represents a limitation of RF, especially when some knowledge is required on the response-predictors relationship from the decision-making perspective. In this work, we aim to explain RF using a Post-Hoc approach, i.e. by extracting a compact and simple list of rules from an estimated RF focusing on a spatial regression context. By means of a spatial dataset, we compare the final sets of rules and discuss the predictive accuracies of the standard RF and its *gold standard* for the case of spatially correlated data.

Keywords: Explainable Machine Learning, inTrees, Post-hoc methods, Rule extraction, RF-GLS

1. Introduction

The Machine Learning (ML) era has given rise to complex and powerful methods that can process vast amounts of data and make predictions with remarkable accuracy. However, the inherent *black-box* nature of some of these techniques has raised concerns about their lack of interpretability. Often the term *interpretability* is used as a synonym for *explainability*, but actually they refer to two different concepts. According to Rudin *et al.* (11), interpretability is referred to models that are built to be interpretable, while explainability is obtained by applying further techniques to non-interpretable models in order to extract information. On the topic of explainable ML methods, the recent paper by Wikle *et al.* (14) is worth to be mentioned. In particular, the authors discuss the use of explainability techniques in spatial ML to understand the role of specific inputs in predicting environmental variables. Even if from a statistical point of view the gold standard would be to use interpretable ML methods, when this is not possible it is a good practice to try to extract information from non-interpretable ML methods that have proven good performance.

In this work, among ML techniques, we consider Random Forest which is well known for its high prediction accuracy. It is a non-parametric supervised algorithm that, thanks to its flexibility, can model complex non-linear relationships between the response variable (categorical or continuous) and the predictors (3). RF is defined as an ensemble model as the result of aggregating a set of decision trees. Each tree is the result of a recursive binary splitting process obtained using re-sampled data and a random set of predictors evaluated at each node as splitting candidates. Given its adaptability, RF has also been widely applied in the spatial framework with different strategies to deal with the spatial autocorrelation of the data. Patelli *et al.* (10) have recently proposed a literature review and a novel taxonomy of the existing strategies adopted to adjust RF for spatially correlated data. In particular, the authors highlight that the most interesting strategy is the RF-GLS method proposed by Saha *et al.* (12), who extend the RF by estimating trees using generalized least squares (GLS). It was proven that RF-GLS outperforms the classical RF in the presence of spatial correlation, thus representing the gold standard to be used in the spatial framework.

In any case, spatially aware or not, RF remains a non-interpretable algorithm. However, it is possible to use specific methods to explain the RF resulting model, as described in the review by Haddouchi and Berrado (7). In particular, "Internal Processing" (IP) methods try to get "insights that are inherent to internal processing" providing a global overview of the model. "Post-Hoc" (PH) methods instead are based on RF post-processing, such as for example the "Rule Extraction" (RE) approaches (see e.g. inTrees (5), SIRUS (2), Node harvest (9) and RuleFit (6) among others). These methods aim to find a limited set of rules (each defined as the combination of predictors and split values) that is common to many trees in the RF and that allow representing the prediction mechanism of RF.

The main aim of this contribution is to verify if, for a spatial regression problem, there exist differences in the rules obtained by using - so far - the inTrees approach applied to two different cases: trees grown by RF-GLS and by a classical RF. We expect that taking or not into account the spatial correlation when implementing RF will have an impact also in its extracted rules. The analysis is carried out by using a dataset regarding daily meteorological records measured by 159 monitoring stations in Croatia. We present here preliminary results followed by a discussion on further steps.

2. Data and methods

The explainability of RF in the spatial framework is illustrated using meteorological daily data from the national network of 159 stations in Croatia for the year 2008, provided by the Croatian National Meteorological Service (available at https://github.com/AleksandarSekulic/RFSI). At this stage of the work, we have not considered the temporal dimension of the data confining the analysis to a single day: 14^{th} June 2008. The locations of the 151 stations working at this date are shown in Fig. 1. In particular, dots and crosses represent training and test data considered to implement the RF-GLS and RF algorithms. For this dataset, we randomly selected 90% of the data (i.e., 135 observations) for training the algorithms and used the remaining 10% of the data (i.e., 16 observations) for testing the algorithms. Croatia is a country located in southeastern Europe, bordering the Adriatic Sea. It has a diverse topography with flat plains in the east, a hilly central region, and mountainous terrain in the west. The response variable is the mean daily temperature[1] [`TEMP`], measured in degrees Celsius (°C). The minimum and maximum observed mean daily temperature values are 1.8°C and 21.5°C, respectively. The highest temperatures are recorded along the coast and at low altitudes. The variables used as predictors are latitude [`lat` (in meters)], longitude [`lon` (in meters)], distance-to-coastline [`HRdsea` (in km)], elevation [`HRDdem` (in meters)], wetness index [`HRtwi`], seasonal fluctuation [`ctd` (in days)], insolation (total incoming solar radiation) [`INSOL` (in Joules)], and Moderate Resolution Imaging Spectroradiometer land surface temperature [`MODIS.LST`] images. The dataset and predictors are detailed in (8) and references therein. In particular, this dataset was used by Sekulić *et al.* (13) to evaluate and compare the performance of a spatial interpolation method they proposed, i.e. the Random Forest Spatial Interpolation.

With the aim of obtaining simple, stable and accurate rules, we implemented the inTrees approach proposed by Deng (5) and implemented in the homonym `R` package `inTrees`[2]. The set of algorithms proposed in the work of Deng (5) can be applied to all tree ensemble methods to perform different tasks: extract, prune, select and summarize the rules. Each step is not mandatory, and the procedure can be tailored based on the specific explanatory necessity.

[1] On most meteorological stations `TEMP` is measured three times a day: at 7 am, 1 pm and 9 pm.

[2] https://cran.r-project.org/web/packages/inTrees/index.html

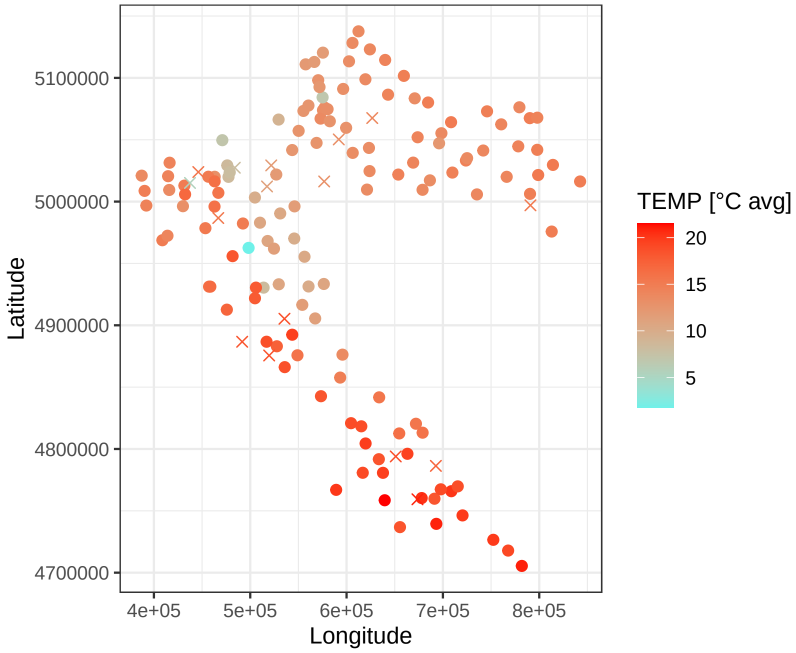

Figure 1: Mean daily temperature recorded on 2008-06-14 in 151 Croatian meteorological stations. Dots represent the mean daily temperature registered in the 135 training sites; crosses represent the mean daily temperature measured in the 16 test sites.

In order to extract and analyze rules by means of `inTrees`, the first step consists in running the chosen RF algorithm to have a collection of trees grown over a set of training data. Each tree results in the combination of all its splits, i.e. the conditions that permit splitting of the predictor space and getting predictions in the final regions. Then the obtained rules can be evaluated by using the relative "frequency" of occurrence, the prediction "error" and their "length" representing the rule complexity.

Using these metrics and considering opportune (decay) functions, the rules can be further simplified by pruning irrelevant predictor-split values. In order to have a compact rule set containing relevant and non-redundant rules, a complexity-guided condition selection method can be used, e.g. guided regularized Random Forest (GRRF) (4). In the end, the extracted rules can also be summarized by a rule-based learner that should be comparable in terms of prediction accuracy to the standard RF but more interpretable, named Simplified Tree Ensemble Learner (STEL). Note that in `inTrees` it is possible to build a STEL only for classification problems.

3. Preliminary empirical results

This section shows our preliminary results by applying the inTrees approach to extract insights from the RF-GLS and RF algorithms applied to the temperature spatial dataset.

We started by training the regression RF-GLS and RF on the same training set, by means of the R packages `randomForestGLS`[3] and `randomForest`[4], respectively. We used the same setting for the hyperparameters. In particular, we have set to 1000 the number of trees (`ntree` in `R`) and to 3 (one-third of the total number of predictors) the number of the variables randomly sampled as candidates at each split (`mtry` in `R`). For the RF-GLS, the covariance function used in modelling the spatial dependence structure among the observations was the default value, i.e. the exponential covariance function (`cov.mat` in `R`). Note that the coordinates [`lat`, `long`], measured in meters, have also been considered as predictors in both algorithms. In order to stabilize the forest structure, we followed the strategy pro-

[3] https://cran.r-project.org/web/packages/RandomForestsGLS/index.html
[4] https://cran.r-project.org/web/packages/randomForest/index.html

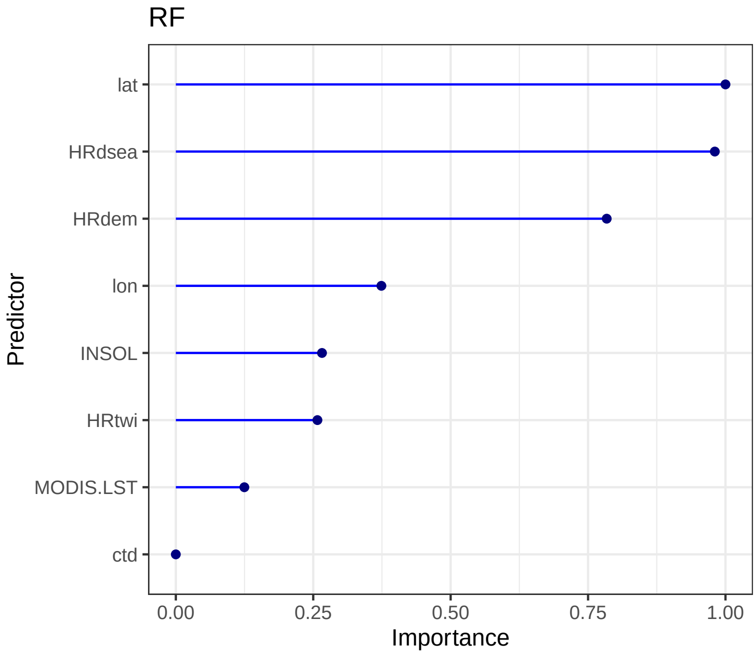

Figure 2: Variable importance plot for RF. The importance index is scaled to a maximum of 1.

posed in Bénard *et al.* (2) for rule generation consisting in restricting the node splits to the q-empirical quantiles of the predictors. This modification to Breiman's original regression tree algorithm is expected to have a small impact on predictive accuracy but is essential for stability.

Table 1 shows the test accuracy in terms of root mean square error and percentage of explained variance of the two algorithms when the node splits are restricted to the 10-empirical quantiles of the predictors. Different values of q will be considered in the next steps of the work. As expected, RF-GLS shows a better predictive performance than RF because it is able to capture the spatial autocorrelation of the response variable.

Table 1: Root Mean Square Error (RMSE) and percentage of explained variance (Var explained) values evaluated for the test dataset.

Algorithm	RMSE [°C]	Var explained [%]
RF-GLS	1.057	93.52
RF	1.357	89.32

Latitude, distance-to-coastline and DEM are the most important predictors for RF (see Fig. 2). This information is not reported for RF-GLS since the `R` package `randomForestGLS` does not provide the variable importance as output yet.

Given the two forests, we applied the inTrees approach described in Section 2. For both algorithms, RF-GLS and RF, we used the same setting for the tuning parameters of the `inTrees` functions. We extracted the rule conditions from the set of trees with a maximum length of 3 (`maxdepth` in `R`) from each tree. The distinct rule conditions extracted from the 1000 trees of RF-GLS and RF were 2,836 and 3,007, respectively. Then, we assigned the outcome values (mean of the response variable values of the training observations that satisfy the condition) [pred] to the conditions and measured the quality of the rules by "frequency" [freq], "error" [err], and "length" [len]. We pruned the extracted rules' irrelevant or redundant variable-value pairs considering the metric "error" and the "relative" decay function. With the irrelevant variable-value pairs being removed, the pruned rules have shorter conditions and a frequency that increases without an increase in error. Finally, we applied the complexity-guided regularized random forest (GRRF) to the set of distinct pruned rules in order to have two compact lists of stable rules (≤ 30)

able to explain the results of both algorithms. We grew 1000 trees, setting the importance threshold to 0.1 and using the default values for the other tuning parameters of the function `selectRuleRFF` in `R`. From a run of this function we obtained a list of 19 and 25 rules starting from the forests grown by the RF-GLS and RF algorithms, respectively. By applying both these lists of rules to test data we obtained a very good predictive performance: the percentage of variance explained was 92.01 and 90.17, respectively.

Table 2 and Table 3 show the two lists of the first ten rules output for the meteorological dataset. The scores [impRF] of the selected conditions are calculated by building an RF on the selected rules. In general, the two lists of selected rules have 17 rules in common. An example is represented by the first rule in Table 2 and Table 3. More specifically, the first rule in both lists shows that the interaction of a low latitude with a low elevation and a low distance to the coastline induces a higher mean daily temperature. The third rule in Table 2 (and then the fifth rule in Table 3) displays that the interaction of low longitude and a high elevation induces a mean daily temperature of about 9°C. This is composed of two conditions (lon<= 589199.5 & HRdem>317.40), and satisfied by the 14.8% of the observations in the training dataset and has an RMSE (the square root of "err") of about 2.2°C. One can notice that rule scores (importance values) and the rules metrics are not related. For instance, the fourth rule in Table 2 (and then the second rule in Table 3) has a larger frequency than the three most important ones.

Table 2: First ten rules extracted, measured, pruned and selected via GRRF, generated by RF-GLS. The rules are ordered by scores (importance value - ImpRF)

rule	len	freq	err	condition	pred	impRF
1	3	0.252	3.082	lat <= 4931735.37 & HRdem <= 609.20 & HRdsea <= 26.14	18.534	1
2	3	0.289	3.998	lon > 457787.2 & HRdem<=609.20 & HRdsea <= 26.14	18.160	0.893
3	2	0.148	4.882	lon<= 589199.5 & HRdem>317.40	9.325	0.673
4	2	0.681	5.564	lat>4780743.3 & HRdsea>1.34	12.885	0.602
5	2	0.230	2.472	lon>457787.2 & HRdsea<=1.34	18.714	0.586
6	3	0.148	4.882	lon<=620344.9 & lat>4873835.0 & HRdem>317.40	9.325	0.577
7	3	0.230	2.690	lat<=4931735.37 & HRdem<=317.40 & HRdsea<=26.14	18.743	0.505
8	3	0.148	4.882	lat>4873835.0 & HRdem>317.40 & HRdsea<=195.44	9.325	0.489
9	2	0.259	5.886	lat<=4931735.37 & HRdsea<=26.14	18.242	0.365
10	2	0.237	2.905	lat<=4931735.37 & HRdem<=317.40	18.645	0.347

Table 3: First ten rules extracted, measured, pruned and selected via GRRF, generated by RF. The rules are ordered by scores (importance value - ImpRF)

n	len	freq	err	condition	pred	impRF
1	3	0.252	3.082	lat<=4931735.37 & HRdem<=609.20 & HRdsea<=26.14	18.534	1
2	2	0.681	5.564	lat>4780743 & HRdsea>1.34	12.885	0.560
3	3	0.148	4.882	lon<=620344.9 & lat>4873835 & HRdem>317.40	9.325	0.487
4	3	0.148	4.882	lat>4873835 & HRdem>317.4 & HRdsea<=195.44	9.325	0.486
5	2	0.148	4.882	lon<=589199.5 & HRdem>317.40	9.325	0.475
6	3	0.267	5.879	lon>457787.2 & lat<=4982676 & HRdsea<=26.14	18.212	0.439
7	2	0.230	2.472	lon>457787.2 & HRdsea<=1.34	18.714	0.433
8	3	0.230	2.690	lat<=4931735 & HRdem<=317.4 & HRdsea<=26.14	18.743	0.366
9	3	0.207	1.914	lon>457787.2 & HRdsea<=1.34 & INSOL>8.082524	18.994	0.359
10	3	0.23	2.963	lon>503554.1 & HRdem<=609.2 & HRdsea<=26.14	18.723	0.294

4. Discussion and next steps

This work represents a first attempt to "open" an RF that is specifically designed for spatially dependent data, i.e. RF-GLS. This algorithm should be the gold standard in a spatial framework. We compared the predictive performance and explainability of RF-GLS and RF applied to a Croatian meteorological dataset. Both algorithms have shown high and similar predictive performance in our application. A cross-validation procedure will be implemented to confirm this result. Among the different approaches existing in the literature to obtain explainability from RF, we focused on the rule extraction methods. In particular, we considered the approach proposed by Deng (5) applying the same constraints to the node splits proposed in Bénard *et al.* (2). We found two compact lists of rules with high predictive performance sharing a large number of rules in common. However, the shared rules have different scores (importance values) within their respective membership lists. As next step, we aim to tune the GRRF hyperparameters to reduce the number of rules in the two lists while maintaining their predictive performance. Moreover, we aim to set up a comparison study considering the main competitors of inTrees, i.e. SIRUS (2), Node harvest (9) and RuleFit (6). Unfortunately, the `R` functions implementing RF-GLS (`RFGLS_estimate_spatial` and `RFGLS_predict_spatial`) return objects that are not valid inputs for the `R` functions implementing the competitor rule extraction methods. This will require further investigation.

References

[1] Aria, M., Cuccurullo, C., Gnasso, A.: A comparison among interpretative proposals for Random Forests. MLWA **6**, 100094 (2021)

[2] Bénard, C., Biau, G., Da Veiga, S., Scornet, E.: Interpretable random forests via rule extraction. In: Proceedings of the 24th International Conference on Artificial Intelligence and Statistics, pp. 937–945 (2021)

[3] Breiman, L.: Random forests. Machine Learning **45**, 5–32 (2001)

[4] Deng, H., Runger, G.: Gene selection with guided regularized random forest. Pattern Recognit. **46**, 3483–3489 (2013)

[5] Deng, H.: Interpreting tree ensembles with intrees. Int J Data Sci Anal. **7**, 277–287 (2019)

[6] Friedman, J. H., Popescu, B. E.: Predictive learning via rule ensembles. Ann Appl Stat. **2**, 916-954 (2008)

[7] Haddouchi, M., Berrado, A.: A survey of methods and tools used for interpreting random forest. In: Proceedings of the 2019 1st International Conference On Smart Systems And Data Science (2019) doi:10.1109/ICSSD47982.2019.9002770

[8] Hengl, T., Heuvelink, G.BM., Perčec Tadić, M., Pebesma, E.J.: Spatio-temporal prediction of daily temperatures using time-series of MODIS LST images. Theor Appl Climatol. **107**, 265–277 (2012)

[9] Meinshausen, N.: Node harvest. Ann Appl Stat. **4**, 2049–2072 (2010)

[10] Patelli, L., Cameletti, C., Golini, N., Ignaccolo, R.: A path in regression Random Forest looking for spatial dependence: a taxonomy and a systematic review. arXiv **2303.04693** (2023)

[11] Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., Zhong, C.: Interpretable machine learning: Fundamental principles and 10 grand challenges. Stat Surv. **16**, 1–85 (2022)

[12] Saha, A., Basu, S. Datta, A.: Random forests for spatially dependent data. JASA **118**, 665–683 (2023)

[13] Sekulić, A., Kilibarda, M., Heuvelink, G. BM, Nikolić, M., Bajat, B.: Random forest spatial interpolation. Remote Sens. **12**, 1687 (2020)

[14] Wikle, C., Datta, A., Hari, B., Boone, E., Sahoo, I., Kavila, I., Castruccio, S., Simmons, S., Burr, W., Chang, W.: An illustration of model agnostic explainability methods applied to environmental data. Environmetrics **34**, e2772 (2023)