The Use of IoT Sensor Data to Dynamically Assess Maintenance Risk in Service Contracts *

Stijn Loeys† [1], Robert N Boute[1,2,3], and Katrien Antonio[1, 4]

[1]Faculty of Economics and Business, KU Leuven

[2]Technology and Operations Management, Vlerick Business School

[3]Flanders Make@KU Leuven

[4]Faculty of Economics and Business, University of Amsterdam, The Netherlands

Abstract

We explore the value of using operational sensor data to improve the risk assessment of service contracts that cover all maintenance-related costs during a fixed period. An initial estimate of the contract risk is determined by predicting the maintenance costs via a gradient-boosting machine based on the machine's and contract's characteristics observable at the onset of the contract period. We then periodically update this risk assessment based on operational sensor data observed throughout the contract period. These sensor data reveal operational machine usage that drives the maintenance risk. We validate our approach on a portfolio of about 4,000 full-service contracts of industrial equipment and show how dynamic sensor data improves risk differentiation.

Keywords: Maintenance, Service Contracts, Predictive Models, Condition Monitoring, Usage-Based Risk Assessment

1. Introduction

In the last two decades, many original equipment manufacturers (OEMs) shifted their focus from traditional product sales to maintenance and after-sales service activities (Neely, 2008). In doing

*To appear in the *European Journal of Operational Research*

†corresponding author, e-mail: stijn.loeys@kuleuven.be

so, OEMs take on the responsibility of maintaining and, in some cases, guaranteeing their products' uptime. This attention shift can be attributed to various factors, including higher expected profits, customer intimacy, and an attempt to avoid the so-called commodity trap. Counterintuitively, this pursuit of servitization may sometimes lead to lower, instead of higher, profits when the proper infrastructure is not in place. An OEM can only thrive in the after-sales market when it deliberately allocates resources to establish the required infrastructures to provide such services.

Numerous avenues are accessible to deliver maintenance services in the after-sales market. In one extreme form, users no longer acquire the equipment. Instead, they only buy the use of it. This concept is known as power by the hour. Smith (2013) discusses a case study where Rolls-Royce provided all maintenance and logistical support for supplied engines to the US Navy and received a fixed price for each hour the engines were operating. Under a less extreme scenario, customers retain equipment ownership but opt for a full-service contract covering all maintenance costs. The service provider is then responsible for all maintenance-related services during the contract.

In a full-service contract, the customer pays a contract fee at the start of each contract period while the service provider bears all maintenance risk. Customers with high maintenance risks incur higher maintenance costs on average. At the start of a contract period, it is vital to assess this maintenance risk, and thus the expected costs, as accurately as possible. The expected costs may fluctuate based on numerous factors, including the machine's physical attributes and running conditions. Charging the same maintenance cost to a portfolio of heterogeneous maintenance contracts that equals the average costs incurred over all machines and all contract periods does not account for the maintenance risk heterogeneity in the portfolio. This may lead to adverse selection where maintenance-light customers opt out of the contract when overcharged and, thus, pay for the maintenance-heavy customers. Adverse selection can be avoided by differentiating the tariff plan based on the customer's maintenance risk profile.

The phenomenon of adverse selection is well known in non-life insurance (NLI), which faces a similar problem. The insurance provider must set contract prices in line with the customer's risk profile to cover the stochastic claim costs incurred during a specified coverage period. In traditional NLI, it is common to assume the number (i.e., frequency) and the magnitude (severity) of the insurance claims as statistically independent (Klugman et al., 2012).[1] Differentiation of the contract fee then relies on predictive models that estimate the frequency and severity of

[1]Without this assumption, the distribution of the total incurred costs would be a mixture of a degenerate point mass at zero and a positive continuous distribution.

the claims based on the customer's features, such as the policyholders' age or vehicle power. Various predictive models exist that use these risk factors to build a tariff structure. The most prominent in the insurance industry are generalized linear models (GLMs, Nelder and Wedderburn, 1972) and generalized additive models (GAMs, Hastie and Tibshirani, 1990), because of their relatively high explainability together with their predictive power. Machine learning models to price insurance contracts have recently gained traction in the literature. Henckaerts et al. (2021) suggest appropriate loss functions for frequency and severity data to be used in tree-based methods such as random forests (Breiman, 2001) and gradient-boosting machines (Friedman, 2001). Noll et al. (2020) compare GLMs, gradient-boosting machines (GBMs), and neural networks (NNs) on a third-party liability dataset containing univariate risk factors. They find that the GBM and NN outperform the GLM regarding predictive accuracy of the claim frequency.

These risk factors, however, are only proxies for the actual risks related to a policyholder. In motor insurance, for instance, two drivers with similar risk factors can have different driving habits and styles, translating into divergent actual risk profiles. The advent of the Internet of Things and big data allows capturing new information revealed during a coverage period via the integrated use of *tele*communications and infor*matics* (Husnjak et al., 2015). These data collected by telematics are dynamic risk factors and are only revealed during the contract. There are different ways to extract features from this telematics information. For instance, basic features related to driving habits (e.g., the mileage driven during a contract period, night time driving) can be derived and can be used to predict the number of claims (see, e.g., Baecke and Bocca, 2017; Ayuso et al., 2019; Verbelen et al., 2018). Alternatively, unsupervised classification techniques can cluster customers with similar driving profiles. Gao and Wüthrich (2018) use clustering techniques on velocity-acceleration heat maps that represent the policyholder's driving style. The results of this clustering based on driving styles can be included as additional features in the predictive models for claim frequencies or severities.

Within a usage-based (UB) risk assessment, a customer's risk is directly influenced by his actual usage and behavior, observed via telematics, rather than relying solely on risk factors known at the start of a contract. This approach allows the insurer to tailor a premium more closely to the observed risk during a period and to avoid so-called premium leakage (Tselentis et al., 2016). Premium leakage refers to the financial loss incurred by an insurer due to inaccurate pricing relative to the actual risk exposure of a customer. A UB pricing scheme uses the telematics data registered in a contract period to adjust the risk assessment of that same period. Henckaerts and Antonio

(2022) analyze a motor third-party liability portfolio that includes telematics-collected data on the driving habits and driving styles (e.g., braking, harsh acceleration or deceleration, time of day during which the driving occurred) of young drivers who agreed on the installation of a telematics box in their car. The authors propose multiplying the baseline price with a correction factor that depends on the telematics information collected about young drivers. As such, the baseline fee is dynamically adjusted based on the vehicle's actual use during the coverage period. This makes the proposed tariff structure truly usage-based.

The rapid development of computer-based monitoring technologies has also found its way into industrial maintenance applications. In recent years, OEMs have collected data on the running conditions of their installed machinery through sensors. This has led to high-dimensional datasets where each dimension pertains to an operational condition of a machine. However, extracting a set of meaningful health indicators is not trivial. Machinery generally exhibits a long-term degradation process from a healthy to a failed state, and preventive maintenance is usually undertaken before actual failure. This means that, in general, there is abundant data on healthy machines yet little data on equipment that has run to failure. A whole stream of literature exists on extracting meaningful low-dimensional health indicators from captured high-dimensional sensor data (see, for example, Lei et al. (2018) for an overview). In this field, neural networks (of many different architectures) are recently gaining traction; see, e.g., de Pater and Mitici (2023) and Liu and Gryllias (2020). These extracted low-dimensional health indicators indicate the level of machine degradation and are used to calibrate condition-based maintenance (CBM) policies. In doing so, preventive maintenance can be scheduled based on a machine's actual condition or health state rather than age or time-based heuristics. An example is Drent et al. (2023), who capture the degradation of a component in real-time. They use the condition data to characterize the optimal replacement policy using real-time learning based on Bayesian theory. We refer to Quatrini et al. (2020) for a comprehensive overview of the literature related to CBM.

In this paper, we do not use the sensor data to install a CBM policy. We use the operational sensor data observed during the contract to capture a customer's machine usage and assess its maintenance risk. The procedure we propose follows the principles of a UB risk assessment and works as follows. Similar to Deprez et al. (2023b), we first predict the baseline maintenance costs based on a set of risk factors known at the contract's start. However, instead of the generalized additive model (GAM) employed by Deprez et al. (2023b), we use a GBM to estimate these maintenance costs. GBMs offer superior predictive power compared to GAMs when high-dimensional

input data, interaction effects, and ample training data are present. Furthermore, GBMs require less manual input during model building compared to GAMs. Other tree-based learners exist, such as regression trees or random forests (for a summary, see Henckaerts et al. (2021)). Still, a GBM is currently considered the gold standard for predictive modeling with tabular datasets, even when compared with several deep learning architectures (Shwartz-Ziv and Armon, 2022). Next, we propose a novel method to incorporate operational sensor data observed during the contract period into the maintenance risk assessment of full-service contracts, leveraging the collected IoT data. This extends the usage-based pricing approach of Henckaerts and Antonio (2022) to our industrial maintenance setting of full-service contracts and makes it different from the approach used by Deprez et al. (2023b), which solely uses risk factors known at the contract's start. Their approach does not consider how the machine is used during a contract. Our results show that some operational sensor data are relevant to assess maintenance risk and allow for better risk differentiation. By incorporating these dynamic risk factors, the contract period's cost estimate is adjusted based on its effective machine usage and better reflects the customer's maintenance risk. This results in a tailored and more refined risk assessment for these service contracts compared to models relying only on risk factors known at the contract's start. We emphasize that adjusting the baseline risk estimate solely depends on the actual usage and not on the maintenance costs realized during the coverage period. Although these maintenance costs are linked to usage, incorporating them into the estimation of maintenance risk contradicts the core principles of an insurance contract, as it would no longer shield customers from adverse events.

This paper is organized as follows: Section 2 outlines our methodology to calculate and subsequently update the baseline risk estimate. In Section 3, we describe the dataset we use to calibrate our models. Section 4 and 5 describe the predictive models used to estimate the contract's baseline maintenance risk and the correction factor, respectively. Section 6 discusses how a maintenance service provider can use this refined risk assessment for identifying high-maintenance customers, fairer pricing, and recommending part replacements. Section 7 concludes.

2. Maintenance cost prediction of full-service contracts

A full-service contract for machine i starts at timestamp $t_{i,0}$ and has a fixed, finite duration during which the OEM covers all planned and unplanned maintenance costs on machine i. We divide the contract duration into multiple fixed-length periods and predict the expected maintenance

costs for each period. This is done in two stages. We first develop a baseline maintenance cost prediction for each contract period j based on the risk factors known at the contract's start at time $t_{i,0}$. We then update the cost prediction periodically based on sensor data measured during each period. Our models can easily be adapted to any period definition, but the period definition should be large enough as we need enough sensor data to update the price with a reasonable amount of confidence. If a contract's duration is a non-integer number of periods, the length of the final period is set to the remainder of the duration.

The maintenance costs incurred in a contract period include the costs of all planned and unplanned maintenance visits. The planned maintenance visits are scheduled at the contract's start and are known as preventive maintenance (PM) visits. Multiple types of PM visits exist, focusing on different repairs or replacements. The sequence of these visits depends upon the projected machine usage. A machine can also require an unexpected maintenance visit during the contract. We distinguish between a minor and a more costly major maintenance visit. A minor maintenance visit is the replacement or repair of non-critical components at, or just before, their point of failure. We denote them as minor corrective visits. Major maintenance visits replace core components and are more expensive. We refer to them as overhaul visits. In our dataset, some contracts cover these overhauls, while others do not.

Define the Number (frequency) of corrective and overhaul visits for machine i in observation period j as $N_{i,j,c}$ and $N_{i,j,o}$, respectively, and let $S_{i,j,c,k}$ be the cost (Severity) of the kth corrective visit and $S_{i,j,o,l}$ the cost of the lth overhaul visit in period j. Furthermore, we define n_{pm} types of PM visits. Since these visits are scheduled at the contract's start, we know the number of visits per type p in period j for machine i, $n_{i,j,p}$, but the costs related to the kth preventive maintenance (PM) visit of type p in period j for machine i, $S_{i,j,p,k}$ are unknown. The total cost incurred during the observation period j for machine i, denoted $C_{i,j}$, is then given by the sum of the costs of all planned and unplanned maintenance visits:

$$C_{i,j} = \sum_{p=1}^{n_{pm}} \sum_{k=1}^{n_{i,j,p}} S_{i,j,p,k} + \sum_{k=1}^{N_{i,j,c}} S_{i,j,c,k} + \sum_{l=1}^{N_{i,j,o}} S_{i,j,o,l}. \tag{1}$$

In what follows, we describe how we develop a baseline maintenance cost prediction based on risk factors known at $t_{i,0}$. Next, we incorporate the operating conditions observed during the contract period j to adjust the baseline prediction for said contract period.

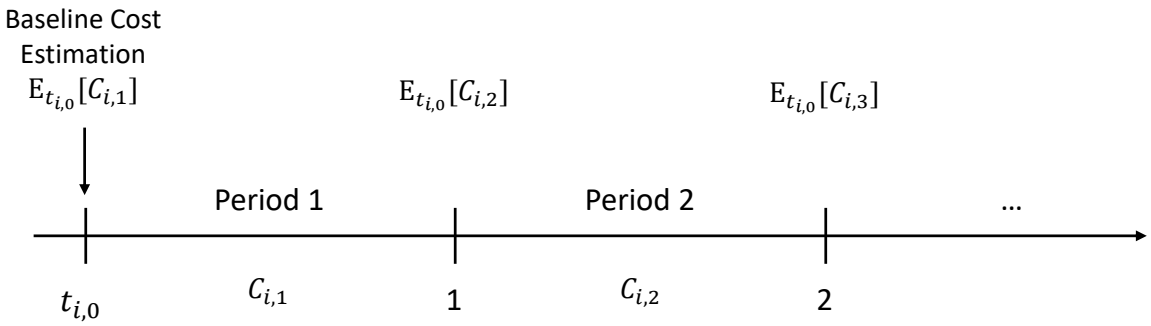


Figure 1: Machine i's baseline cost prediction for period j, $\mathbb{E}_{t_{i,0}}[C_{i,j}]$, uses risk factors known at the contract start.

Baseline cost prediction using risk factors known at the contract's start Figure 1 outlines our procedure to determine the baseline cost prediction based on information known at the contract start. To make our analysis tractable, we model the frequency and severity of the maintenance visits as statistically independent. This results in five components that need to be estimated: the frequencies of minor corrective visits and overhaul visits, and the severities of minor corrective visits, overhauls, and preventive maintenance visits. Period j's baseline cost estimation for machine i then evaluates to

$$\mathbb{E}_{t_{i,0}}[C_{i,j}] = \sum_{p=1}^{n_{pm}} \sum_{k=1}^{n_{i,j,p}} \mathbb{E}[S_{i,j,p,k}] + \mathbb{E}[N_{i,j,c}]\,\mathbb{E}[S_{i,j,c}] + \mathbb{E}[N_{i,j,o}]\,\mathbb{E}[S_{i,j,o}]\,. \tag{2}$$

Usage-based risk assessment based on sensor data We adjust the estimated baseline cost prediction for period j with $j > 0$, of a certain machine i based on observed operational data during period j. At the end of each period j, we have observed a set of operational conditions and events on machine i, denoted by $\mathbf{z}_{i,j}$. Each component of the baseline cost prediction $\mathbb{E}_{t_{i,0}}[C_{i,j}]$ (see Eq. (2)) is then multiplied with its respective correction factor $\delta(\mathbf{z}_{i,j})$. The resulting impact on the baseline cost prediction, $\mathbb{E}_{t_{i,0}}[C_{i,j}]$, is given by a global correction factor $\Delta(\mathbf{z}_{i,j})$, such that the updated usage-based cost prediction $\mathbb{E}_{t_{i,j}}[C_{i,j}]$ equals

$$\mathbb{E}_{t_{i,j}}[C_{i,j}] = \mathbb{E}_{t_{i,0}}[C_{i,j}] \cdot \Delta(\mathbf{z}_{i,j}). \tag{3}$$

This cost prediction adjustment attempts to reflect the change in estimated maintenance risk induced by the effective machine usage. For instance, should customers use their machine more extensively or contrary to the anticipated usage, their maintenance risk is adjusted via $\Delta(\mathbf{z}_{i,j}) > 1$.

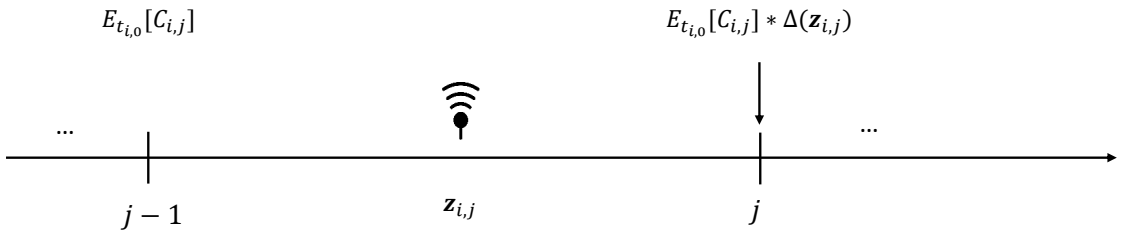


Figure 2: We adjust machine i's baseline cost prediction for period j, $\mathbb{E}_{t_{i,0}}[C_{i,j}]$, using the operational data $z_{i,j}$ observed in period j.

If the machine is well taken care of and usage is better than foreseen at the contract's start, their maintenance risk is reduced via $\Delta(\mathbf{z}_{i,j}) < 1$. This correction factor is evaluated at the end of every period j during the contract.

3. Dataset description

We validate our risk assessment approach with a dataset of an OEM who serves machines under a full-service maintenance contract. The dataset comprises about 4,000 contracts[2] observed over five years, located in multiple countries. We note that one machine can have multiple contracts over its lifespan and that one contract typically comprises multiple contract periods. We identified our risk factors and sensor data after ample discussions with the OEM, considering data availability and expert opinion. These discussions provided an initial set of risk factors and sensor data deemed most relevant. Appendix A describes the different pre-processing steps taken on this dataset.

Risk factors known at the contract's start Table 1 lists the features and feature groups available in our dataset for baseline cost estimation. On the one hand, we have risk factors that do not change over time. Examples are the machine type, regulation type, or power level. The calendar year refers to the year of the contract period's start. On the other hand, we have risk factors known at the contract initialization that can be different in each period. Examples include the preventive maintenance schedule or the expected running hours per period. We provide a full list of the feature groups in Table 1.

[2]The exact number is disguised for confidentiality reasons.

Feature Group	
Static	Time-varying
Machine category	Planned PM information
Machine power	PM history
Regulation type	Foreseen running hours
Country	Total registered and expected cumulative running hours
Service contract information	
Controller type	
Calendar year	

Table 1: Features available in the dataset for baseline cost estimation, categorized into static and time-varying feature groups. Static features do not change over contract periods, whereas time-varying features do. Some feature groups only include one feature.

Sensor data The dataset contains dynamic sensor data about the running conditions of a subset of machines. These data relate to the machine's internal or external operating conditions. Examples of internal data encompass but are not restricted to the machine's temperature, pressure differentials, or operating energy level. External data might consider factors such as ventilation levels or room temperature. Finding meaningful patterns in this raw, continuous data proved challenging. We use (discrete) events created by experts at the OEM, likely to be correlated with our target variables (i.e., the frequency and severity of visits). As such, we rely on their expertise to preprocess the continuous, raw data into meaningful low-dimensional indicators. If this is unavailable, one can rely on specialized literature to extract such health indicators from high-dimensional operational data. An example of an event is a specific temperature limit being exceeded multiple times in a period or the machine shutting down unexpectedly frequently.

The occurrence of such an event is immediately known to the customer and the OEM. An event is triggered when a pre-defined set of conditions is met and is revealed to the customer and the OEM in real time. However, this information is not immediately used to update the contract. Updating the risk assessment each time an event occurs would make the risk assessment too volatile. Furthermore, the fact that no events happened over a certain period is information in itself. The collected event data is used periodically at the end of each contract period for a given machine.

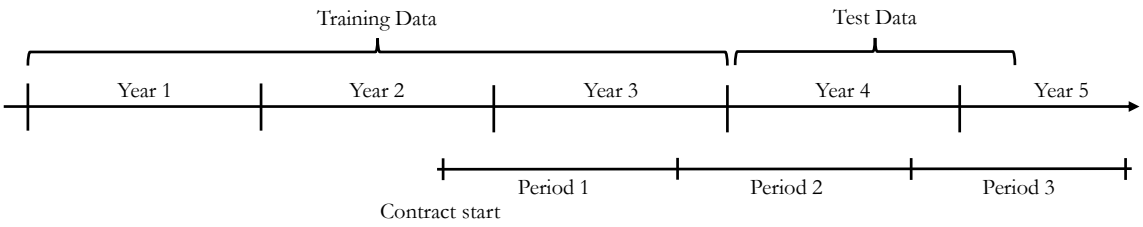


Figure 3: All contract periods starting before calendar year 4 are contained in the training set. The test set contains the contract periods starting after year 4. The figure illustrates a contract where the first two contract periods are contained in the training set, and the third contract period is in the test set.

Train-test split We split our dataset into a training and test set. We adopt the splitting strategy proposed in Henckaerts and Antonio (2022). The training set is all the contract periods that start in the first three calendar years. Contract periods starting in the fourth or fifth calendar year are contained in the test set. This strategy mimics reality in that a new set of observations becomes available every year, and a model, calibrated on historical data, makes predictions for these new observations. Figure 3 provides an example. This machine's first two contract periods start before the end of year 4 and are contained in the training set. Period 3 starts in year 4 and is included in the test set. This period is right-censored since we do not consider information past mid-year 5.

4. Predictive models to estimate the baseline costs

To predict the maintenance costs using risk factors known at the contract's start (see Eq. (2)), our predictive models predict a target variable $y_{i,j}$ for a machine i and contract period j, with the target variable either the frequency of minor corrective visits or overhauls, or the severity of minor corrective visits, overhauls, and preventive visits. Two types of risk factors are at our disposal as predictor variables. The risk factors that do not change over time, such as the machine type, regulation type, or power level, are denoted $\mathbf{x}_i$. The risk factors known at the contract initialization that can be different in each period j, such as the preventive maintenance schedule or the expected running hours per period, are indicated by $\mathbf{x}_{ij}$.

4.1 Gradient-boosting machine

A GBM is an iterative statistical method that combines a sequence of weak learners, such as shallow decision trees (Friedman, 2001), into one powerful predictor. At each iteration, the pseudo-residuals are used to identify the observations for which the current predictor performs poorly. The pseudo-residual for observation i in each iteration is calculated as the negative gradient of the loss function evaluated at the current model fit. The boosting method learns slowly by fitting shallow decision trees of a certain depth to these pseudo-residuals. Per leaf node, the update is calculated as the constant that has to be added to the previous model fit to minimize the loss function. A shrinkage parameter controls the learning speed. Finally, the collection of trees is used to make predictions. Stochasticity is introduced in the learning process by randomly sampling a subset of all available columns and rows of the training data. We use the XGBoost implementation of Chen et al. (2015), which also features a regularization parameter. A higher value of this regularisation parameter prevents overfitting and results in simpler models.

All hyperparameters must be carefully selected in the GBM training process. To select best-performing values for these parameters, we perform a full grid search in combination with a 10-fold cross-validation[3] on the training set defined in Section 3. We account for different distributions of the target variable by modifying the loss function. Henckaerts et al. (2021); Wuthrich and Buser (2023) suggest using the Poisson and gamma deviance as the loss function for, respectively, the frequency and severity of claims. The contribution to the Poisson deviance, $D_{i,j}^{POI}$, of a period j on machine i is given by

$$D_{i,j}^{POI}(y_{i,j}, f_{freq}(\mathbf{x}_i, \mathbf{x}_{i,j})) = 2\left(y_{i,j} \cdot \log\left(\frac{y_{i,j}}{f_{freq}(\mathbf{x}_i, \mathbf{x}_{i,j}) \cdot \Delta t_{i,j}}\right) - (y_{i,j} - f_{freq}(\mathbf{x}_i, \mathbf{x}_{i,j}) \cdot \Delta t_{i,j})\right), \tag{4}$$

where $y_{i,j}$ is either the number of failure or overhaul visits and $f_{freq}(\mathbf{x}_i, \mathbf{x}_{i,j})$ is the predicted value. The predictions are scaled by the exposure term $\Delta t_{i,j}$. $\Delta t_{i,j}$ equals one if the observation considers a full period. Should a period be shorter, $\Delta t_{i,j}$ reflects the fraction of a maximum length period. The loss function is defined as the gamma deviance for the severity models. A contract period j on machine i contributes

$$D_{i,j}^{GAM}(y_{i,j}, f_{sev}(\mathbf{x}_i, \mathbf{x}_{i,j})) = 2\left(\frac{y_{i,j} - f_{sev}(\mathbf{x}_i, \mathbf{x}_{i,j})}{f_{sev}(\mathbf{x}_i, \mathbf{x}_{i,j})} - \log\left(\frac{y_{i,j}}{f_{sev}(\mathbf{x}_i, \mathbf{x}_{i,j})}\right)\right) \tag{5}$$

[3] A 10-fold cross-validation is widely adopted in the machine learning literature as a standard technique. Using a higher number of folds decreases the size of the validation set, leading to more variable results and potential overfitting. Decreasing the number of folds lowers the amount of training data, leading to worse-performing models.

Rank	Minor corrective visit frequency		Overhaul frequency		Minor corrective visit severity	
	Feature	%	Feature	%	Feature	%
1	PM history	18.4	Planned PM information	17.78	Country	22.78
2	Total cumulative registered running hours	16.54	PM history	17.16	PM history	17.22
3	Foreseen running hours in next period	15.73	Foreseen running hours in next period	16.81	Machine power	14.34
4	Machine power	15.63	Total cumulative registered running hours	16.39	Total cumulative registered running hours	14.1
5	Planned PM information	12.46	Machine power	13.6	Foreseen running hours in next period	12.76
$\sum$		78.76		81.74		81.2

Rank	Overhaul severity		PM severity	
	Feature	%	Feature	%
1	Machine power	22.94	PM history	16.31
2	PM history	17.17	Machine power	14.44
3	Country	14.37	Planned PM information	12.6
4	Total cumulative registered running hours	14.08	Country	10.57
5	Foreseen running hours in next period	13.43	Foreseen running hours in next period	10.27
$\sum$		81.99		66.19

Table 2: The 5 most important feature groups in the training data per component of the baseline cost estimation in Eq. (2) and their information share.

to the gamma deviance, where $y_{i,j}$ represents either the severity of a minor corrective visit, an overhaul, or a PM visit and $f_{sev}(\mathbf{x}_i, \mathbf{x}_{i,j})$ is the predicted value. To measure out-of-sample predictive performance, we use the Poisson and gamma deviance for, respectively, the frequency and severity of maintenance visits.

We calibrate a GBM per component of Eq. (2) and combine their predictions to arrive at a baseline cost prediction per contract period. Table 5 in Appendix B gives an overview of the parameters we tuned in combination with their optimal values for each model.

4.2 Insights in the GBMs

Table 2 lists the five most important feature groups per estimated GBM. Historical information about maintenance visits, the foreseen running hours in the next contract period, and the machine power are among the most important features in all GBMs. In all components of Eq (2), the top five feature groups carry around 80 % of the total information contained in the collection of 11 feature groups (see Table 2), except for the PM severity component, where about 66% is explained.

Deprez et al. (2023b) use GAMs to estimate the baseline price. A GAM assumes a distribution from the exponential family for the response variable $y_{i,j}$. It relates the expected value of this response variable via a link function $g(\cdot)$, e.g., a log function, to the predictor variables, i.e., the risk factors known at the contract's start. GAMs trade in predictive power for superior explainability compared to GBMs. In Appendix B, we briefly introduce GAMs and report detailed

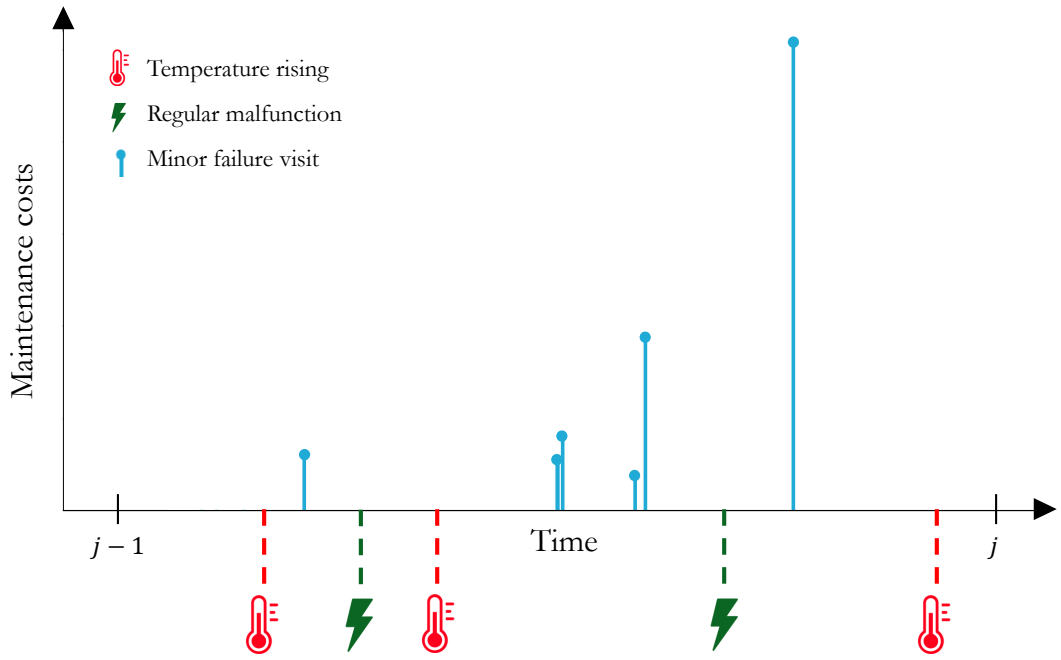


Figure 4: Example of a contract period where minor corrective visits (and their associated costs) and two types of operational events are registered, i.e., *regular malfunction* (twice) and *temperature rising* (three times).

results comparing our GBM models against GAM. Our GBMs outperform or equal their GAM counterparts in predicting the frequency of unplanned visits and the severity of corrective visits and preventive maintenance but not the severity of overhauls.

5. Predictive models for usage-based risk assessment

5.1 Data cleaning and pre-processing

We consider the actual machine usage to update the baseline risk assessment during the considered period. The OEM with whom we collaborate captures operational sensor data with a sampling frequency of a few minutes. Performing feature engineering on this big data without domain knowledge goes beyond the scope of this paper. Instead, we focus on different types of 'events' predefined by the OEM. These events delineate conditions that considerably impact the machine's health. Figure 4 visualizes the timeline of one specific contract period in our dataset, registering two types of operational events, one having been registered three times, and the other two times. The minor corrective visits and their related costs are also displayed.

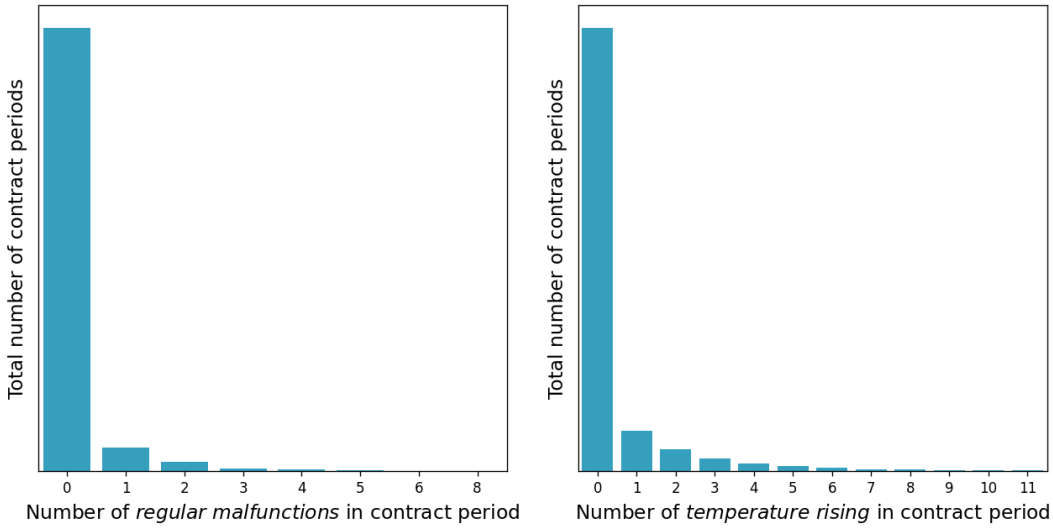
We made the following considerations to identify the events we use to adjust the baseline risk assessment. First, some events were only introduced in the last few years of the observation period (defining the test set). Therefore, we restricted the pool of possible events to those defined before the end of year 2, so that the event existed for at least one full observation period in the training set. Second, the event had to be active on the machine. An event is inactive when it is impossible on that specific machine type or when the related sensor data are not captured.

This reasoning leaves us with two events derived from sensor data: *regular malfunction* and *temperature rising*. These events indicate whether the machine exceeded a defined temperature limit multiple times in a period or the machine experienced frequent malfunctions, respectively. These events are registered on only one machine type, however. This results in a dataset of about 5,000 contract periods. Figures 5a and 5b show that both events occur rarely. Moreover, Figures 5c and 5d reveal that more occurrences of these events correspond with more corrective visits in a contract period. As the data indicates that most events rarely occur more than once or twice in a contract period, there is no pattern of occurrence (such as a uniform, increasing or decreasing pattern of occurrences). Should events with more occurrence be introduced, our correction factor could additionally consider a covariate capturing the concentration or "burstiness" (Jo and Hiraoka, 2023).

5.2 Model

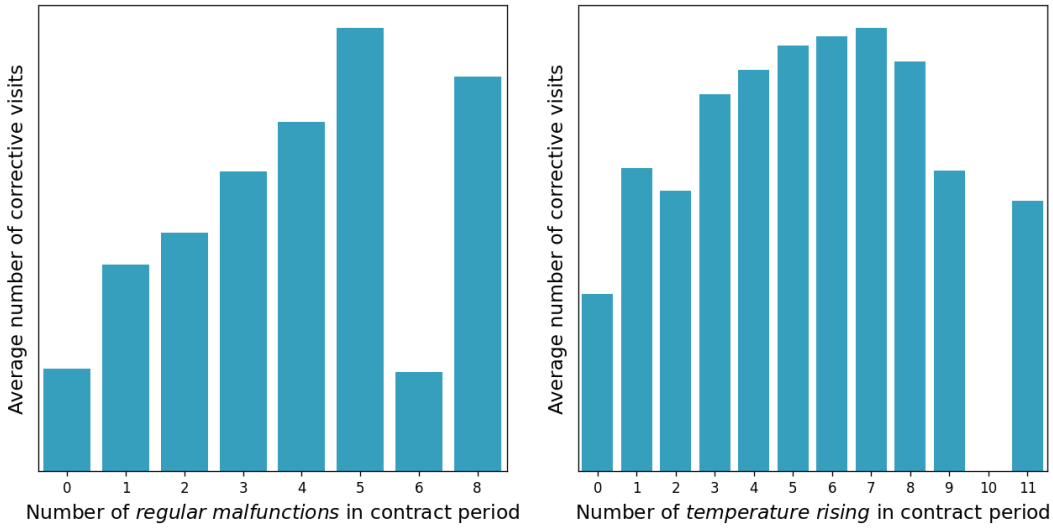
The usage-based cost prediction $\mathbb{E}_{t_{i,j}}[C_{i,j}]$ after observing period j is obtained by multiplying the maintenance visit's frequencies and severities in the baseline estimate $\mathbb{E}_{t_{i,0}}[C_{i,j}]$ of Eq. (2) with their respective correction factor $\delta(\mathbf{z}_{i,j})$. We calibrate the updating factor for each component of Eq. (2) via a generalized linear model (GLM, Nelder and Wedderburn, 1972). GLMs generalize linear models by allowing the response variable to follow any distribution from the exponential family. The Poisson and gamma distributions are part of this family. Furthermore, in a GLM the so-called link function maps the linear predictor to the expected value of the response variable. In our GLM specifications, we use a log-link and take the baseline prediction given by the GBM in Section 4.1 into account via an offset, a term with a coefficient fixed to one, in the linear predictor of the GLM. As such, at the level of the linear predictor, the GLM specification for a target variable $y_{i,j}$ is

$$\ln(\mathbb{E}[y_{i,j}|\mathbf{x}_i, \mathbf{x}_{i,j}, \mathbf{z}_{i,j}]) = \ln(f(\mathbf{x}_i, \mathbf{x}_{i,j})) + \beta_0 + \sum_k \beta_k z_{i,j,k} + \sum_k \sum_{l,k\neq l} \beta_{k,l} z_{i,j,k} z_{i,j,l}, \tag{6}$$



(a) Histogram of *regular malfunction*

(b) Histogram of *temperature rising*



(c) Average number of minor corrective visits per contract period as a function of the number of *regular malfunction* per contract period

(d) Average number of minor corrective visits per contract period as a function of the number of *temperature rising* per contract period

Figure 5: Figures (a) and (b) show the total number of contract periods for which a certain number of events have been registered. Figures (c) and (d) display the average number of corrective visits per contract period when a certain number of respectively *regular malfunction* or *temperature rising* have occurred. Both events positively correlate with the number of corrective visits in a contract period.

or, after exponentiating both sides, we obtain at the level of the target variable

$$\mathbb{E}\left[y_{i,j}|\mathbf{x}_i, \mathbf{x}_{i,j}, \mathbf{z}_{i,j}\right] = f(\mathbf{x}_i, \mathbf{x}_{i,j}) \underbrace{\times \exp(\beta_0) \times \prod_k \exp(\beta_k z_{i,j,k}) \times \prod_k \prod_{l,k\neq l} \exp(\beta_{k,l} z_{i,j,k} z_{i,j,l})}_{\delta(\mathbf{z}_{i,j})}. \quad (7)$$

The linear predictor is thus additive, and the prediction at the level of the target variable is multiplicative in the following factors:

- the baseline prediction, $f(\mathbf{x}_i, \mathbf{x}_{i,j})$, obtained from the GBM described in Section 4.1,
- a general updating factor, $\exp(\beta_0)$, when no events are registered,
- the contribution of each operational event, $\exp(\beta_k \cdot z_{i,j,k})$,
- the contribution of each one-way interaction between two events, $\exp(\beta_{k,l} \cdot z_{i,j,k} \cdot z_{i,j,l})$.

With a log-link function, the expected value of the response variable, i.e., the frequency or severity of the different types of maintenance visits, depends multiplicatively on the events derived from the operational sensor data, captured by $\delta(\mathbf{z}_{i,j})$ in (7). This correction factor adjusts the baseline prediction $f(\mathbf{x}_i, \mathbf{x}_{i,j})$ obtained with the GBM.

The calibration of a GLM with canonical link function and maximum likelihood estimation (MLE) satisfies the balance property, which implies that the sum of the fitted values equals the sum of the observations on the training data (Wüthrich, 2019, Corollary 2.4). The total predicted cost of the portfolio remains unchanged after considering the dynamic risk sensor data. Our proposed UB methodology results in a pure redistribution of estimated costs on the training set.

We apply the Least Absolute Shrinkage and Selection Operator or LASSO (Tibshirani, 1996) to perform an initial selection of features and their interactions to be included in the correction factor. We recognize that employing a LASSO regression may not be necessary in this context, given the small number of events derived from the gathered sensor data. We use it, however, to show how our framework can be used even when the number of available features increases. LASSO shrinks the model coefficients $\boldsymbol{\beta}$ towards 0 by introducing a regularization penalty $\lambda\|\boldsymbol{\beta}\|_1$ in the likelihood function, thereby promoting sparsity and effectively setting some coefficients to zero. The degree of sparsity depends on the regularization parameter λ. Each variable $z_{i,j}$ is standardized before applying LASSO to allow for a fair trade-off between events.

To tune λ, we specify a grid of possible values. Next, we perform 10-fold cross-validation on the training set with each possible λ. The λ that, on average, yields the best prediction across the

	Event	Coefficient	s.e.	p-value
Intercept		-0.0711	0.02	< 0.001
TR	Temperature rising	0.0514	0.0118	< 0.001
RM	Regular malfunctions	0.1922	0.0269	< 0.001
TR:RM	Interaction between TR and RM	-0.0291	0.0101	0.004

Table 3: Estimated coefficients, standard error (s.e.), and p-value of events and their one-way interactions for minor corrective visit frequency component of Eq. (2)

holdout sets is selected. Next, we fit a GLM using the variables and interaction terms selected by the LASSO model when using the optimal λ. Finally, we remove all statistically non-significant terms, i.e., $p \geq 0.05$, from this GLM and re-calibrate the model using only the statistically significant terms.

5.3 Results

In what follows, we describe how the sensor information adjusts the prediction of the expected number of the frequency of minor corrective visits. The estimated correction factors of the usage-based model (see Eq. (7)) for the other components of the total cost prediction in Eq. (2) are obtained similarly. We find that only the frequency of minor corrective visits and the severity of preventive maintenance (PM) components have statistically significant effects on reassessing the maintenance risk. We describe the results in Appendix C.

First, the selection of features and interaction effects using LASSO with the optimal λ (obtained through 10-fold cross-validation) reveals that both the features *regular malfunction* and *temperature rising* and their interaction effect should be included in the GLM. Next, we fit a GLM that uses these variables. Table 3 presents their estimated coefficients, the corresponding standard error, and the p-value. According to the model fit, all retained variables allow for a statistically significant adjustment of the expected number of minor corrective visit frequency. The general updating factor or intercept, β_0, is estimated at -0.0711. This implies that when neither a *regular malfunction* nor a *temperature rising* is registered in a period, the correction factor $\delta(\mathbf{z}_{i,j})$ for the minor corrective visit frequency prediction is $\exp(-0.071) = 0.9315$. This downward adjustment signals that, ceteris paribus on the other components of Eq. (2), the initial maintenance risk was estimated too high when none of these events were registered. The coefficients of *regular malfunction* and *temperature rising* are estimated at 0.1922 and 0.0514. Both

events have a statistically significant upward effect on the expected frequency of minor corrective visits, with the effect of a *regular malfunction* around 4 times higher than that of a *temperature rising*. The estimated coefficient of the interaction term is -0.0291. The negative coefficient indicates that when both a *regular malfunction* and a *temperature rising* occur, their respective coefficients overestimate the effect on the minor corrective visit frequency.

The linear predictor being additive in its coefficients facilitates the interpretation of the estimated coefficients of our GLMs. This supports the explainability of the effect of sensor information on the perceived maintenance risk. Figure 6 provides a graphical method visualizing the different terms of the linear predictor (see Eq. (6)). This waterfall plot indicates how the frequency prediction of minor corrective visits is adjusted for a specific contract period when during the contract period, two *temperature rising* (TR) and three *regular malfunction* (RM) events were registered. For this contract period, the baseline estimate given by our GBM indicates an expected number of 0.93 minor corrective visits in that contract period. The waterfall in Figure 6 then outlines how the linear predictor is adjusted for this contract period: The logarithm of the baseline prediction $ln(0.93) = -0.072$ is updated with a general updating factor $\beta_0 = -0.07$ (Intercept), the occurrence of two temperature rising (TR) events resulting in a contribution of 0.1, three regular malfunctions (RM) with contribution 0.58, and their interaction effect of -0.17. This results in a total linear predictor for the minor corrective visit frequency of

$$ln(0.93) - 0.07 + 0.1 + 0.58 - 0.17 = 0.368.$$

The revised estimate of the minor corrective visit frequency based on this sensor information amounts to $\exp(0.368) = 1.45$ minor corrective visits for that period, an increase of 56% compared to the initial estimate provided by the GBM.

We analyze the distribution of the correction factors $\delta(\mathbf{z}_{i,j})$ to the different components of the baseline risk assessment in Eq. (2), together with the global correction factor to the entire baseline risk assessment, $\Delta(\mathbf{z}_{i,j})$. Figure 7 shows the distribution of these correction factors and the global correction factor. The distribution for both the minor corrective visit frequency and the PM severity contains a large point mass that represents about 80% of the observations where neither a *regular malfunction* nor a *temperature rising* is observed. This point mass is located at the exponent of the intercept, respectively at $\exp(-0.0711) = 0.9313$ and $\exp(-0.0054) = 0.9946$ for the minor corrective visit frequency and PM severity. The distribution of the global correction factor, $\Delta(\mathbf{z}_{i,j})$, does not have a point mass. This is because a contract period can have no, one, or multiple PM visits, each time resulting in a different global correction factor in the absence of

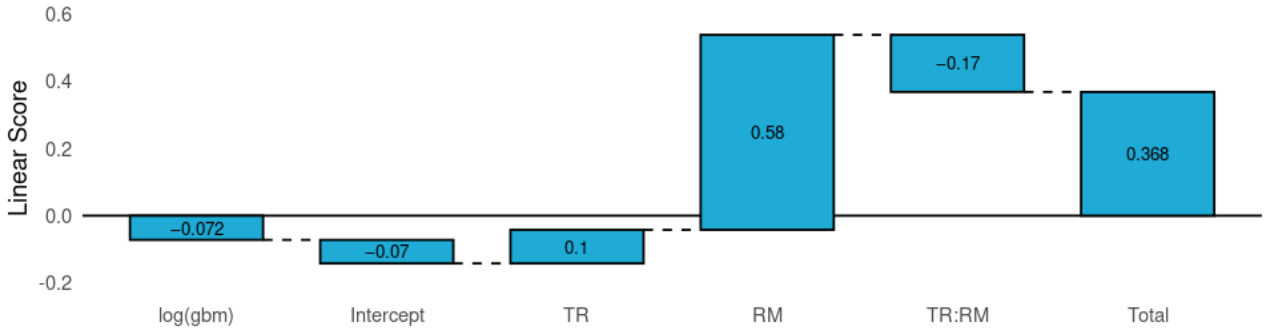


Figure 6: The waterfall plot visualizes the individual components of the linear predictor resulting in the final linear predictor (Total) for the minor corrective visit frequency of a specific contract period. In this contract period, two *temperature rising* (TR) and three *regular malfunction* (RM) are registered. Using the coefficients from table 3, the baseline estimate of $\exp(-0.072) = 0.93$ is adjusted upward to $\exp(0.368) = 1.45$, an increase of 56%.

event registrations.

We observe that for about 90% of the contract periods, the initial maintenance risk, and thus baseline cost estimate, is slightly over-estimated, while the baseline cost estimate is an underestimate of the risk for about 10% of the contract periods (indicating their actual maintenance risk is higher due to adverse operating conditions). This aligns with the fact that both *regular malfunction* and *temperature rising* are not observed in most periods, as shown in Figure 5, and that all the calibrated GLMs suggest either a non-significant intercept or a non-positive one. Furthermore, we note that the balance property of GLMs (Wüthrich, 2019, Corollary 2.4) guarantees that the positive adjustments completely cover the negative ones per component on the training set. If the distribution of the observations found in the test set is similar to those found in the training set, this property will still hold approximately.

6. Discussion

This paper proposes a methodology to predict the maintenance costs under full-service contracts based on a maintenance risk assessment. This risk assessment is initially based on risk factors known at the contract's start. As the actual maintenance risk also depends on the machine usage during the contract, our methodology captures these dynamic risk factors and adjusts the initial cost estimate periodically during the contract. If this dynamic information had been known at

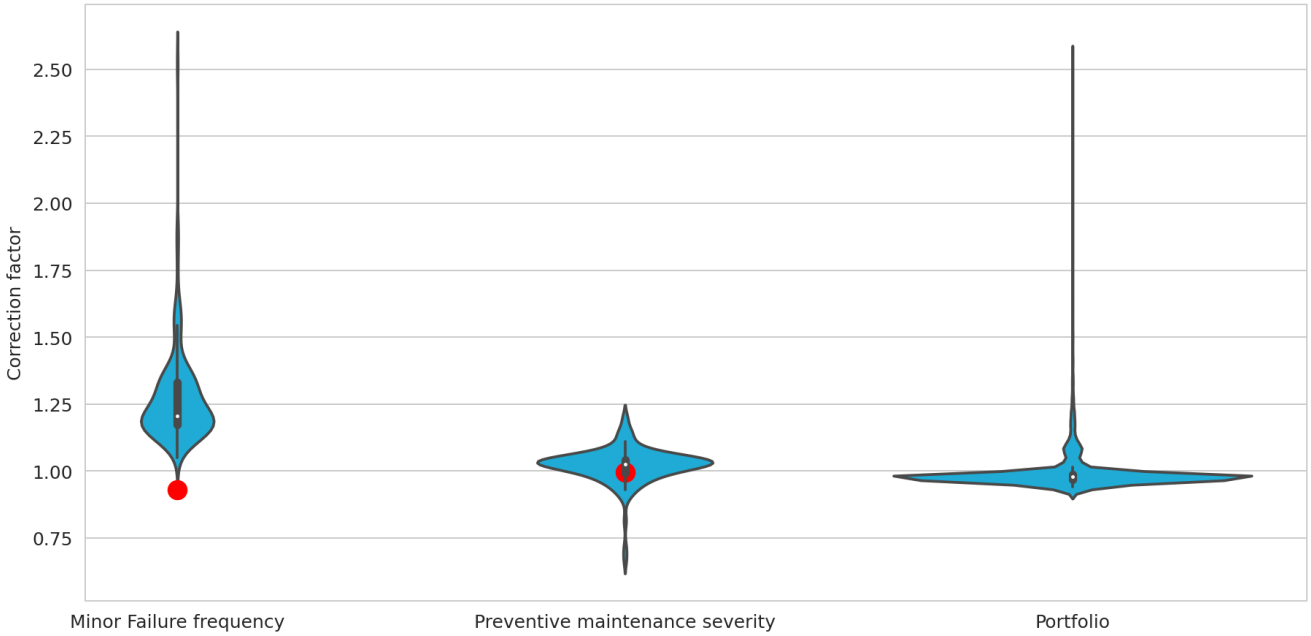


Figure 7: These violin plots show the distribution of the correction factors of the minor corrective visit frequency, the preventive maintenance severity, and the global correction factor $\Delta(\mathbf{z}_{i,j})$. The red dot indicates a point mass located at the correction factor applied when neither a *regular malfunction* nor a *temperature rising* event occurs.

the contract's start, it could have been incorporated into the initial risk assessment.

One of the benefits of updating the maintenance risk assessment throughout the contract's duration is that it allows maintenance service providers to identify customers who take good (or better) care of their machines. Holding customers accountable for their operational usage reduces the moral hazard of a full-service contract. Our proposed usage-based risk assessment thus incentivizes customers to care for their machines, indirectly reducing maintenance risks. An intuitive dashboard, similar to the waterfall plot shown in Figure 6, for example, reveals how each operational event impacts the risk assessment. The adjustments to machine i's baseline cost estimate for period j are quantified through the correction factor, $\Delta(\mathbf{z}_{i,j})$. When $\Delta(\mathbf{z}_{i,j}) > 1$, the initial cost estimate was too low to reflect the maintenance risk in period j. When $\Delta(\mathbf{z}_{i,j}) < 1$, the contract period costs were estimated too high with respect to the risk based on the machine usage. This correction factor makes it transparent to the customer how their risk is impacted. Customers with a high perceived maintenance risk profile could then identify the reasons and modify their behavior accordingly. This, in turn, will reduce the maintenance costs in the portfolio, making the service provider more competitive.

The contract provider can additionally consider adjusting the contract price through a penalty

or rebatement compared to the corresponding (past) contract period. We emphasize that such price adjustment is not based on the actual maintenance costs incurred during the said period but on the operational risk corresponding to machine usage. However, we caution against a pure price adjustment based on each contract period's correction factor. First, because the correction factor is unbounded, this can result in considerable penalties or rebatements. This offers customers little certainty and is undesirable from a commercial viewpoint, as certainty is one of the primary reasons for purchasing a full-service contract. Bounding the correction factor could give back some of this certainty. Second, the service contract provider could decide only to offer discounts when the operational data indicates a lower maintenance risk than estimated by the baseline models. Henckaerts and Antonio (2022) propose a methodology to do so. Instead of a monetary price adjustment, customers can be offered the opportunity to nullify the penalty if, for example, they install a better cooling element that reduces maintenance risk. This scenario is beneficial for both parties. The installed machine experiences enhanced uptime and, on average, fewer breakdowns, leading to lower maintenance costs for the contract provider and higher customer satisfaction.

Finally, the operational data can also be employed to assess the maintenance risk in future periods. For instance, the correction factor can be used to price future contract periods, assuming future machine usage will be similar to past periods. Alternatively, it can be used in the underwriting process to decide whether to offer a customer a full-service contract. Operational data exposes the customer's risk profile and provides additional information for customer selection. The service provider can refrain from offering a full-service contract to a customer if the operational data indicate that basic care is not given to the machine or it is placed in a damaging environment.

7. Conclusion

This paper discusses the maintenance risk estimation and price-setting of full-service contracts that cover all future preventive and corrective maintenance costs during a contract period. On a portfolio level, the contract fees should cover the expected maintenance costs to ensure the contract is not loss-making. Moreover, they should properly reflect the maintenance risk as some contracts are expected to incur higher costs than others based on their risk profile. We explore the benefits of using operational sensor data to improve this risk assessment.

We analyze a dataset of about 4,000 full-service maintenance contracts of industrial equipment. For all contracts in the portfolio, we first estimate the baseline costs for each contract period. This

is done by calibrating a gradient-boosting machine on risk factors known at the contract start, e.g., the machine type, oil type, or service history. In settings with high-dimensional input datasets, i.e., many covariates, GBMs can eliminate the need for a manual feature and interaction selection needed by more traditional statistical methods such as a generalized additive model. These risk factors allow an initial risk differentiation in the contract portfolio. However, the maintenance risk also depends on the machine usage. Two machines with similar risk factors can have divergent risk profiles when deployed in different operating conditions. We propose a usage-based risk assessment that updates the baseline cost estimation based on operational data captured by sensors using a generalized linear model. This correction factor reflects the operational risk factors registered throughout a period. In addition to using observable operational data, we can allow for additional heterogeneity among machines by incorporating a random or latent variable, similar to Drent et al. (2023). This is a topic for future research.

We identified two operational 'events' in our dataset each delineating a combination of machine conditions with a meaningful impact on the machine's health. The (lack of) registration of these events allows a better assessment of the frequency of minor corrective maintenance visits and the severity of preventive maintenance visits. This leads to an adjustment to the baseline cost estimation. In our dataset, we observe that the baseline cost was slightly overestimated for about 90% of the contract periods. In 10% of the contract periods, the initial cost estimate was adjusted upward to reflect the higher maintenance risk due to the occurrence of (at least one) of these events. The contract provider can decide whether and how this price adjustment is implemented, for instance, through a penalty or rebatement, either on the past period's contract price or on future periods' contract prices.

With the increased collection of operational sensor data, the LASSO procedure can identify more events to improve the risk assessment of full-service contracts. This may improve the competitiveness of the contract portfolio by attracting maintenance-light customers at lower prices and correctly charging maintenance-heavy customers without adversely impacting profitability.

Where this paper assumes a fixed preventive maintenance policy known at the contract's start, a potential extension of our model dynamically updates the PM policy based on reassessing the maintenance risk (Gilardoni et al., 2016). A higher-risk machine should receive more preventive maintenance. This will lower the expected costs of a contract, making it more competitive or lucrative. Vanderschueren et al. (2023) use causal inference to determine the effect of different PM strategies on the frequencies of corrective and overhaul visits. This approach estimates the

effect of different PM policies on total cost. Assessing the maintenance risk using causal learning with operational sensor data would be an interesting avenue to explore in future research.

Finally, we emphasize the scalability of our approach. Both a GBM and the LASSO procedure promote scalability. GBMs thrive with larger datasets and offer superior predictive power compared to GAMs when high-dimensional input data, interaction effects, and ample training data are present. Furthermore, GBMs require less manual input to build a model than GAMs, promoting scalability. The LASSO procedure introduced to identify the significant registered events also ensures scalability, offering automatic feature selection and regularization. LASSO reduces model complexity by shrinking non-relevant regression coefficients to exactly zero, allowing the service provider to identify the most relevant events. Thus, it ensures scalability when the number of registered events during a contract increases. Our methodology is also readily adaptable to similar contexts where the objective is to evaluate maintenance risks within a full-service contract with operational information collected as the contract progresses. An example is service contracts for cars, where service providers can access a high-dimensional set of initial risk factors at the contract's start. Operational telematics data affecting maintenance risks, such as heavy braking events, seasonal conditions, and cumulative mileage, is revealed over the contract's duration.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Ayuso, M., Guillen, M., and Nielsen, J. P. (2019). Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. *Transportation*, 46:735–752.

Baecke, P. and Bocca, L. (2017). The value of vehicle telematics data in insurance risk selection processes. *Decision Support Systems*, 98:69–79.

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

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4.

de Pater, I. and Mitici, M. (2023). Developing health indicators and rul prognostics for sys-

tems with few failure instances and varying operating conditions using a LSTM autoencoder. *Engineering Applications of Artificial Intelligence*, 117:105582.

Deprez, L., Antonio, K., Arts, J., and Boute, R. (2023a). Data-driven preventive maintenance for a heterogeneous machine portfolio. *Operations Research Letters*, 51(2):163–170.

Deprez, L., Antonio, K., and Boute, R. (2021). Pricing service maintenance contracts using predictive analytics. *European Journal of Operational Research*, 290(2):530–545.

Deprez, L., Antonio, K., and Boute, R. (2023b). Empirical risk assessment of maintenance costs under full-service contracts. *European Journal of Operational Research*, 304(2):476–493.

Drent, C., Drent, M., Arts, J., and Kapodistria, S. (2023). Real-time integrated learning and decision making for cumulative shock degradation. *Manufacturing & Service Operations Management*, 25(1):235–253.

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

Gao, G. and Wüthrich, M. V. (2018). Feature extraction from telematics car driving heatmaps. *European Actuarial Journal*, 8:383–406.

Gilardoni, G. L., de Toledo, M. L. G., Freitas, M. A., and Colosimo, E. A. (2016). Dynamics of an optimal maintenance policy for imperfect repair models. *European Journal of Operational Research*, 248(3):1104–1112.

Hastie, T. J. and Tibshirani, R. J. (1990). Generalized additive models, volume 43 of. *Monographs on statistics and applied probability*, 15.

Henckaerts, R. and Antonio, K. (2022). The added value of dynamically updating motor insurance prices with telematics collected driving behavior data. *Insurance: Mathematics and Economics*, 105:79–95.

Henckaerts, R., Côté, M.-P., Antonio, K., and Verbelen, R. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal*, 25(2):255–285.

Husnjak, S., Peraković, D., Forenbacher, I., and Mumdziev, M. (2015). Telematics system in usage based motor insurance. *Procedia Engineering*, 100:816–825.

Jo, H.-H. and Hiraoka, T. (2023). Bursty time series analysis for temporal networks. In *Temporal network theory*, pages 165–183. Springer.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss models: from data to decisions*, volume 715. John Wiley & Sons.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing*, 104:799–834.

Liu, C. and Gryllias, K. (2020). Unsupervised domain adaptation based remaining useful life prediction of rolling element bearings. In *PHM Society European Conference*, pages 10–10.

Neely, A. (2008). Exploring the financial consequences of the servitization of manufacturing. *Operations management research*, 1:103–118.

Nelder, J. A. and Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 135(3):370–384.

Noll, A., Salzmann, R., and Wüthrich, M. (2020). Case study: French motor third-party liability claims. Available at SSRN: 10.2139/ssrn.3164764.

Quatrini, E., Costantino, F., Di Gravio, G., and Patriarca, R. (2020). Condition-based maintenance—an extensive literature review. *Machines*, 8(2):31.

Shwartz-Ziv, R. and Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90.

Smith, D. J. (2013). Power-by-the-hour: the role of technology in reshaping business strategy at rolls-royce. *Technology analysis & strategic management*, 25(8):987–1007.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Tselentis, D. I., Yannis, G., and Vlahogianni, E. I. (2016). Innovative insurance schemes: pay as/how you drive. *Transportation Research Procedia*, 14:362–371.

Vanderschueren, T., Boute, R., Verdonck, T., Baesens, B., and Verbeke, W. (2023). Optimizing the preventive maintenance frequency with causal machine learning. *International Journal of Production Economics*, 258:108798.

Verbelen, R., Antonio, K., and Claeskens, G. (2018). Unravelling the predictive power of telematics data in car insurance pricing. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 67(5):1275–1304.

Wüthrich, M. V. (2019). From generalized linear models to neural networks, and back. Available at SSRN: 10.2139/ssrn.3491790.

Wuthrich, M. V. and Buser, C. (2023). Data analytics for non-life insurance pricing. *Swiss Finance Institute Research Paper*.

A. Data cleaning and pre-processing

A.1 Maintenance costs

To maintain cost estimations within a reasonable range, we establish, in collaboration with the OEM, both a lower and upper limit for the cost associated with each type of visit. The lower bound is set to a fixed amount. The cost of a PM visit was capped at four times the average cost of that type. Furthermore, all corrective and overhaul visit costs above the 95th percentile are deemed excessive and thus removed from the cost dataset.

As our dataset spans multiple periods, the inflation rate has to be accounted for. We employ the same strategy presented in Deprez et al. (2023a). First, the average cost per maintenance visit type in the reference year 2019 is calculated. Second, all costs related to a maintenance visit of a certain type are divided by the average cost of said type. As such, these costs can be compared without considering their visit type. Third, the average cost per year, $c_{y,calendar_yr}$ is calculated, and the logarithm is taken. Finally, we fit a weighted linear model on the tuples $(calendar_yr, \log(c_{y,calendar_yr}))$ with as weights the number of observations during each period. This procedure gives us an estimate of $r = 1.72$. All costs are scaled using this inflation rate.

A.2 Running hours

The last and arguably most important risk factor is the running hours (RHS) of a machine. Every few minutes, the cumulative amount of running hours since installation for a specific machine is logged. Due to the malfunctioning of sensors, this data can get corrupted. Two cleaning steps are undertaken. First, all decreases in the cumulative RHS are discarded. Second, if the RHS in a period is larger than the period itself, the increase is capped at the length of the period. We incorporate the expected RHS at the start of a period, and the expected RHS logged during the period as static covariates in the frequency models. The severity models are supplemented with the expected RHS at the start and, if it concerns a PM visit when the visit took place.

Since we only have access to the actual running hours, we assume they are equal to the expected running hours. However, in practice, this assumption does not hold. To address this, we can adopt an approach where we either gather or estimate the expected RHS per customer. These values can then be incorporated into our model, and an additional variable in the dynamic model can capture the discrepancies between predicted and actual RHS.

There are two primary reasons for not conducting this analysis. First, we do not have access to well-informed data about expected RHS per customer. Second, building a rudimentary model, such as linear regression, for RHS estimation based on historical data, is feasible only for older machines. Given that a significant proportion, approximately 50%, of our machines is relatively new, we would have to delve into more advanced statistical methods. But this is beyond our scope.

B. Comparison of a gradient-boosting machine and a generalized additive model

B.1 Generalized additive model

Consider period j on a machine i. In the GAM framework, the response variable, $y_{i,j}$, follows a distribution from the exponential family. We use the Poisson and gamma distributions, both part of the exponential family, to respectively model the frequency and severity of maintenance events. Both types of static risk factors, i.e. either variant, denoted by $X_{i,j}$ or invariant risk factors over different periods for a machine i, denoted by X_i, have categorical variables, respectively denoted by the vectors $\mathbf{x}_i^{cat}$ and $\mathbf{x}_{i,j}^{cat}$, and continuous variables, respectively denoted by the vectors $\mathbf{x}_i^{con}$ and $\mathbf{x}_{i,j}^{con}$. Furthermore, we apply smooth functions $f^{con}(\cdot)$ to capture non-linear effects of continuous

	Loss function	GBM	GAM	$-\%\Delta$
N_c	D_{POI}	**1.59**	1.78	10.67
N_o	D_{POI}	**0.23**	0.27	14.81
S_c	D_{GAM}	**0.78**	**0.78**	0
S_o	D_{GAM}	0.31	**0.30**	- 3.33
S_{pm}	D_{GAM}	**0.28**	0.32	12.5

Table 4: Poisson and gamma Deviance Statistic for GBM and GLM models. The best values are highlighted in bold.

risk factors. These smooth functions are usually represented using spline functions, which are piecewise-defined polynomials. A GAM then has the following specification:

$$\mathbb{E}\left[y_{i,j}|x_i^{cat}, x_i^{con}, x_{i,j}^{cat}, x_{i,j}^{con}\right] = g^{-1}\left(\beta_0 + \sum_k \beta_k x_{i,k}^{cat} + \sum_l \beta_l x_{i,j,l}^{cat} + \sum_p \beta_p f_p^{con}(x_{i,p}^{con}) + \sum_s \beta_s f_s^{con}(x_{i,j,s}^{con})\right).$$

We adopt the GAM calibration strategy presented in Deprez et al. (2021). To select the optimal model specification, we calibrate a GAM with all possible risk factors and remove the non-significant ones. Non-significant levels are merged with neighboring levels. To further improve their GAMs, interaction selection is done, limited to a single interaction term, based on the Akaike information criterion (AIC, Akaike (1974)). To measure out-of-sample performance, we use the Poisson and gamma deviance, defined in Section 4.1, for respectively the frequency and severity components.

B.2 Benchmark against GBM

Table 4 shows the out-of-sample Poisson and gamma deviance statistics obtained from the GBMs and GAMs for the frequency and severity components of Eq. (2). Table 5 summarizes the optimal tuning parameter values used for the GBM to generate these results. For both the frequency and the preventive maintenance severity model, the GBM outperforms the GAM by a significant margin. This is because ample data is available to calibrate the models, which benefits the GBM more. It is also an indication that higher-order interaction effects are present. Hunting for these effects in a GAM setting quickly becomes infeasible, even with limited available covariates.

The GAM does not outperform the GBM in predicting the frequency of corrective maintenance visits. On the overhaul severity model, the GAM outperforms the GBM slightly. It is a well-known fact in non-life insurance pricing that the estimation of severity is considerably more difficult than

	Distribution	ntrees	learn_rate	max_depth	sample_rate	col_sample_rate	lambda
N_c	Poisson	2300	0.01	7	1	0.5	90
N_o	Poisson	1700	0.01	7	1	0.5	50
S_c	gamma	300	0.01	3	1	0.25	20
S_o	gamma	1100	0.01	3	1	0.5	25
S_{pm}	gamma	1500	0.01	5	1	0.25	20

Table 5: Optimal settings of the XGBoost tuning parameters for the different predictive models

frequency modeling, with most literature focusing on the latter. This can be attributed to two main factors. First, maintenance cost structures inherently demonstrate more stochastic behavior. Second, we often only have a limited amount of observations available for model training. Indeed, our dataset only contains about 500 overhaul visits, of which 300 can be found in the test set. This severely limits the GBM's learning capability. A GAM is more robust to small data, as indicated by the results. Note that the severity of preventive maintenance does not adhere to both factors. The cost structure is less stochastic compared to their corrective and overhaul counterparts since in advance it is more or less defined what will be done during a preventive maintenance visit of a certain type. Additionally, the quantity of preventive maintenance surpasses the availability of corrective and overhaul data.

C. Estimated coefficients of usage-based models of all cost components

The severity of the preventive maintenance visits is the only other model where some selected operational events were considered significant enough by the variable selection procedure described in Section 5.1.

	Event	Coefficient	s.e.	p-value
Intercept		-0.00642	0.00521	0.2179
TR	Temperature rising	0.00054	0.0065	0.93309
RM	Regular malfunctions	0.06264	0.01677	< 0.0001
TR:RM	Interaction between TR and RM	-0.01629	0.00644	0.0115

Table 6: Estimated coefficients of events and their one-way interactions for the preventive maintenance severity model.