Are Outages Homogeneous Among Nuclear Reactor Technologies? Evidence from Machine Learning Approaches

Stefan Seifert[a], Nicolas Landry[b], Christian von Hirschhausen[c], Astrid Cullmann[b], and Ben Wealer[c]

[a] *University of Bonn, Corresponding author: sseifert@uni-bonn.de*
[b] *DIW Berlin*
[c] *Berlin University of Technology*

Abstract

Nuclear electricity production has been constrained over the last decades by a low availability factor of plants, about only 2/3 on average. The factors underlying outages are not very well understood, but research thereupon is intensifying, both from an institutional and a technical perspective. This paper focuses on unplanned outages that particularly affect the availability of nuclear power plants adversely. We distinguish two broad types of reactors that have emerged in the industry, i.e., light-water reactors and gas-cooled reactors: whereas light-water reactors were designed to generate electricity only, gas-cooled reactors (in combination with graphite moderation) have a more complex design to obtain economies of scope between electricity generation and the extraction of plutonium. We dispose of a unique sample of 2534 reactor-year observations between 2003 and 2015 of plants in France, Germany, Japan, Spain, Switzerland, the UK, and the US. 15 classes of unplanned outages are considered. To identify outage-reactor relationships, machine learning algorithms for classification are used. Results show that unplanned outages are not homogeneous among different reactor technologies, but outage classes can be used to identify the underlying reactor technologies with considerable accuracy. These results are useful for a better understanding of nuclear technologies and their past, and perhaps future, developments.

JEL-Codes: L94, Q40, C45, C55

Keywords: Nuclear power plants, reactor technologies, unplanned outages, machine learning, tree-based classification

---

This paper combines a programmes on nuclear power and production economics at DIW Berlin and Berlin University of Technology. We thank the participants of the DIW Nuclear Economics Workshop as well as Chloé Le Coq, Anne Neumann, Julia Rechlitz, Sebastian Schwenen, and Aleksandar Zaklan for research support and comments. The usual disclaimer applies.

1 Introduction

Nuclear power was one of the most outstanding innovations of the 21st century, and continues to be an important energy source in many countries, such as the United States, Japan, and in many European countries. However, nuclear power plants have also been plagued by complex technical challenges that have limited their economic success; overall, the availability factor of nuclear power plants in the 1970s was only 65%, i.e., not only two thirds. Clearly, a better understanding of the outages of nuclear power plants is important.

The large number of nuclear power plants, over 600 have been built since the 1950s, and the detailed technical information collected and published, also make it an attractive field for data-driven research. The International Atomic Energy Agency (IAEA, Vienna, Austria) collects and publishes a significant amount of technical data, amongst them a very detailed analysis of outages, in not less than 15 sub-categories. Yet, a quantitative comparative analysis of this "big data" has not yet been undertaken.

This paper applies machine learning algorithms to analyze the nature of unplanned technical outages of nuclear power plants, and to look for a link between unplanned outages and the respective reactor technology. Our hypothesis, based on anecdotal evidence, is that due to different modes of operation, the unplanned outages of light-water reactors (the "standard" and one-purpose only technology) are less important than those of graphite-moderated gas-cooled reactors that have a more complex design. While the light water reactors were designed to generate electricity only, gas-cooled (and graphite moderated) reactors were developed after World War II to co-produce electricity and plutonium simultaneously to benefit from economies of scope (Hirschhausen, 2017). We focus on unplanned outages because these are likely to reveal insights into the underlying production technology; the structure and the large number of available data invites to use machine learning algorithms.

There is some literature on the determinants of outages in nuclear power plants, and it has recently attracted particular attention. Thus, Davis and Wolfram (2012) show that most variation in reactor performance can be explained by variation in outages. Further, due to the low variable costs of nuclear power plant operation, outages lead to considerable losses of operating profits (Hausman, 2014). However, studies (Deutch et al., 2003, 2009) show that outages are not sheer bad luck, but a result of effort. And indeed, Zhang (2007) and Davis and Wolfram (2012) find considerably increased capacity factors since the 1980, i.e., a higher availability of nuclear power plants and reduced outages. Thereby, Zhang (2007) argues that longer shutdowns (exceeding half a year) are used for investments resulting in increased plant availability.[1] Results by Davis and Wolfram (2012) also indicate that increasing competition, deregulation, and consolidation improve plant availability. The authors argue that plant operators' responsibility for costs of outages incentivizes better plant management, including especially the employment of highly qualified labor (with scale effects) as a major determinant of outages. Lastly, studies indicate that maintenance that reduces outages has complementaries with nuclear power plant safety (Hausman, 2014; Deutch et al., 2003, 2009). As Hausman (2014) shows, divestiture and privatization may therefore increase power plant safety due to the incentives to reduce outages.

From a methodological perspective power plant outages data are challenging to analyze due to the the strong heterogeneity in outage occurence, due to a large number of different outage types, and due to potential interactions of outage types depending on reactor technology. Machine learning approaches are one way to overcome such limitations and have been used more recently also in the analysis of energy generation technologies (see, e.g., Voyant et al., 2017; Sharma et al., 2011). However, a large number of such algorithms exists and *ex-ante* it is not clear which perform best (Ruiz-Gazen and Villa, 2007; Grömping, 2012; Hu et al., 2012). This paper analyzes whether unplanned outages are homogeneous among different types of reactor technologies. Thereby, we argue that if reactor technology prediction from outage types is possible with considerable accuracy, unplanned outages are not homogeneous among technologies. For this purpose, three tree-based machine learning algorithms for classification, namely classification trees (Breiman et al., 1984), random forests (Breiman, 2001), and boosting (Freund and Schapire, 1997; Wang, 2011), are used. Each of these methodologies trains a classifier using only a subset of the data (training set). This classifier is then used to predict the reactor type with the second part of the sample (test set), for which the actual reactor type is assumed to be unknown. If correct predictions with the classifiers are possible, outages are not homogeneous but are specific to the reactor technologies. To optimize algorithms' prediction accuracy, for each estimator a large number simulations are carried out with varying parameter settings, accounting for characteristics of the sample and characteristics of the classifiers.

The empirical application uses a newly collected sample of 2534 reactor-year observations of light-water (LWR) and gas-cooled reactors (GCR) in France, Germany, Japan, Spain, Switzerland, the UK, and the US between 2003 and 2015. In total, 165 GCRs and 2369 LWRs reactor-year observations are included and cover in total between 127 and 260 GW GW total annual gross capacity, i.e., 30 to 69% of total world wide nu-

[1] While investments increase current availability, power plants needs to be evaluated with lifetime availability, as investments must be covered over the lifetime (Joskow and Parsons, 2009; Linares and Conchado, 2013).

clear capacity. For each observation, annual outage hours for 15 different categories are considered and differentiate in detail the sources of outages, such as, e.g., reactor instrumentation and control systems (I & C) or reactor cooling.
Our results show that an optimized tree can predict the reactor type from a random outage profile with considerable accuracy and prediction accuracy lies between 75 and 82 percent, with partly very low standard deviations. This means that outage types are not homogeneous among different reactor types, but that outage profiles vary considerably between technologies. Methodologically, our results show that, in the given setting, more complex algorithms do not automatically outperform their "simpler" counterparts. However, our results underline that a thorough exploration of potential parameter settings is necessary to calibrate the algorithms to fully use their potential.
The remainder of the paper is structured as follows: Section 2 describes the methods used in the paper and outlines the implementation of the methods. Section 3 describes the data. Results are presented and discussed in section 4, and section 5 concludes.

2 Nuclear Power Technologies and Outage Rates

2.1 Nuclear power technologies

Nuclear fission energy is released when an atom of uranium is split, producing several new chemical products, radioactivity plus a large amount of energy (heat); the fast neutrons cause a chain reaction as they hit other uranium atoms. The speed of this chain reaction can be controlled by a "moderator" that slows down the neutrons; moderators include graphite and light water. The spent fuel contains a large portion of unused fuel (uranium238 and uranium235), plutonium, and other products such as barium, strontium, and cesium. Nuclear bombs can be produced from both highly enriched uranium and plutonium. When analyzing nuclear technologies one needs to go back to the origins of nuclear power, i.e. the late 1940s and the 1950s. Nuclear technologies developed after World War II had to pursue multiple objectives, and this is still the case today for many countries. On the one hand, nuclear fission allows generating electricity, but on the other hand, it can also be used to produce fissile material used in nuclear bombs. The priorities of a country, therefore, determine the choice of technologies, i.e. towards a more civilian or a more military use of nuclear power (Hirschhausen, 2017).
Concretely, nuclear reactor technologies can be distinguished by the fuel used (either natural uranium238 with approx. 1% of fissible U235, or enriched uranium, U238 with about 3-5% U235), and three basic combinations of moderators, that is, elements

steering the chain reaction, and coolants, that is, elements transporting the heat to be used for electricity production. Gas-cooled graphite moderated reactors (GCRs) are able to produce a broad range of electricity: plutonium ratios, thanks to the possibility to change the fuel rods in continuous mode. GCRs were the dominant technology in the 1950s, and were deployed, amongst others, in the Soviet Union, the UK, and France. On the other hand, light water reactors are designed to produce "only" electricity, and the extraction of plutonium is much more complicated (though still theoretically possible). Light water reactors were developed quite some time after the GCRs, by the United States, but have taken the lead in the meantime, as the dominant technology; they are used, amongst others, in the United States, France, and Japan.

2.2 Capacity utilization and power plant outages

Between 1970 and 2014, about 77,044 TWh of electricity have been produced using different nuclear reactor technologies. Surprisingly, the capacity utilization of all nuclear power plants between 1970 - 2014 was only 63%. In other words, about three fifths of the available capacity of nuclear power plants was used over the years. The mean capacity utilization ratio of different technologies provides some support to the hypothesis of nuclear power as mainly military driven: we observe an inverse relation between the capacity utilization and the ease of plutonium production.[2] In addition to running below capacity, the main explanation for these low figures are the high outage rates of nuclear technology (see below for details).

Graphite and light water moderated reactors show very different characteristics, when it comes to the capacity utilization (this observation in fact motivated our research). Both the means and the structure of the distribution are significantly different from each other: a t-test on the different of means has a significant result (t-value: 10.054, p-value: 2.2e-16). The Kolmogovor-Smirnov test, too, suggests that the distribution of the two technologies is different and that the technologies are used differently (p-value of 0.02566).

Unused capacities seem to prevail in all nuclear technologies used, but there seems to be a structural difference between GCRs and light water reactors. This has not only technical relevance, but economic implications as well: a nuclear power plant constrained by outages is less profitable. An enquiry about the nature of outages therefore seems justified.

[2]In fact, graphite reactors were used for electricity generation not even half of their capacity (48%), whereas light water reactors clearly lead the pack, with 65%, just short of 2/3 of their available capacity.

3 Machine Learning Algorithms

To analyze the relationship of outage types and reactor technologies, this paper uses three tree-based machine learning algorithms for classification, namely classification and regression trees (Breiman et al., 1984), random forests (, Breiman, 2001), and boosting (Freund and Schapire, 1997; Wang, 2011). The general idea of such supervised learning methods is to train an algorithm to predict an outcome variable Y with a set of p features x_p ($x = (x_1, \ldots, x_p)$). To construct a prediction model, the training of the algorithms uses a share of the total sample for which Y and x are observed (training set). The derived prediction model is then used with the remaining share of the sample (test set), for which only x is used to predict Y for this subset. The prediction accuracy of this second stage informs about the generalizability of the first stage results.

These statistical learning methods have several advantages over standard regression models. First, they can easily deal with continuous, categorical, and binary data. Second, and contrary to standard regression models, the algorithms are non-parametric and can account for nonlinearities without explicitly modeling them. Thus, the methods can deal with cases in which explanatory variables are not correlated with the observed outcome, although a relationship exists. This allows to also account for interdependencies of predictors. And third, these algorithms are popular because they can handle cases with a large number of variables and a small number of observations.

3.1 Classification Trees

Classification and regression trees (CART), first introduced by Breiman et al. (1984) derives a decision tree as a predictive model by partitioning the sample into more and more homogeneous groups. Figure 1 illustrates the concept of a decision tree with a simple example containing a two-dimensional outcome ($Y = 1, 2$) and a two-dimensional feature space ($x = (x_1, x_2)$). In this figure, the dataset sits above the tree, and observations that satisfy the condition at a junction are assigned to the left branch, others are assigned to the right branch. The sample is partitioned down to the terminal nodes (leaves), referred to as regions, that contain predictions of the outcome Y. Classification trees are usually using binary splitting, i.e., splitting into only two branches at each node. This ensures that sufficient data for further splits is available at the next node. Further, using a series of binary splits can also resemble multiway splits, as shown in Figure 1 by the repeated use of x_1 to partition the data. Once the tree has been built, the outcome variable Y can be predicted for each observation by following the criteria at the junction.

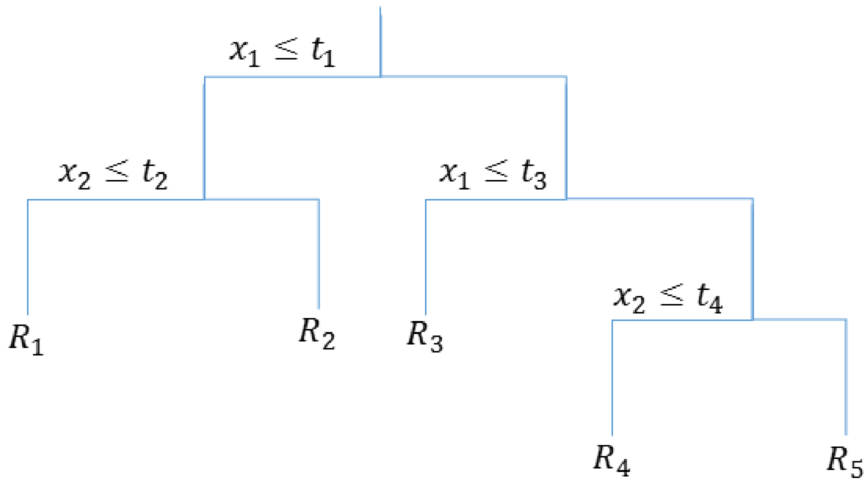

Figure 1: Examplary classification tree following Friedman et al. (2009)

Formally, a tree $T = T(x; \theta)$ is characterized by the predictor variables x, and by θ containing split variables, cut points at the nodes, and terminal node values. To derive the tree, CART first grows a large tree and defines the splitting criteria at the junctions. Thereby, CART guarantees a certain number of observations in each node and splitting only if a certain number of observation are available to split. Both parameters influence tree complexity and are control parameters to avoid overfitting. To determine the splitting criteria, CART seeks to minimize node impurity, whereas impurity of a node can be evaluated with several measures, such as the misclassification error, the Gini index, or as cross-entropy (Friedman et al., 2009). In a second step, to derive the final tree, the initial tree is pruned, i.e., branches are cut off, if they do not add predictive performance according to a pruning criterion. For this purpose, cost-complexity pruning is usually used, which accounts for trade-off between a large tree (complexity) and the goodness of fit with the data. Thereby, the pruning algorithm considers costs of misspecification and costs of tree size weighted with a tree complexity parameter α. For small α the costs of tree size obtain a low weight resulting in a larger tree, and vice versa.

3.2 Random Forests

A random forests (RF, Breiman, 2001) is a collection of multiple trees. The aim of RF is to reduce the variance of the prediction function by reducing the correlation among the different trees. To do so, two random factors are introduced in the construction of a tree: first, only a subset of observations is used in each tree. And second, only a random subset of the explanatory variables is used for the splits. As a result, the trees are rather unstable and deliver different predictions. However, overall prediction of the forest, derived as average over the different trees, is stable with a reduced variance.

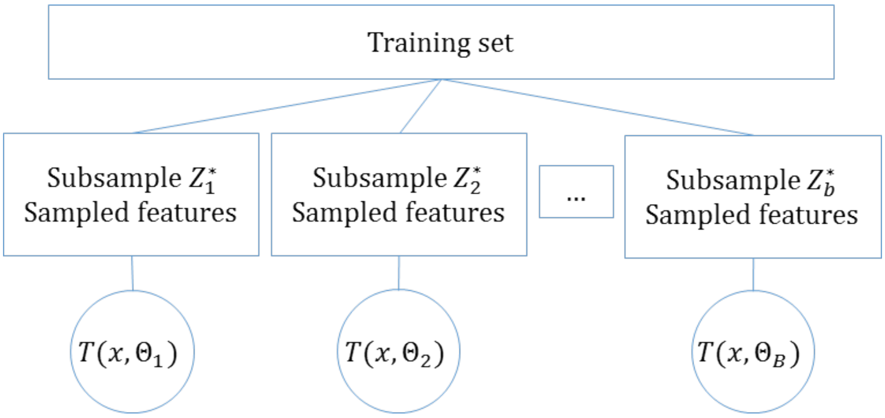

Figure 2: Structure of random forest estimation

More formally, RF grows B $(b = 1, \ldots, B)$ trees $T(x, \Theta_b)$. For each b, a bootstrapped sample Z_b^* of size N is drawn from the training data. To grow $T(x, \Theta_b)$, at each terminal node m variables are randomly sampled from the p predictor variables, the best variable/split-point among the m is derived, and the node is split into two daughter nodes. This is repeated until a minimum node size is reached. The resulting set of trees, $\{T_b\}_1^B$, is the random forest used for prediction. In a regression context, prediction is derived as the average of the predictions of the B trees. In a classification context, the most frequent class prediction is used (majority vote, see Friedman et al., 2009).[3]
While the prediction accuracy of a RF model depends on several parameters, some issues concerning the parameter setting should be noted. First, the accuracy of RF crucially depends on the number of variables selected for growing the tree, m. With a

[3]To analyze the impact of features on the outcome, variable importance measures can be used (Grömping, 2012). However, it should be noted that the power of variable importance measures depend on the scaling of the data (Strobl et al., 2007), which is, however, not an issue in our application.

smaller m correlation among the B trees decreases, which also decreases the variance of the average. However, a too small m, e.g., $m = 1$, would randomly select the splitting variable without taking into account predictive power. In contrast, to large m, e.g., $m = p$, reduces the randomness of the forest. Second, the minimum node size n_{min} necessary to further split a node influences the size of a tree. Increasing node size can strengthen prediction accuracy (Segal et al., 2004), however, increasing node size will also lead to a stronger dependence on m. Both parameters therefore need to be set carefully and simultaneously. And third, the size of the bootstrapped subsample to grow the tree as well as the number of trees are parameters that influence prediction accuracy. For prediction, out of bag (OOB) samples can be used, such that prediction for an observation uses only trees in which this observation is not contained to derive the tree. This allows the analysis of misclassification errors with B, and training can be terminated once the OOB error stabilizes.

3.3 Boosting

Boosting (BO) algorithms aim on combining outputs of many weak classifiers to produce a powerful committee (Friedman et al., 2009; Freund and Schapire, 1997).[4] BO grows, similarly to RF, a larger number of trees. However, contrary to RF, these trees are not independent from each, but are sequentially generated using sequentially modified version of the data. As a result, boosted trees with very weak classifiers, such as shallow trees (trees of low depth) or stumps (trees with only one node), can still outperform other classification methods.
Boosting procedures try to identify a final prediction $G(x)$ as the weighted average of L classifiers G_l such that $G(x) = \text{sign}(\sum_{l=1}^{L} \gamma_l G_l(x))$. Thereby, the boosting algorithm computes the weights $\gamma_1, ..., \gamma_L$ with the aim to put higher weight on more accurate predictors (compare Figure 3).

During each of the L boosting steps, the dataset is modified using weights $w_1, ..., w_N$ for each of the N observations. Initially all observations are equally weighted, i.e., $w_n = 1/N$. After each evaluation of a classifier $G_{l-1}(x)$ the weights are adjusted and misclassified observations obtain a higher weight in $G_l(x)$, while weights decrease for correctly classified observations. As a result, $G_m(x)$ focuses especially on those observations, for which prediction failed in the step before. Thus, the updated weights change especially in cases when prediction accuracy is high, because more importance is put on few observations.

[4] On overview of recent methodological developments of boosting algorithms is provided by Mayr et al. (2014a) and Mayr et al. (2014b).

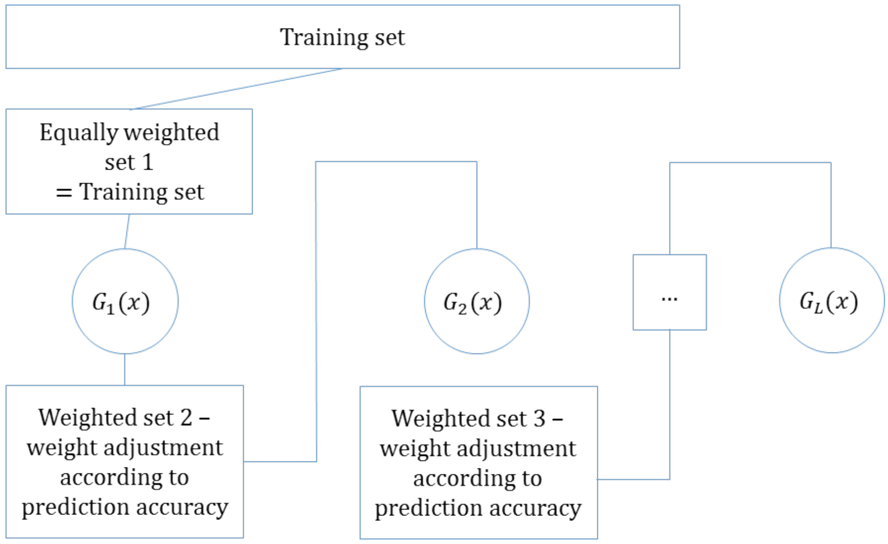

Figure 3: Structure of boosting algorithms

Several parameters influence the accuracy of the resulting weighted tree. First, the scheme used to re-weight the observations influences the results of the next tree, and several re-weighting mechanisms are discussed in the literature (see, e.g., Freund and Schapire, 1997; Zhu et al., 2009; Breiman, 1998). Second, the number of iterations of the boosting procedure influences the prediction accuracy. As in each iteration performance for the training set increases, a large number of iterations may overfit the training data leading to worse prediction accuracy with the test set. Thus, the optimal number of iterations is usually chosen to be the point at which prediction accuracy is maximized. Alternatively, shrinking approaches limit the the learning speed of the algorithm by slowing down the adaptation of the classifiers between iterations. This, however, demands a larger number of iterations to obtain optimized prediction (Friedman, 2001). Third, boosted trees should be of right size as too large trees are computationally burdensome and degrade performance. Therefore, tree size is usually restricted such that it allows for interaction between multiple predictor variables (increasing tree size) while limiting maximum tree size.

The empirical analysis in this paper is based on HingeBoost as introduced by Wang (2011). This gradient-based approach minimizes a hinge-loss function, which is approximately equal to maximizing the area under the curve (AUC), as shown by Steck (2007). An advantage of using HingeBoost is, first, that it allows to incorporate unequal misclassification costs, i.e., false positive and false negative predictions can be unequally weighted. Second, and contrary to, e.g., the widely used AdaBoost, HingeBoost allows a deeper understanding of the relationship between different features to predict the outcome (Mayr et al., 2014b).

3.4 Implementation

As outlined above, the prediction accuracy of the three ensemble learning methods used in this paper depend all on a large number of different parameters. To obtain parameter settings that deliver optimal predictions, simulations are used for a large range of parameter combinations. In total 198.000 classification trees, 9625 random forests with up to 2800 trees each, and 123.600 boosted trees are calculated with 30 replications each. All simulations and calculations are carried out with R (R Core Team, 2015) and the packages rpart (Therneau et al., 2015), randomForest (Liaw and Wiener, 2002), and bst (Wang, 2013).

An overview of the methodology-specific parameters settings is provided in Table 1.[5] For all three methodologies, simulations with different sizes of the training set are carried out. To account for differences in the sample sizes, sampling techniques are used to upscaling the smaller subsample and downscaling the larger subsample (Liu et al., 2006). To account for randomness as, e.g., differences in the sampling, 30 repetitions are calculated for each potential parameter combination. Parameter combinations leading to infeasible models to estimate are dropped.

To evaluate prediction accuracy of these many trees, we simultaneously consider the share of correct predictions in each of our technologies using two measures. Sensitivity ($Sens$) gives the share of correct predictions of the subsample of GCRs, while specificity ($Spec$) gives the share of correct predictions of the subsample of LWRs. Mean sensitivity $\overline{Sens}$, mean specificity $\overline{Spec}$, and their standard deviations (σ_{Sens} and σ_{Spec}) are then calculated for each parameter setting given the 30 replications. As potential best predictions, we first consider only those cases that predict both technologies similarly well, such that $|\overline{Sens} - \overline{Spec}| < min(\sigma_{Sens}, \sigma_{Spec})$. The reported results are then those with maximized prediction accuracy, $\overline{Sens} + \overline{Spec}$.

4 Data

To analyze a potential relationship between unplanned outages and reactor technologies, we dispose of a sample of nuclear power plants (NPPs) in France, Germany, Japan, Spain, Switzerland, the UK, and the USA. Data has been obtained from the International Atomic Energy Agency (IAEA) database on NPPs, the Power Reactor Information System (PRIS, IAEA, 2013), and from IAEA's annual reports on the operating experience with nuclear power plants (OPEX reports, IAEA, 2016). Table 2 provides summary statistics for the aggregate sample. Detailed descriptive statistics

[5] Not-listed parameters follow the default values of the used software packages.

	Parameter	Simulated Range
CART	Min. number of obs. to split	[3,5,...,39]
	Min. number of obs. in leaves	[1,3,...,19]
	Tree complexity	[10^{-5}, 10^{-4},...,10^{-1}]
	Training set size upsampling	[100,200,...,1100]
	Training set size downsampling	[10,30,...,190]
	Share of GCRs	[0.1,0.2,...,0.8]
RF	Number of trees	[200,300,...,2800]
	Number of candidates at each split	[1,2,...,7]
	Class priors	[16,32,...,240]
	Max. number of terminal nodes	[2,4,...,14]
	Subsample size upsampling	[50,125,...,1100]
	Subsample size downsampling	[50,75,...,200]
	Share of GCRs	[0.1,0.2,...,0.8]
BO	Iterations	[10,20,...,100]
	Training set size upsampling	[300,600,...,1200]
	Training set size downsampling	[50,100,...,300]
	Tree complexity	[0.001,0.01,0.1]
	Share of GCRs	[0.2,0.4,0.6]
	Shrinkage Parameter	[0.1,0.2,...,0.9]
	Missclassification cost parameter	[0.3,0.4,...,0.7]

Table 1: Simulated parameter settings

are provided in the appendix.
In total, the sample consists 2534 reactor-year observations between 2003 and 2015 and includes 271 different reactors. The sample is unbalanced and contains 87 to 170 observations in every year. It accounts for shutdowns of NPPs in Japan (after the Fukushima Daichii incident) and in Germany in 2012. In total, between 127 and 260 GW gross capacities are covered in each year, representing between 30 and 68% of total capacities of worldwide nuclear power plants. 165 reactor-year observations are GCRs, while the remaining major share of observations are LWR. All GCRs are located in UK. Except for Japan's NPPs, GCRs show a comparably low average availability during the observation period.
To separate different sources of outages, we differentiate outage categories following the IAEA classification of the operating experience with nuclear power plants. This allows to separate external reasons for outages, i.e., outages that are not under control of the management, from unplanned outages due to failures coming from the NPP itself. 15 different categories of unplanned outages are considered and detailed definitions of the categories are in the appendix, Table 8. Descriptive statistics of the different outage types separate for LWRs and GCRs are given in Table 3. Generally, data shows that only about 7% (LWRs) to 10% (GCRs) of the NPPs have neither a planned nor an unplanned outage throughout a year. While planned outages are nearly of identical occurrence, unplanned outages have generally a higher duration among GCRs. However, the share of NPPs with no planned outages is about 30% of the GCRS, while only about 19% of the LWRs. On the contrary, the share of GCRs without any unplanned outages is already considerably smaller (19%) than for the LWR category (30%). Further, data indicates reactor-technology specific accumulation of outages in two categories, with Reactor & Accessories and Steam Generation showing the most severe discrepancies. However, the different outage categories show overall very low correlations, independent from the reactor type. Similarly, the data also does not indicate noteworthy correlations between planned outages and different types of unplanned outages or total unplanned outages of a plant.
For the estimation, only observations with non-zero outages are used. This reduces the size of the total sample to 1544 reactor-year observations, with 119 GCRs and 1455 LWRs. Further, the model specification of the machine learning algorithms uses the reactor types as categorical variable to be predicted, with the category LWR containing BWRs (boiling water reactor) and PWR (pressurized water reactors). The features of the machine learning models, i.e., the outages, are incorporated as outage shares relative to the hours per year (controlling for leap years), corrected for planned outages,

and corrected for other outages. These shares are calculated as

$$shareOutage_{pt} = \frac{Outage_{pt}}{Hours_t - plannedOutages_t - \sum_{q \neq p} Outage_q}$$

with all variables measured in hours. The rational behind using these corrected shares is the following: First, controlling for planned outages controls also for the (country-specific) regulatory environment. Second, during times in which an NPP has a planned outage, it cannot have an unplanned outage. Third, a plant that is shutdown due to an outage of category q cannot shut down for another reason. Thus, the estimation uses only the shares of the different outage categories relative to the amount of time an outage of a certain category can actually occur.

	CHE	ESP	FRA	GER	JAP	UK	USA
#NPPs	5	9	58	17	55	23	104
GCR obs.	0	0	0	0	0	165	0
LWR obs.	65	106	696	136	426	9	931
Avg. capacity (MW)	697.0	963.5	1135.9	1307.3	892.4	614.6	1049.0
Avg. availability (%)	88.4	86.9	78.2	81.6	64.3	67.2	90.0
Avg. age	34.2	26.3	23.9	27.2	22.4	29.3	30.7

Table 2: Aggregate descriptive statistics 2003 - 2015

Outage category	GCR Mean	GCR Max	GCR #0	LWR Mean	LWR Max	LWR #0
Total Planned	1087.05	8784	32	1075.32	8784	724
Total Unplanned	1210.30	8784	56	430.24	8784	412
Reactor & Accessories	277.39	8784	148	34.18	8784	2190
Reactor I&C	15.80	386	147	23.24	4155	2001
Reactor Auxiliary	7.19	744	160	16.10	7392	2156
Safety Systems	1.48	144	163	10.86	3357	2209
Reactor Cooling	28.36	1374	155	39.67	5976	2128
Steam Generation	241.49	4368	131	20.53	8046	2227
Safety I&C	1.44	237	164	0.33	486	2359
Fuel & Storage	79.38	3810	145	3.13	1200	2242
Turbine & Auxiliaries	54.90	1217	137	42.34	8453	1759
FeedWater	38.65	1042	135	23.82	8784	1996
Circulating Water	65.81	4332	151	3.72	1296	2304
All others I&C	12.80	2112	164	1.90	789	2290
Main Generator	52.88	816	139	35.10	3716	2045
Electrical Power Supply	65.48	1449	133	31.41	7297	2050
Others	6.95	864	161	29.99	2099	2066

Table 3: Descriptive statistics: Outages in h per outage category per year, 2003 - 2015

5 Results

5.1 Simulation results

The results of the simulations are summarized as sensitivity/specificity plots for all three methods in Figure 4. In each of these plots, one dot indicates specificity (correct predictions of LWRs) and sensitivity (correct predictions of GCRs) averaged over the 30 replications for this parameter settings. Thus, an optimal prediction would be located in the top-right corner, in which all observations of the test set are correctly predicted with the classifier based on the training set.

The plots show some interesting patterns for each methodology. For CART, we see that prediction of LWRs is generally better than the prediction of GCRs. However, prediction accuracy exceeding 75% accuracy for both technologies is possible with large, complex trees with a high (bootstrapped) share of GCRs to outweigh the skewed distribution of the two reactor types. Thus, upscaling of the smaller subsample leads to generally improved results. Further, the plots indicate that under certain parameter settings, optimal predictions of both technologies are possible, i.e., specificity or sensitivity equal 1. Thereby, an optimal prediction of LWRs is likely with settings that

favor the larger subsample, e.g., if LWRs are overrepresented in the sample (i.e., there is a low share of GCRs in the training set), or if the tree demands final nodes with many observations. On the contrary, GCR's can only be only optimally predicted, if training sets are very small and their share in the training set is greater than 50%.
For random forest, Figure 4 indicates a similar distribution of potential specificity and sensitivity values. However, compared to CART, slightly better predictions of both technologies are possible, which, additionally possess considerably smaller variance. Generally, as expected, prediction accuracy increases with the number of trees in the forest, although these gains are already small with more than 1000 trees. However, noteworthy gains in terms of reduced standard deviation can still be achieved afterwards. Also more complex trees (increased number of maximum final nodes), with a larger set of candidates at each split can lead to gains in prediction accuracy, while no considerable increase in terms of variance can be observed. Similar to CART, prediction accuracy improves for upsampling of the smaller subsample. For RF, the results also underline the importance of optimized parameter settings: among the 100 best predictions (in terms of the sum of specificity and sensitivity) stem from only 13 different scenarios with slight variations in only one parameter.
For boosting, result show different patterns compared to CART and random forests. Figure 4 indicates boosting predicts typically the larger subsample of LWRs much more accurately for the the chosen parameter settings.[6] An analysis of the impact of the different parameters on predictions shows a complex interplay of the different determinants. We observe that an increasing training set size improves prediction accuracy. However, an increase of training set size with bootstrapped upsampling generally only increases predicitve power regarding LWRs, even with a high share of GCRs. Thus, prediction accuracy seems to be already limited by sample properties. The parameter to set costs for misspecification has the expected effect, and higher costs for misclassifying one technology increases its share of correct predictions. On the contrary, we find only small effects of decreasing the shrinkage parameter and also no considerable effect of using a very large number of trees. With this respect, boosting seems to obtain stable results fairly quickly, whereas their quality depends crucially on the characteristics of the training set, leading also to partially considerable variance of the results.
To summarize, the simulation results underline the importance of optimal parameter settings to obtain classifiers with predictive power. Each methodology needs calibration and different parameters need to be set carefully, and simultaneously. In term of prediction accuracy, our results suggest that all methods can perform well. However, random forests are preferred if results with low variances are favored.

[6]Extreme values for misclassification have not been simulated, but may also allow to correctly predict only GCRs correctly.

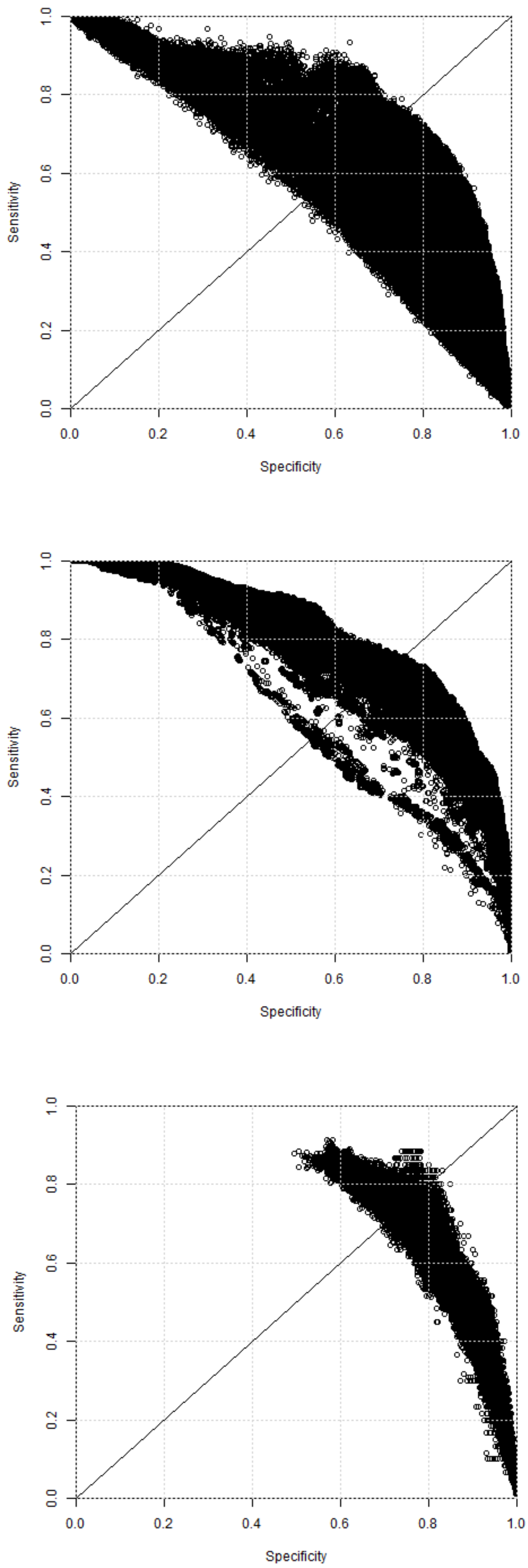

Figure 4: Simulation results: Sensitivity and specificity for CART (top), RF (center), BO (bottom)

5.2 Best predictions

The best predictions of the three machine learning algorithms are collected in form of confusion matrices in Table 4. Best predictions satisfy the criteria outlined above: as we aim on simultaneously predicting both reactor types, only those predictions are taken into account, for which $|\overline{Sens} - \overline{Spec}| < min(\sigma_{Sens}, \sigma_{Spec})$. The confusion matrix reads as follows: the main diagonal of each matrix contains the shares of correctly identified reactor types from the test set, i.e., GCRs that are predicted to be GCRs, and LWRs correctly identified as LWRs (specificity and sensitivity). On the contrary, the counterdiagonal contains the missclassified cases.

For CART, Table 4 shows prediction accuracies between 75 and 74% for LWRs and GCRs for both sampling techniques, meaning that based on the tree derived with the training set, observations from the test set are correctly identified in about three out of four cases. Thus, a fairly accurate prediction of reactor technologies on the basis of outage categories is possible already with simple classification trees, however, also with a certain variation in the accuracy. The table also indicates that the upsampling of the smaller GCR category helps correctly identifying both technologies, although gains are small. The parameters of the optimal trees, given in Table 5, indicate identical tree complexity independent of the sampling methods. However, due to the much larger number of observation under upsampling, considerably larger leaf and split sizes are necessary. Further, with upsamling the share of GCRs in the training set is upsampled to exceed LWR frequence considerably.

For RF with upsampling, models show a predictive power comparable to the simple classification trees, while results under downsampling are comparably worse compared to CART. The results using upsampling are, however, due to the large number of replications with additional randomization coming from subsampling and random variable selection, much more robust than results obtained with CART and show strongly reduced variance. This variance reduction also explains that random forests do not outperform classification trees: by introducing a variance-based threshold for the selection of the best predictions, a considerable number of repetitions has been dropped. As a result, for CART around 5% of all simulated trees are considered to identify the best prediction, compared to only 1% of the simulated settings with random forest. This result does not hold under downsampling, in which the results indicate deteriorating performance with increased variance. Table 5 indicates that under upsampling results favor a large number of trees, which also reduce variance. On the contrary, with downsampling, optimal results are obtained already with small forests.

For boosting with downsampling results are summarized in the Table 4, while no predictions with upsampling fulfill the criterion of equally predicting both reactor technologies. As outlined above, we observe that in upsampled settings LWRs are generally better predicted with small variance, while accuracy for GCRs is low, leading to cases that violate the criterion despite the large variance of the results. Downsampling results, however, indicate on average a better prediction accuracy compared to all CART and random forests, with, however, considerably higher variance as the other two methods. The results indicate that prediction accuracy depends especially on the sample size of the smaller subsample, while the combination of many parameters settings allow to achieve high prediction accuracy. This is an interesting finding because boosting algorithms were especially developed for weak classifiers and our results support this underlying idea. However, our results indicate that also more complex classifiers (more complex trees) can lead to similar predictions and, if computational power is not a burden, might avoid oversimplification.

			Upsampling		Downsampling	
			GCR	LWR	GCR	LWR
CART	Prediction	GCR	76.31% (6.30 %)	23.02% (4.72 %)	76.49% (12.19 %)	26.24% (4.34 %)
		LWR	23.69% (6.30 %)	76.98% (4.72 %)	23.51% (12.19 %)	73.76% (4.34 %)
RF	Prediction	GCR	75.69% (1.27 %)	24.13% (1.88 %)	68.80% (5.87 %)	27.56% (9.01 %)
		LWR	24.31% (1.27 %)	75.87% (1.88 %)	31.20% (5.87 %)	72.44% (9.01 %)
BO	Prediction	GCR	NA (0.00 %)	NA (0.00 %)	83.33% (27.33 %)	16.99% (2.03 %)
		LWR	NA (0.00 %)	NA (0.00 %)	16.67% (27.33 %)	83.01% (2.03 %)

Table 4: Best predictions

	Parameter	Upsampling	Downsampling
CART	Min. number of obs. to split	35	3
	Min. number of obs. in leaves	13	3
	Tree Complexity	0.01	0.01
	Training set size	1060	190
	Share GCR in training set	70%	50%
RF	Number of trees	2400	200
	Number of candidates at each split	5	5
	Class priors	216	1
	Max. number of terminal nodes	8	4
	Subsample size	950	200
	Share of GCRs	50%	10%
BO	Iterations		20
	Training set size upsampling		275
	Training set size downsampling		200
	Complexity		0.01
	Share of GCRs		50 %
	Shrinkage Parameter		0.3
	Missclassification cost parameter		0.4

Table 5: Parameter settings for optimal prediction accuracy

5.3 Classification threshold analysis

In the next step, we analyze the sensitivity of our results with respect to the chosen classification threshold. This threshold determines the necessary probability to be classified as GCR given the features (outages) of an observation and the classifier. By sliding this threshold, additional gains in prediction accuracy are possible, especially in cases with severe sample imbalances as in our case (Ruiz-Gazen and Villa, 2007; Lemmens and Croux, 2006).

To implement these models, we replicate the simulations with a subset of the parameter settings used before accounting for sample skewness using up- and downsampling and stochasticity using 30 replications per parameter setting.[7] In these simulations, the classification threshold is chosen *ex-post* such that it optimizes prediction accuracy for the binary classification GCR/LWR for a given parameter setting. For CART and boosting, the sliding threshold is implemented to optimize prediction with the training set for the given tree. This threshold is then applied to the test set to evaluate prediction accuracy. For random forests, we chose the optimized threshold using the in-bag observations, which is then applied to the out-of bag sample.

Figure 5 shows the sensitivity / specificity plots for the three methodologies with flexible classification thresholds. The figures show that with the sliding threshold simultaneous prediction of both technologies increases, while cases with maximized prediction of only one technology do not occur. Further, the plots already indicate that prediction accuracy can increase. Especially for boosting we observe that results with sliding threshold favor much more the prediction of the GCR class.

The best predictions following our criterion to simultaneously predict both technologies are summarized in Table 6. For CART, we observe an slight increase of prediction accuracy in the upsampled cases, while prediction accuracy decreases to the standard settings with downsampling, but remains in one standard deviation of the results without the flexible threshold. For random forest, our results show a strong increase of prediction accuracy with both upsampling and downsampling. In both settings, both technologies are predicted with more than 82% accuracy, indicating that a distinction of technologies based on outages is possible. Finally, with boosting results using downsampling indicate again prediction accuracy of around 80%, with, however, considerable variance. The reported upsampling result is only one of two parameter settings under which the criterion is fulfilled, which clearly results from the large standard deviations. We therefore conclude that upsampling in boosting algorithms does not improve pre-

[7] 198.000 CART trees, 15.000 random forests, and 123.600 boosting procedures are estimated with 30 replications each. Parameter settings are motivated from the first part of the study with fixed thresholds.

			Upsampling		Downsampling	
			GCR	LWR	GCR	LWR
CART	Prediction	GCR	76.31% (6.30 %)	23.02% (4.72 %)	76.32% (12.27 %)	25.71% (4.41 %)
		LWR	23.69% (6.30 %)	76.98% (4.72 %)	23.68% (12.27 %)	74.29% (4.41 %)
RF	Prediction	GCR	0.00% (2.61 %)	15.29% (2.81 %)	82.86% (3.27 %)	15.87% (2.76 %)
		LWR	16.53% (2.61 %)	84.71% (2.81 %)	17.14% (3.27 %)	84.13% (2.76 %)
BO	Prediction	GCR	43.44% (30.03 %)	27.65% (36.09 %)	83.33% (30.32 %)	20.05% (4.14 %)
		LWR	56.56% (30.03 %)	72.35% (36.09 %)	16.67% (28.16 %)	79.95% (5.96 %)

Table 6: Best predictions with flexible threshold

diction accuracy with the data used in this paper.
The impact of the parameter settings is very similar to the results presented in sections 5.1 and 5.2. Moreover, the effect of the parameters aiming on reducing the impact of sample size disparities, namely class weights for random forests and the misclassification cost parameter for boosting have the same effect using the flexible threshold as before. Since we observe a considerable increase in prediction accuracy especially for random forest a combination of different approaches to tackle such sample issues seems to be useful.

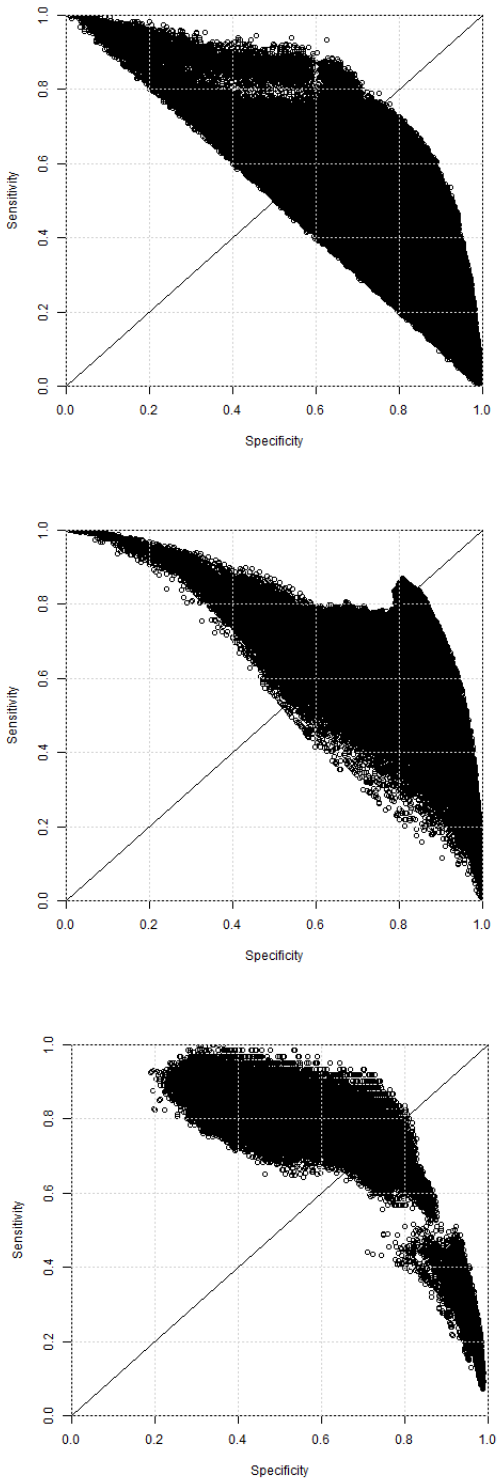

Figure 5: Sensitivity and specificity with flexible threshold for CART (top), RF (center), BO (bottom)

6 Conclusion

Nuclear power is fascinating, but not an easy technology, and attempts to generate nuclear electricity economically, i.e. competitive with other technologies, have been unsuccessful thus far. In particular, nuclear reactors feature low availability rates ($\sim$2/3), so that a better understanding of outage times is necessary. To work in this direction, this paper analyzes the link between unplanned outages and nuclear power plants' reactor technologies. We focus on two stylized technologies that have emerged after World War II, i.e. light-water reactors (LWR) and gas-cooled reactors (GCR). We hypothesize that outage profiles can be used to predict reactor technologies if outages are heterogeneous among reactor technologies. Based on a sample of 2534 reactor-year observations of light-water and gas-cooled reactors between 2003 and 2015, 15 different categories of unplanned outages are analyzed with three tree-based methods, namely classification trees, random forests and boosting. For calibration, a large number of simulations with varying parameter settings are carried out. Different sampling approaches and analyses of classification thresholds are carried out to overcome limitations of an unbalanced sample. To obtain optimal predictions, a strict criterion is derived to focus on the simultaneous prediction of both reactor technologies.

The paper provides two results, and suggests further research: On nuclear technologies, it suggests that there is indeed a structural difference between the two big types of reactors, as indicated by the different origins of unplanned outages. All three methods indicate that a prediction of a reactors' technology using only the outage profile in one year is possible with considerable accuracy. With fixed classification thresholds, our results indicate that well-calibrated trees predict the correct technology with more than 83% using boosting, but already less complex classification trees and random forests obtain accuracies exceeding 75%. A tuning of classification thresholds further increased these accuracies especially for random forests.

From a methodological perspective, our results also underline the importance of a precise calibration of the three used machine learning algorithms. The results of the calibration process show the large range of potential classifiers that can be obtained, but indicate also considerable interdependencies of different parameters. Therefore, our results emphasize that for optimal predictions the parameter space needs to be explored carefully, interdependencies of parameters needs to be considered, and the structure of the data needs to be taken into account. Using a clear criterion to judge best predictions thereby helps to evaluate parameter settings and can give a clear guideline for calibration of methodologies.

References

Breiman, L. (1998). Arcing classifier (with discussion and a rejoinder by the author). *Annals of Statistics*, 26:801–849.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press, Boca Raton.

Davis, L. W. and Wolfram, C. (2012). Deregulation, consolidation, and efficiency: Evidence from US nuclear power. *American Economic Journal: Applied Economics*, 4(4):194–225.

Deutch, J., Moniz, E., Ansolabehere, S., Driscoll, M., Gray, P., Holdren, J., Joskow, P., Lester, R., and Todreas, N. (2003). The future of nuclear power. *an MIT Interdisciplinary Study, http://web. mit. edu/nuclearpower*.

Deutch, J. M., Forsberg, C. W., Kadak, A. C., Kazimi, M. S., Moniz, E. J., and Parsons, J. E. (2009). Update of the MIT 2003 future of nuclear power. *Cambridge, Mass.: Report for Massachusetts Institute of Technology. Retrieved September*, 17:2009.

Freund, Y. and Schapire, R. E. (1997). A desicion-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139.

Friedman, J., Hastie, T., and Tibshirani, R. (2009). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2nd edition edition.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 29(5):1189–1232.

Grömping, U. (2012). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*.

Hausman, C. (2014). Corporate incentives and nuclear safety. *American Economic Journal: Economic Policy*, 6(3):178–206.

Hirschhausen, v. C. (2017). Nuclear power in the twenty-first century: An assessment (Part I). DIW Berlin Discussion Paper 1700, DIW Berlin, Berlin.

Hu, C., Youn, B. D., Wang, P., and Yoon, J. T. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering & System Safety*, 103:120 – 135.

IAEA (2013). *PRIS-Statistics: Power Reactor Information System Statistical Reports*. International Atomic Energy Agency, Vienna, computer manual series no. 22 edition.

IAEA (2016). *Operating Experience with Nuclear Power Stations in Member States in 2015*. International Atomic Energy Agency, Vienna.

Joskow, P. L. and Parsons, J. E. (2009). The economic future of nuclear power. *Daedalus*, 138(4):45–59.

Lemmens, A. and Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2):276–286.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, 2(3):18–22.

Linares, P. and Conchado, A. (2013). The economics of new nuclear power plants in liberalized electricity markets. *Energy Economics*, 40, Supplement 1:S119 – S125. Supplement Issue: Fifth Atlantic Workshop in Energy and Environmental Economics.

Liu, Y., Chawla, N. V., Harper, M. P., Shriberg, E., and Stolcke, A. (2006). A study in machine learning from imbalanced data for sentence boundary detection in speech. *Computer Speech & Language*, 20(4):468–494.

Mayr, A., Binder, H., Gefeller, O., and Schmid, M. (2014a). The evolution of boosting algorithms. *Methods of Information in Medicine*, 53(6):419–427.

Mayr, A., Binder, H., Gefeller, O., Schmid, M., et al. (2014b). Extending statistical boosting. *Methods of Information in Medicine*, 53(6):428–435.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ruiz-Gazen, A. and Villa, N. (2007). Storms prediction: logistic regression vs random forest for unbalanced data. *Case Studies in Business, Industry and Government Statistics*, 1(2):91–101.

Segal, M. R., Barbour, J. D., and Grant, R. M. (2004). Relating hiv-1 sequence variation to replication capacity via trees and forests. *Statistical Applications in Genetics and Molecular Biology*, 3(1):1–18.

Sharma, N., Sharma, P., Irwin, D., and Shenoy, P. (2011). Predicting solar generation from weather forecasts using machine learning. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pages 528–533. IEEE.

Steck, H. (2007). Hinge rank loss and the area under the roc curve. In *European Conference on Machine Learning*, pages 347–358. Springer.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC bioinformatics*, 8(1):1.

Therneau, T., Atkinson, B., and Ripley, B. (2015). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-10.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., and Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105:569 – 582.

Wang, Z. (2011). HingeBoost: ROC-Based Boost for Classification and Variable Selection. *The International Journal of Biostatistics*, 7(1):1–30.

Wang, Z. (2013). *bst: Gradient Boosting*. R package version 0.3-3.

Zhang, F. (2007). Does electricity restructuring work? Evidence from the U.S. nuclear energy industry. *Journal of Industrial Economics*, 55(3):397–418.

Zhu, J., Zou, H., Rosset, S., and Hastie, T. (2009). Multi-class adaboost. *Statistics and its Interface*, 2(3):349–360.

7 Appendix

7.1 Descriptive statistics

Country	Year	GCR	PWR	BWR	Avg. avail.	Total capacity	Reactor age (avg.)	Planned outages	Unplanned outages
CHE	2003	0	3	2	0.92	3333	28.20	520.20	20.40
CHE	2004	0	3	2	0.91	3333	29.20	658.60	39.20
CHE	2005	0	3	2	0.84	3333	30.20	615.00	751.20
CHE	2006	0	3	2	0.93	3333	31.20	526.40	0.00
CHE	2007	0	3	2	0.93	3333	32.20	506.80	22.00
CHE	2008	0	3	2	0.93	3333	33.20	573.80	4.20
CHE	2009	0	3	2	0.92	3333	34.20	626.00	20.00
CHE	2010	0	3	2	0.89	3333	35.20	887.60	19.20
CHE	2011	0	3	2	0.88	3333	36.20	939.40	33.40
CHE	2012	0	3	2	0.86	3333	37.20	781.40	353.60
CHE	2013	0	3	2	0.88	3333	38.20	762.40	190.40
CHE	2014	0	3	2	0.92	3333	39.20	545.00	89.00
CHE	2015	0	3	2	0.69	3333	40.20	1678.80	969.20
ESP	2003	0	7	2	0.89	7708	21.89	593.11	164.22
ESP	2004	0	7	2	0.92	7708	22.89	353.33	175.78
ESP	2005	0	7	2	0.84	7708	23.89	672.78	462.22
ESP	2006	0	7	2	0.89	7708	24.89	412.33	327.89
ESP	2007	0	6	2	0.82	7567	24.25	786.50	543.75
ESP	2008	0	6	2	0.87	7567	25.25	390.75	550.50
ESP	2009	0	6	2	0.78	7567	26.25	1238.50	437.75
ESP	2010	0	6	2	0.90	7567	27.25	421.00	230.88
ESP	2011	0	6	2	0.84	7567	28.25	873.25	275.38
ESP	2012	0	6	2	0.89	7567	29.25	656.88	128.12
ESP	2013	0	6	2	0.87	7567	30.25	1140.12	231.50
ESP	2014	0	6	2	0.88	7567	31.25	692.71	214.86
ESP	2015	0	6	1	0.88	7121	30.57	775.43	73.29
FRA	2003	0	58	0	0.79	63260	18.63	1044.52	353.25
FRA	2004	0	58	0	0.81	63260	19.63	965.86	388.12
FRA	2005	0	58	0	0.81	63260	20.63	937.27	218.12
FRA	2006	0	58	0	0.81	63260	21.63	915.88	346.34
FRA	2007	0	58	0	0.78	63260	22.63	1049.02	612.98
FRA	2008	0	58	0	0.78	63260	23.63	961.27	701.24
FRA	2009	0	58	0	0.72	63260	24.63	1043.42	1082.76
FRA	2010	0	58	0	0.75	63260	25.63	959.80	806.31
FRA	2011	0	58	0	0.79	63130	26.43	1087.93	565.98
FRA	2012	0	58	0	0.76	63130	27.43	920.50	758.31
FRA	2013	0	58	0	0.75	63130	28.43	1118.12	757.98
FRA	2014	0	58	0	0.79	63130	29.43	1240.00	351.48
FRA	2015	0	58	0	0.78	63130	30.43	1245.88	366.86
GER	2006	0	11	6	0.89	20496	24.41	664.12	118.06
GER	2007	0	11	6	0.75	20496	25.41	1519.76	577.76
GER	2008	0	11	6	0.78	20496	26.41	581.24	1177.41
GER	2009	0	11	6	0.72	20496	27.41	1320.88	1101.18
GER	2010	0	11	6	0.75	20496	28.41	980.41	1086.00
GER	2011	0	11	6	0.79	20496	29.41	466.47	1917.24
GER	2012	0	7	2	0.90	12074	26.67	609.00	187.44
GER	2013	0	7	2	0.89	12074	27.67	673.11	270.11
GER	2014	0	7	2	0.89	12074	28.67	583.33	285.44
GER	2015	0	7	2	0.89	12074	29.67	668.38	91.25

Country	Year	GCR	PWR	BWR	Avg. avail.	Total capacity	Reactor age (avg.)	Planned outages	Unplanned outages
JAP	2003	0	23	29	0.59	44283	19.46	2046.33	1641.69
JAP	2004	0	23	30	0.69	45460	19.98	1518.19	1212.88
JAP	2005	0	23	32	0.68	47635	20.23	2132.29	695.82
JAP	2006	0	23	32	0.69	47635	21.23	2025.58	651.49
JAP	2007	0	23	32	0.63	47635	22.23	2120.07	734.67
JAP	2008	0	23	32	0.58	47635	23.23	3308.13	286.65
JAP	2009	0	24	32	0.63	48501	23.81	2505.35	338.89
JAP	2010	0	24	30	0.68	47180	24.47	2162.11	588.08
UK	2004	26	1	0	0.70	11359	28.78	1655.75	490.50
UK	2006	22	1	0	0.67	11167	28.00	1321.61	1011.00
UK	2007	18	1	0	0.57	10297	26.37	1344.42	1596.53
UK	2008	18	1	0	0.52	10297	27.37	1458.71	2390.94
UK	2009	18	1	0	0.69	10297	28.37	1163.26	772.68
UK	2010	18	1	0	0.66	10297	29.37	602.63	1702.74
UK	2011	18	1	0	0.70	10297	30.37	906.37	998.05
UK	2012	17	1	0	0.76	10080	30.67	682.22	754.56
UK	2013	15	1	0	0.79	9373	30.12	686.88	643.75
UK	2014	15	1	0	0.69	9373	31.12	1074.25	1090.19
UK	2015	14	1	0	0.76	8883	31.33	1095.13	291.20
USA	2006	0	69	35	0.91	103366	26.66	661.16	118.06
USA	2007	0	69	35	0.92	103366	27.66	553.83	121.07
USA	2008	0	69	35	0.91	103366	28.66	639.47	116.34
USA	2009	0	69	35	0.90	103366	29.66	648.77	213.92
USA	2010	0	69	35	0.91	103366	30.66	642.35	116.06
USA	2011	0	69	35	0.89	103366	31.66	851.17	90.65
USA	2012	0	69	35	0.86	103366	32.66	1004.80	225.44
USA	2013	0	69	35	0.88	103366	33.66	624.98	268.48
USA	2014	0	65	35	0.92	99790	34.65	616.38	86.52
USA	2015	0	65	34	0.92	99185	35.58	587.83	114.69

Table 7: Annual country-specific descriptive statistics

Outage Category	Plant systems affected
Reactor and Accessories	Reactor vessel and main shielding (including penetrations and nozzles), reactor core (including fuel assemblies), reactor internals (including steam separators/dryers - BWR, graphite, pressure tubes), auxiliary shielding and heat insulation, moderator and auxiliaries (PHWR), annulus gas system (PHWR/RBMK), none of the above systems
Reactor I&C Systems	Control and safety rods (including drives and special power supply), Neutron monitoring (in-core and ex-core), reactor instrumentation (except neutron), reactor control system, Reactor protection system, Process computer, Reactor recirculation control (BWR)
Reactor Auxiliary Systems	Primary coolant treatment and clean-up system, Chemical and volume control system, Residual heat removal system (including heat exchangers), Component cooling system, Gaseous, liquid and solid radwaste treatment systems, Nuclear building ventilation and containment inerting system, Nuclear equipment venting and drainage system (including room floor drainage), Borated or refuelling water storage system, CO2 injection and storage system (GCR), Sodium heating system (FBR), Primary pump oil system (including RCP or make-up pump oil), D2O leakage collection and dryer system (PHWR), Essential auxiliary systems (GCR)
Safety Systems	Emergency core cooling systems (including accumulators and core spray system), High pressure safety injection and emergency poisoning system, Auxiliary and emergency feedwater system, Containment spray system (active), Containment pressure suppression system (passive), Containment isolation system (isolation valves, doors, locks and penetrations), Containment structures, Fire protection system, None of the above systems
Reactor Cooling Systems	Reactor coolant pumps/blowers and drives, Reactor coolant piping (including associated valves), Reactor coolant safety and relief valves (including relief tank), Reactor coolant pressure control system, Main steam piping and isolation valves (BWR)
Steam Generation Systems	Steam generator (PWR), boiler (PHWR, AGR), steam drum vessel (RBMK, BWR), Steam generator blowdown system, Steam drum level control system (RBMK, BWR)
Safety I&C Systems (excluding reactor I&C)	Engineered safeguard feature actuation system, Fire detection system, Containment isolation function, Main steam/feedwater isolation function, Main steam pressure emergency control system (turbine bypass and steam dump valve control), Failed fuel detection system (DN monitoring system for PHWR), RCS integrity monitoring system (RBMK)

Outage Category	Plant systems affected
Fuel Handling and Storage Facilities	On-power refuelling machine, Fuel transfer system, Storage facilities, including treatment plant and final loading and cask handling facilities
Turbine and auxiliaries	Turbine, Moisture separator and reheater, Turbine control valves and stop valves, Main condenser (including vacuum system), Turbine by-pass valves, Turbine auxiliaries (lubricating oil, gland steam, steam extraction), Turbine control and protection system
Feedwater and Main Steam System	Main steam piping and valves, Main steam safety and relief valves, Feedwater system (including feedwater tank, piping, pumps and heaters), Condensate system (including condensate pumps, piping and heaters), Condensate treatment system
Circulating Water System	Circulating water system (pumps and piping/ducts excluding heat sink system), Cooling towers / heat sink system, Emergency ultimate heat sink system
Miscellaneous Systems	Compressed air (essential and non-essential / high-pressure and lowpressure), Gas storage, supply and cleanup systems (nitrogen, hydrogen, carbon dioxide etc.), Service water / process water supply system (including water treatment), Demineralized water supply system (including water treatment), Auxiliary steam supply system (including boilers and pressure control equipment), Non-nuclear area ventilation (including main control room), Chilled water supply system, Chemical additive injection and makeup systems, Non-nuclear equipment venting and drainage system, Communication system
All other I&C Systems	Plant process monitoring systems (excluding process computer), Leak monitoring systems, Alarm annunciation system, Plant radiation monitoring system, Plant process control systems, None of the above systems
Main Generator Systems	Generator and exciter (including generator output breaker), Sealing oil system, Rotor cooling gas system, Stator cooling water system, Main generator control and protection system
Electrical Power Supply Systems	Main transformers, Unit self-consumption transformers (station, auxiliary, house reserve etc.), Vital AC and DC plant power supply systems (medium and low voltage), Non-vital AC plant power supply system (medium and low voltage), Emergency power generation system (e.g. emergency diesel generator and auxiliaries), Power supply system logics (including load shed logic, emergency bus transfer logic, load sequencer logic, breaker trip logic etc.), Plant switchyard equipment

Table 8: Outage category definition