Data-Driven Predictive Maintenance for Gas Distribution Networks

Wolfgang Betz[a,*], Iason Papaioannou[a,b], Tobias Zeh[c], Dominik Hesping[a], Tobias Krauss[c], Daniel Straub[a,b]

[a] *Eracons GmbH, Oberanger 42, 80331, München, Germany*
[b] *Engineering Risk Analysis Group, Technische Universität München, Germany*
[c] *Netze BW GmbH, Schelmenwasenstraße 15, 70567 Stuttgart, Germany*

Abstract

A generic data-driven approach is presented that employs machine learning to predict the future reliability of components in utility networks. The proposed approach enables utilities to implement a predictive maintenance strategy that optimizes life-cycle cost without compromising safety or creating environmental issues. Any machine learning technique that qualifies as a probabilistic classifier can be employed within the proposed approach. To identify the data-driven model that performs best, a practical metric to assess the performance of the competing models is proposed. This metric is specifically designed to quantify the forecasting performance with respect to maintenance planning. Additionally, a data-driven sensitivity analysis approach is discussed that allows assessing the influence of the different features on the model prediction. Through an application example, it is demonstrated how the proposed approach can be applied to predict future defect rates of pipe sections for maintenance planning in a large gas distribution network.

1. Introduction

A *gas distribution network* distributes the natural gas delivered over long distances by *trunk gas pipelines* to consumers. Gas distribution networks are operated by *natural gas utilities*. A key responsibility of natural gas utilities is to maintain the gas distribution network that they operate.

Unless regulated by national standards, a gas utility can pursue different maintenance policies [33, 11]: In a *reactive/corrective* maintenance policy, components are repaired or replaced once they fail. Contrary to that, a *proactive* maintenance policy aims at reducing the occurrences of unplanned defects [58]. Proactive maintenance policies include preventive and predictive policies. In a *preventive* maintenance policy, maintenance or inspection actions are performed at regular intervals. In a *predictive* maintenance policy, maintenance is performed based on the estimated state of the components [25, 26].

A problem of a reactive/corrective maintenance policy is that maintenance actions cannot be scheduled to balance the workload. Instead, the incurring workload is often unbalanced, which makes resource planning difficult. As a consequence, a reactive maintenance action is usually more expensive than a comparable scheduled maintenance action. Moreover, a comparatively high rate of observed component defects could potentially bring the operating utility company into disrepute and hinder the company from acquiring or renewing network licenses in the future.

For a *proactive* maintenance policy, resource planning can be done efficiently, but unplanned component defects can only be reduced and not totally prevented. With a proactive policy, more maintenance actions need to be performed on average than for a corrective policy, often resulting in higher network reliability. Indeed, preventive maintenance is a common strategy for gas networks [33, 52], whereby the individual components are inspected/maintained periodically.

A central objective of maintenance planning is to optimize the life-cycle cost without compromising safety or creating environmental issues [25, 30]. In this regard, predictive maintenance [51] has a larger potential compared to a periodically performed preventive maintenance [1]. For a predictive maintenance, the most critical components to maintain preferentially are identified based on a prediction of the future trend in the development of the conditions of network components [33]. The criticality can be expressed in terms of the expected current or future reliability (*reliability*-based maintenance) or risk (*risk*-based maintenance). Especially risk-based procedures are highly suitable to minimize life-cycle cost [1]. If the maintenance is performed reliability- or risk-based, the rate of observed defects tends to gradually decline compared to most other maintenance strategies [33, 30].

However, assessing the reliability and the associated risk of components in natural gas distribution networks is challenging due to incomplete knowledge about the state of the component (partly due to a lack of suitable deterioration models) and the unknown state of the surrounding environment [33, 64, 48, 21, 62]. For example, most pipes in gas distribution networks are buried underground and the present state of the pipe as well as the properties of the surrounding soil are uncertain. Therefore, a stochastic model is required to quantify the involved uncertainties and to derive the expected reliability and/or risk associated with the component.

In order to reduce the uncertainty in such a stochastic model,

*Corresponding author
Email addresses: betz@eracons.com (Wolfgang Betz), papaioannou@eracons.com (Iason Papaioannou), t.zeh@netze-bw.de (Tobias Zeh), hesping@eracons.com (Dominik Hesping), t.krauss@netze-bw.de (Tobias Krauss), straub@eracons.com (Daniel Straub)

Manuscript, published in ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng., Volume 8, Issue 2, March 2022 *DOI: 10.1061/AJRUA6.0001237*

historical data on the gas distribution network and on external factors that influence the state of the network components is essential. For example, collected historical data can help to identify the main causes of unexpected breakdowns [33, 58, 38, 44]. This requires a utility to invest in collecting relevant data that can help to reduce modeling uncertainties. Such data is already collected by many utilities [41]. Additionally, geospatial data on soil types, location-specific data on weather/climate conditions and data on traffic intensity can be readily acquired from geographic information systems for the coordinates of interest [16, 28].

So far, data-driven predictive maintenance of pipes in gas distribution networks has not received much attention in the literature. Contributions that discuss predictive maintenance for other utility networks include [32, 4, 17, 20, 2, 10]. Other work on data-driven predictive maintenance includes [52, 32, 36, 5, 14, 31, 13, 15, 59]. Contributions on artificial intelligence in the context of gas distribution networks include [65, 56, 63].

This contribution presents a novel data-driven approach that employs machine learning to predict the future reliability of pipes in gas distribution networks. In principle, any machine learning technique that qualifies as a *probabilistic classifier* can be employed within the proposed approach. We propose a practical metric to identify the data-driven model that performs best, in case different competing modeling approaches are investigated. Moreover, we introduce a data-driven global sensitivity analysis approach that allows assessing the influence of the different features on the model prediction. With some modifications the proposed approach is applicable to other component types or other types of utility networks.

2. Problem Statement

2.1. The feature vector and recorded past defects

To each pipe section S in the network, one can attribute a feature vector $\mathbf{x}$. The feature vector $\mathbf{x}$ is a collection of properties or characteristics associated with the particular pipe section S [8]. Characteristics of S can for example include its length, age, diameter, material type, defects that occurred in the past, location or associated district of the network. Features can also be based on geospatial data. This can include data on soil types, weather/climate conditions, vegetation in the pipe's vicinity or traffic intensity.

Note that the feature vector $\mathbf{x}$ does not only depend on S, but also on a point in time. In the following, we treat $\mathbf{x}$ as a function of S and y (i.e., $\mathbf{x}(S, y)$), where y denotes a specific calendar year. This implies that in order to predict the performance of S in year $y + 1$, one can embed information available up to (and including) year y in the feature vector $\mathbf{x}$.

Additional to the feature vector $\mathbf{x}$, it is essential that a (comprehensive) record of defects that occurred in the past on the network's pipe sections is available for the analysis. At least for a certain observation period or network section, this record must be complete; i.e., it must contain all (detected) defects.

2.2. The quantity to predict

The employed data-driven machine learning approach predicts for a specific pipe section S and calendar year y with feature vector $\mathbf{x}(S, y)$ the probability $\Pr[D_{S,y+1}|\mathbf{x}(S, y)]$ that defects will occur within calendar year $y + 1$.

We distinguish *three* outcome classes $D_{S,y} \in \{0, 1, 2\}$ for defects occurring within calendar year y on pipe section S:

$D_{S,y} = 0$**:** *No* defects occur.

$D_{S,y} = 1$**:** Exactly *one* defect occurs.

$D_{S,y} = 2$**:** *More than one* defects occur.

The outcome classes $D_{S,y} = 1$ and $D_{S,y} = 2$ should be combined into a single class if the recorded data contain almost no occurrences of multiple defects per pipe section and per year: In such case, a good prediction performance for $D_{S,y} = 2$ should not be expected and differentiating between $D_{S,y} = 1$ and $D_{S,y} = 2$ could actually lead to a decrease in model performance. The occurrence of defects at any two pipe sections S_a and S_b is modeled as conditionally independent given $\mathbf{x}(S_a, y)$ and $\mathbf{x}(S_b, y)$.

The associated risk [47] can be derived by multiplying $\Pr[D_{S,y+1}|\mathbf{x}(S, y)]$ with the consequences and costs associated with $D_{S,y+1}$ as a function of $\mathbf{x}(S, y)$ and by evaluating the expectation over all states of $D_{S,y+1}$.

2.3. How the analysis can support predictive maintenance

The derived predictive quantity (either probability or risk) is a useful input for maintenance planning. In *short-term* maintenance planning, the predictive quantity can be used for execution planning to decide which pipe sections to maintain/replace next in order to make optimal use of the available budget [43, 29, 34, 37], e.g., by ranking the network sections with respect to their predicted defect rates. In *medium-term* maintenance planning, the predictive quantity can be used to decide how much budget should be allocated to the individual sectors of the network for future maintenance actions [23], e.g., by assessing how many pipe sections need to be replaced in order to achieve a specific target rate of pipe defects. The predictive quantity could also be used for *inspection* planning, e.g., by using a risk-based approach to decide which pipe sections should be inspected next.

2.4. Restrictions and challenges of the modeling approach

The approach proposed in this work is illustrated by considering *pipe sections* as the network components of interest for the predictive modeling; other component types of the gas distribution network (e.g., valves and fittings) are not considered. The proposed approach could also be applied to predict the performance of other types of network components. However, in such a case, one must reassess which features to consider/include in the predictive analysis. Moreover, it is essential that a detailed list of past defects is available for the type of network component to be considered. Past defects are the basis upon which machine learning strategies predict the occurrence of defects in the future.

Preferably, only similar material types should be considered jointly within an analysis (e.g, different types of polyethylene (PE) pipes). If, for two material types, degradation depends on entirely different effects or processes (e.g., for steel and PE pipes), it is often better to assess their performance in two separate analyses (i.e., with two separate data-driven models).

In principle, the presented approach can also be applied to other types of utility networks. However, one needs to take into account that the term "failure" can have a different interpretation and different consequences in different utility networks. For natural gas distribution networks, the network typically remains operational even if a failure occurs at a pipe section. This may also hold for water distribution networks, but usually not for electricity networks. For gas distribution networks and for electricity networks, it is often clearly defined what a failure event is, whereas for water distribution networks a failure can be either a leak or a burst; e.g., [25].

3. Predictive modeling of pipe defects

3.1. Separation of data into a training *and a* test *data set*

The data-driven prediction model will forecast the probability $\Pr[D_{S,y_d+1}|\mathbf{x}(S,y_d)]$ of defects in calendar year y_d+1 on a specific pipe segment S with feature vector $\mathbf{x}(S,y_d)$, where y_d is the calendar year up to which recorded data (i.e., past observations) is available. We remark that the initial calendar year differs for each pipe section. Note that it is straightforward to modify the model so that it predicts defects in calendar year y_d+n, with $n \in \{1,2,\ldots\}$. For the final model that is used to generate the predictions, all data available up to year y_d will be used to train the model. However, for model development, in order to perform feature engineering, to identify the most suitable machine learning technique and to adjust the parameters of the data-driven model, the available data is split into a *training* and a *test* data set. During model development, the data-driven models are learned based on the *training* data set and the final model performance of the different modeling approaches is assessed and compared using the *test* data set [39, 8].

The splitting of the *training* and the *test* data set should be performed such that it corresponds well with the actual forecasting problem to solve. In the forecasting problem at hand, we use all available information up to a calendar year y_d to predict the probability of defects in the following calendar year y_d+1. Thus, the splitting of the *training* and the *test* data set should be done with respect to calendar years; i.e., all data up to and including calendar year y_s are assigned to the *training* data set and all data starting from year y_s+1 are assigned to the *test* data seft, where $y_s+1 \leq y_d$ must be maintained. The calendar year y_s for the split should be selected such that most data is assigned to the *training* data set. However, the data assigned to the *test* data set should still give a good representation of the expected average defect rates of the individual pipe sections. The statistical uncertainty in the defects observed in just a single year (i.e., for $y_s = y_d - 1$) is often relatively high.

Note: The splitting of the *training* and the *test* data set should not be done with respect to the pipe sections, e.g., through assigning 80% of the sections to the *training* data set and the remaining 20% to the *test* data set. For such a splitting approach, the focus of the *test* data set is on predicting the probability of the pipe section being defective over the entire recorded service life (whereas in practice we are only interested in the future defect rate of the pipe section), where (future) information on the performance of similar pipes is already available (which is generally not the case in practice).

3.2. Learning of data-driven models

After having separated the data into a *training* and a *test* data set, data-driven models can be developed. To train a model, all recorded data up to year y_s are used. The target variable is arranged as follows: For each operation year $y < y_s$ of pipe segment S, the model predicts the probability $\Pr[D_{S,y+1}|\mathbf{x}(S,y)]$ of *no* defects ($D_{S,y+1}=0$), *one* defect ($D_{S,y+1}=1$), or more than *one* defects ($D_{S,y+1}=2$) on pipe section S within year $y+1$; where the cases $D_{S,y+1}=1$ and $D_{S,y+1}=2$ can be combined if almost no observations of $D_{S,y+1}=2$ are available. The feature vector $\mathbf{x}(S,y)$ is the model input and the observed defects $D_{S,y+1}$ of the corresponding pipe section S in year $y+1$ are treated as model output. The parameters of the data-driven model are inferred such that the evaluated probability $\Pr[D_{S,y+1}|\mathbf{x}(S,y)]$ explains best the model output $D_{S,y+1}$ as a function of the model input $\mathbf{x}(S,y)$ for all $y < y_s$ and all sections S. The precise learning approach depends on the particular machine learning model.

Any *probabilistic classification* technique that can handle data as described above can be applied within our proposed approach. Consequently, within a specific project, a selection of different machine learning techniques can be employed to produce a variety of competing data-driven models. In the present context, we recommend gradient boosting machines [18, 12] and random forest [9] as probabilistic classifiers. However, other probabilistic classifiers like logistic regression or naive Bayes can also be investigated within a comparative performance assessment. Some probabilistic classifiers are known to distort the predicted probabilities, which can be calibrated in a post-processing step [42].

Additionally, instead of the standard probabilistic classifiers, a Bayesian model can be developed to probabilistically describe the occurrence of defects. Such a model is tailored to a specific project and the data observed in that project. Furthermore, it is based on engineering considerations of the processes that potentially cause defects. However, from a computational point of view, Bayesian models are typically significantly more expensive to learn than standard probabilistic classifiers.

3.3. Assessment of the prediction capabilities through prediction curves

Once different data-driven models are set up and the corresponding model parameters are learned using the *training* data set, the prediction capabilities of the models should be assessed and compared based on the model performance on the *test* data set. Instead of using standard machine learning measures to quantify the model performance, we propose a case-specific

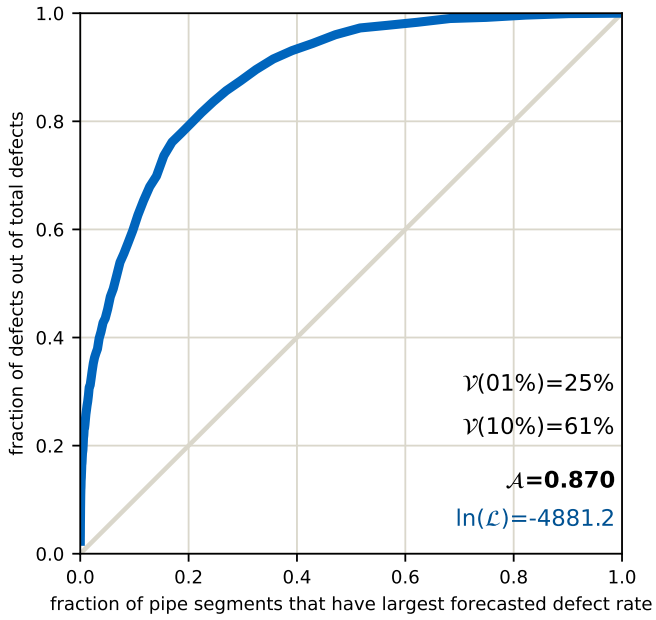

Fig. 1. The *prediction curve* of the *gradient boosting* model investigated in the presented application example.

measure. For a predictive model supporting maintenance planning, in the ideal case, the model would be able to correctly identify all pipe segments that will be defective in a certain future operation year. Thus, a model should be rated higher if it is able to identify a larger number of defective pipe segments .

For each pipe section S that is still operational in year $y_s + 1$ (i.e., each pipe section that appears at least once in the *test* data set), we evaluate the probability p_{d,S,y_s+1}:

$$p_{d,S,y_s+1} = \Pr\left[D_{S,y_s+1} = d|\mathbf{x}(S, y_s)\right], \tag{1}$$

for $d = 1$ and $d = 2$. The defect rate λ_{S,y_s+1} of pipe section S in year $y_s + 1$ can then be approximated as:

$$\lambda_{S,y_s+1} \approx \frac{p_{d=1,S,y_s+1} + p_{d=2,S,y_s+1} \cdot n_{D=2}}{l(S)}, \tag{2}$$

where $l(S)$ is the length of pipe section S. $n_{D=2}$ is the average number of observed defects on a section in a year, conditional on at least two defects occurring within that year, evaluated based on the *training* data set.

As a tailored measure for model performance, we introduce a so-called *prediction curve*. For a prediction curve, the horizontal and vertical axes of the plot range between $[0, 1]$. Points on the *prediction curve* are obtained as follows:

1. First, all n_{total} pipe sections $S(k)$, $k \in \{1, \ldots, n_{\text{total}}\}$, appearing in the *test* data set have to be sorted in descending order according to their future predicted defect rate $\lambda_{S(k),y_s+1}$; i.e., the pipe section $S(1)$ with the largest $\lambda_{S(1),y_s+1}$ is at the top of the list and the pipe section $S(n_{\text{total}})$ with the lowest rate $\lambda_{S(n_{\text{total}}),y_s+1}$ is at the end of the list. Instead of using the predicted defect rate $\lambda_{S(k),y_s+1}$ for calendar year $y_s + 1$, in principle, also the predicted average defect rate within the pipe's years of operation in the *test* data set could be used.
2. Next, for each pipe section $S(k)$, the total number $n_{S(k)}$ of defects recorded in the *test* data set is evaluated.
3. For the k pipe segments that have the highest predicted defect rate, the joint length of the pipe segments is evaluated, relative to the total length of all investigated pipe segments:
$$x_k = \frac{\sum_{i=1}^{k} l(S(i))}{\sum_{i=1}^{n_{\text{total}}} l(S(i))}, \tag{3}$$
where $l(S(i))$ is the length of the pipe section with the ith largest predicted defect length. This fraction x_k gives the value of the point on the horizontal axis that is associated with k selected samples.
4. To get the value z_k of the point on the vertical axis that is associated with k selected samples, the number of defects recorded for the selected pipes in the *test* data set is counted relative to the total number of defects recorded in the *test* data set:
$$z_k = \frac{\sum_{i=1}^{k} n_{S(i)}}{\sum_{i=1}^{n_{\text{total}}} n_{S(i)}}. \tag{4}$$

An exemplary *prediction curve* is shown in Fig. 1.

The *prediction* curve allows to visually assess and compare the performance of a data-driven model. To quantitatively compare different models, the point on the curve that is most relevant for the predictive maintenance problem at hand is selected. For example, if approximately 1% of all pipes are maintained/replaced each calendar year, then the vertical value of the curve $\mathcal{V}(1\%)$ that belongs to the 1% value on the horizontal curve is used. The $\mathcal{V}(1\%)$ states what fraction of defects could potentially be avoided if 1% of the identified most critical pipes are replaced. For example, with the predictive model considered in Fig. 1, replacing the 1% most critical pipe sections leads to a reduction of 25% of pipe failures. Besides the $\mathcal{V}(1\%)$ value, any other fraction can be used; e.g., $\mathcal{V}(10\%)$.

Alternatively, the area $\mathcal{A}$ under the *prediction curve* can be used to quantify the overall predictive capability of a given model, similar to the AUC of an ROC curve [40]. The value of $\mathcal{A}$ should be within $[0.5, 1)$. A value of 0.5 signifies that the model is not better than chance. The larger the value of $\mathcal{A}$, the better the overall predictive performance of the model. However, even the ideal model cannot achieve a value of 1 in practice, since defects are not concentrated on a single pipe section.

The *log-likelihood* $\ln(\mathcal{L})$ of the prediction on the *test* data set can also be used to assess model performance:

$$\ln(\mathcal{L}) = \sum_{i=1}^{n_{\text{total}}} \sum_{\forall y \in \mathbf{Y}(S(i))} \ln\left(\Pr\left[D_{S(i),y}|\mathbf{x}(S, y-1)\right]\right), \tag{5}$$

where $\mathbf{Y}(S(i))$ is the set operation years contained in the *test* data set for pipe section $S(i)$. The larger $\ln(\mathcal{L})$, the better the model predicted the actual observations $D_{S(i),y}$.

For some machine learning techniques, a probability of *zero* could be associated with a certain outcome class $D_{S(i),y}$. If

a different outcome class is observed, a value of $-\infty$ will be associated with $\ln(\mathcal{L})$. In such cases the probability calibration strategies mentioned in [42] can help to obtain meaningful values for $\ln(\mathcal{L})$.

The *log-likelihood* can also be evaluated with respect to the *training* data set, which we will in the following refer to as *data-fit*.

3.4. Sensitivity analysis

Conducting a sensitivity analysis helps us to understand (i) which features have the largest influence on the predicted damage rate or (ii) which features could potentially be omitted as their impact on the damage rate is small [50]. For the proposed data-driven approach, we suggest to use a global sensitivity analysis method that allows to treat the model as a black-box; i.e., a sensitivity method that is not specific to the used machine learning model. Such a generally applicable approach allows us to easily compare and assess the behavior of a set of (fundamentally) different predictive classifiers within a common framework.

We suggest to employ a variance-based global sensitivity analysis [49, 54] that estimates the first-order Sobol' indices [53]. The first-order sensitivity index S_i measures the contribution of feature X_i on the variance of the output, where the impact of interactions between X_i and other input features is not considered. The index S_i for input feature X_i is defined as:

$$S_i = \frac{\mathrm{Var}_{X_i}\left[\mathrm{E}_{\mathbf{x}_{-i}}\left[\lambda_{S,y+1}(\mathbf{x})\middle|X_i\right]\right]}{\mathrm{Var}\left[\lambda_{S,y+1}(\mathbf{x})\right]}, \tag{6}$$

where $\mathrm{E}_{\mathbf{x}_{-i}}\left[\lambda_{S,y+1}(\mathbf{x})\middle|X_i\right]$ is the expected defect rate with respect to all input features except X_i, which is fixed. The indices can be used for factor prioritization; i.e., to identify which factors have the largest influence on damage rate predicted by the model (see point (i) above). The suggested procedure to conduct the sensitivity analysis is:

1. All feature vectors of the *training* data set are treated as input samples to the sensitivity analysis.
   Alternatively, the combined set of feature vectors of both the *training* and the *test* data set could be used as input to reduce the statistical uncertainty in the estimated sensitivities.
2. For each feature vector $\mathbf{x}$ of pipe section S in year y, the defect rate $\lambda_{S,y+1}$ is evaluated with the data-driven model of interest.
   Alternatively, for initial feature selection and extraction, the recorded defects can be used instead of the defect rates. This allows to assess the sensitivities of the recorded defects without an underlying model.
3. Using the matrix of collocated feature vectors as input, and the vector of associated defect rates as output, the first order Sobol' indices can be directly estimated with the algorithm explained in [35], without further model evaluations.
4. The relative importance of each feature is obtained by normalizing the vector of first order Sobol' indices.

3.5. Forecasting of future defect rates

For forecasting the future defect rates $\lambda_{S,y_\mathrm{d}+1}$, the selected probabilistic classifier is trained on the *full* data set. Using the resulting data-driven model, the probabilities $\Pr[D_{S,y_\mathrm{d}+1}|\mathbf{x}(S,y_\mathrm{d})]$ of defects in calendar year $y_\mathrm{d}+1$ are evaluated for all active pipe sections of the network. In a post-processing step, the defect rates $\lambda_{S,y_\mathrm{d}+1}$ are obtained according to Eq. (2).

Both $\Pr[D_{S,y_\mathrm{d}+1}|\mathbf{x}(S,y_\mathrm{d})]$ and $\lambda_{S,y_\mathrm{d}+1}$ can then be used as input for maintenance planning. Alternatively, the risk associated with each pipe section S can be calculated as basis for maintenance planning.

4. Case study network

As application example for the practical demonstration of the proposed approach, we use a large gas distribution network located in southwest Germany. The available data set includes all relevant information starting from 1910 up to the end of 2019. Only *steel* pipes are considered here; pipes made out of *PE* or *gray cast iron* are not included. The total length of *steel* pipes in the database is 2.7×10^3 km. The total observed joint operation time of all pipe segments is 1.7×10^6 years. Further properties of the data set are summarized in Table 1.

For each pipe section, we take into account the past history, as well as the pipe's characteristics and the available geospatial data at the pipe's location. This includes: number of defects observed in each year of operation, length of the section, year of construction, material group (this also includes the pipe's coating), years of operation, diameter, associated district of the network, corrosion prevention measures, type of soil in the pipe's vicinity, type of land use in the pipe's vicinity (urban, rural, industrial), traffic volume on ground surface, vegetation in the pipe's proximity, third-party influences and years of critical findings during inspections. A *critical finding* is recorded if there is (due to construction work in the vicinity) an opportunity for a visual inspection of a pipe section and if the conducted visual inspection shows that the pipe is in an unfavorable state. For the *type of soil* in the pipe's vicinity, we only take the predominant soil type into account if the pipe traverses different soil layers. A *third-party influence* is logged if in the associated district of the network a pipe is damaged in an external event (e.g., due to an excavation). Thus, third-party influences are not directly associated with a specific pipe section, but with an entire district of the network. It could be used as an indicator for the intensity of construction activity in the pipe's vicinity. Data on the *vegetation* in the pipe's proximity was only available for roughly a third of the entire pipe sections.

4.1. Separation of data into a training and a test data set

We assign the last *five* years of recorded data to the *test* data set; i.e., $y_s = 2014$. A comparison between the full data set, the *training* data set and the *test* data set is given in Table 1. For the application example at hand, we can use roughly 10^3 recorded defects (around 8% of all recorded defects) to assess the forecasting performance of the developed data-driven model.

Table 1. Key data of the pipe network investigated in the application example.

	full data set	**training data set**	**test data set**
number of pipe segments	4.4×10^4	4.3×10^4	3.0×10^4
total pipe length [km]	2.7×10^3	2.7×10^3	2.2×10^3
joint operation time [years]	1.7×10^6	1.5×10^6	0.2×10^6
total observed "size" [km years]	1.1×10^5	1.0×10^5	0.1×10^5
total recorded defects	1.2×10^4	1.1×10^4	0.1×10^4

If the *size* of a data set is quantified as the sum over the product of *length* and *operation years* of the pipe segments, about 9% of the data is assigned to the *test* data set.

Note that in Table 1, for the quantities "number of pipe segments" and "total pipe length", the values given for the *training* and *test* data set do not add up to the values given for the *full* data set. This is due to splitting up the data with respect to the year $y_s = 2014$. Consequently, a pipe can be both in the *training* data set (for the operation time up to 2014) and in the *test* data set (for the operation time since 2015).

4.2. Learning of data-driven models

As predictive classifiers, we investigate *gradient boosting*, *random forest* and *naive Bayes*. Additionally, we set up a predictive Bayesian model tailored to the problem at hand. A short summary of implementation specific details of each data-driven model is given in the following:

gradient boosting: The *gradient boosting* classifier [22, 18, 19] from the *Python* library *scikit-learn* (version 0.23.2) is employed [45]. We use 500 for the number of weak learners and a learning rate of 0.1. The early stopping criterion is activated when the validation score does not improve for 5 subsequent iterations with a validation fraction of 0.1.

random forest: The *random forest* classifier [9] from the *Python* library *scikit-learn* (version 0.23.2) is employed [45]. We use a total of 100 trees. All features are considered when looking for the best split, where the quality of a split is measured in terms of the *entropy*.

naive Bayes: We use our own Python implementation of *naive Bayes*. In the final model, only the features *year of operation*, *length of the section*, *material group* and associated *district* of the network are considered. The data of the features *year of operation* and *length of the section* were both discretized into 7 categories. The data of the *material group* and the associated *district* of the network is already discrete.

Bayesian modeling approach: An expert-based parametric model that employs Gaussian process regression is used. The uncertain model parameters are inferred with a Bayesian approach. An overview of the Bayesian model is presented in Appendix I.

4.3. Assessment of the prediction capabilities through prediction curves

Fig. 1 presents the *prediction curve* of the investigated *gradient boosting* model. Characteristic values that quantify the model performance are listed in Table 2 for all investigated models. In terms of $\mathcal{V}(10\%)$, the *gradient boosting* model exhibits the best performance. With respect to $\mathcal{V}(1\%)$, both *gradient boosting* and the full *Bayesian* model perform equally well. With respect to $\mathcal{A}$, *gradient boosting*, *naive Bayes* and the full *Bayesian* model perform equally well. For the *log-likelihood* on the *test* data set and the *data-fit* on the *training* data set, the full *Bayesian* model performs best. The *naive Bayes* model performs surprisingly well with respect to $\mathcal{V}(1\%)$, $\mathcal{V}(10\%)$ and $\mathcal{A}$. However, unlike *gradient Boosting* and the full *Bayesian* model, we observed that in general the model performance of *naive Bayes* is relatively volatile for alternative separations of the data set.

For the example at hand, *gradient boosting* appears to be the most suitable machine learning algorithm. Even though the Bayesian model exhibits a similar performance as *gradient boosting*, the former model is more involved to set up and more computationally expensive to learn.

4.4. Sensitivity analysis

We apply the approach outlined in Section "Sensitivity Analysis" to estimate the first-order Sobol' indices for the gradient boosting model. The resulting sensitivities are illustrated in Fig. 2.

Four of the five features with the largest impact on the predicted defect rate in the gradient boosting model can be linked to the history of damages of the pipe section: the length of the pipe section, the past defects, critical findings and third-party influences. The *length* is the feature with the largest sensitivity. This effect is artificially induced by the management of the history of the pipes in the database: If part of the pipe section on which a defect occurred needs to be replaced with a new pipe, each obtained sub-section of the original pipe is assigned a new ID number. Thus, even the part of the pipe that is not replaced is assigned a new ID number in this process. Consequently, the shorter the pipe, the more likely it is that it was split-up before, which in turn means that shorter pipes are likely problematic pipe sections. Besides the features that can be linked to the damage history, the *years of operation* is the feature with the third-largest sensitivity. All other features not mentioned here have a relatively small sensitivity index.

Table 2. Characteristic values that quantify the performance of the models investigated in the presented application example.

	test data set				*training* data set
	$\ln(\mathcal{L})$	$\mathcal{V}(1\%)$	$\mathcal{V}(10\%)$	$\mathcal{A}$	data-fit
gradient boosting	-4.8×10^3	25%	60%	0.87	-4.9×10^4
random forest	–	16%	49%	0.77	–
naive Bayes	-6.3×10^3	23%	57%	0.87	-4.6×10^4
Bayesian model	-4.7×10^3	25%	59%	0.87	-3.2×10^4

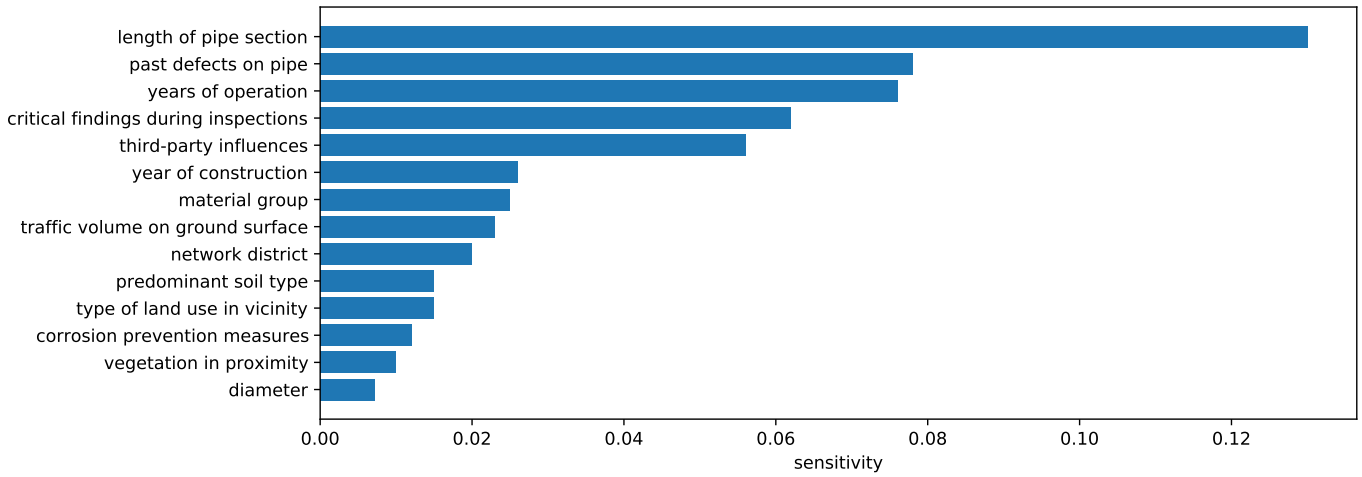

Fig. 2. Sensitivities obtained with the first-order Sobol' analysis conducted on the gradient boosting model for the presented application example.

4.5. Forecasting of future defect rates

The forecasted defect rates λ_{S,y_d+1}, obtained with gradient boosting trained on the *full* data set, are used to rank all active pipe sections of the network. Such a ranking of the pipe sections in the network is crucial for implementing an efficient maintenance plan [33]. Based on this ranking, the most critical pipes in the network can be identified and prioritized for upcoming maintenance and repair actions.

Up to now, the maintenance in the investigated gas distribution network is based on the ranking obtained from a heuristic measure. This heuristic ranking is calculated based on a formula that blends practical experience with available damage data. In particular, the state of a pipe section is estimated based on (i) parameters from Weibull statistics and (ii) hazard components like traffic or vegetation in proximity to the grid. The heuristic ranking is obtained as the product of these two influence groups, where each influence group is computed using a weighted summation. The goal is to replace the employed heuristic rating of pipes with the forecasted future defect rates. In order to visually compare the performance of the proposed data-driven model, with the one of the current heuristic approach, we plot in Fig. 3 the forecasted defect rates over the heuristic rating for all pipe sections in the *test* data set, with orange points indicating defective pipe sections. The plot shows that a criticality ranking based on the forecasted defect rates will differ considerably from a criticality ranking based on the heuristic rating. Considerably more defective pipe sections are amongst the most critical pipes identified based on the defect rate predicted with machine learning (25% of all defective pipes), compared to the critical pipe sections identified based on the heuristic rating (2% of all defective pipes).

5. Concluding remarks

The proposed data-driven approach is applicable to assess the condition of components in utility networks. It is not linked to a particular machine learning algorithm, but works with any probabilistic classifier. The proposed approach allows to assess and compare the performance of different probabilistic classifiers based on a practical metric, which can be summarized in prediction curves. This metric is specifically designed to quantify the forecasting performance with respect to maintenance planning.

We demonstrate the applicability of the proposed approach by assessing the defect probability and rate of steel pipe sections in a large gas distribution network. For the investigated application example, *gradient boosting* exhibits the best performance of all investigated machine learning techniques. The model input features with the largest impact are the ones for which a direct link to the history of damages of the pipe sections can be established. Geospatial data is found to have a relatively moderate impact on the forecasted future damage rate.

Data Availability Statement

Some or all data, models, or code generated or used during the study are proprietary or confidential in nature and may only

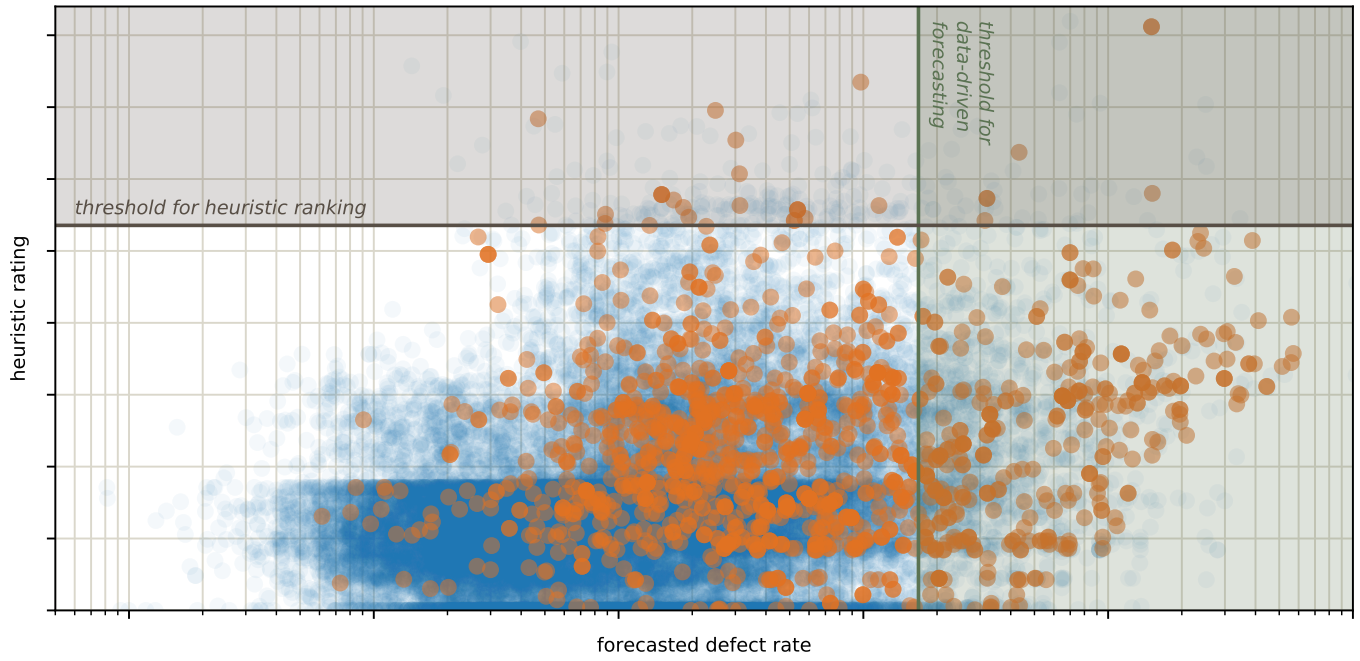

Fig. 3. Comparison of the forecasted defect rate (using the gradient boosting model) with the heuristic ranking for the presented application example. The *blue* dots represent pipe sections without defects, the *orange* dots represent pipe sections on which defects occurred. The pipe sections in the area shaded in green are the most critical pipe sections according to the forecasted defect rate that cover 1% of the entire network length. The pipe sections in the area shaded in brown are the most critical pipe sections according to the heuristic ranking used up to now that cover 1% of the entire network length.

be provided with restrictions.

All data used in the application example are proprietary and confidential. The data was provided by the utility service provider 'NetzeBW' and cannot be shared.

Appendix A. Overview of the Bayesian model used in the case study

In the case study, a Bayesian modeling approach that employs Gaussian process regression is used for one of the data-driven models. The Bayesian approach is employed to infer the uncertain model parameters of an expert-based parametric model. The Bayesian inference is performed numerically using the aBUS [7] approach in a manner similar to [60]. aBUS was selected as inference method, as its performance does not depend on the number of uncertain model input parameters; i.e., aBUS can cope with high dimensional inference problems. In the BUS (Bayesian Updating with Structural reliability methods) approach, the space of uncertain model parameters is augmented by a uniform random variable and the Bayesian inference problem is interpreted as a structural reliability problem [55]. The a posteriori samples are obtained as samples conditioned on the failure domain of the structural reliability problem. To obtain this failure domain, the likelihood function needs to be scaled by a constant. In aBUS, the structural reliability problem is solved with Subset Simulation [3] and the scaling of the likelihood function [6] is performed adaptively.

The parametric model is gradually improved by sequentially adding features to the model. Initially, for each *material group*, four one-dimensional Gaussian processes [61] are introduced. Conditional on the respective material group, the corresponding four Gaussian processes are superposed to predict the probability of $D_{S,y+1}$ as a function of the *year of operation*. Two Gaussian processes are superposed to model the probability of one defect in a year ($D_{S,y+1} = 1$) and the other two are superposed to model the probability of more than one defect in a year ($D_{S,y+1} = 2$). In each superposition, the outcome of one of the two Gaussian processes is multiplied with the *length of the pipe section*. The separation into two Gaussian processes allows to account (i) for effects for which the length is directly proportional to the occurrence rate of defects and (ii) for effects for which the occurrence rate of defects does not depend on the length of a pipe section. Higher order influences of the length of pipe sections were not considered. The Gaussian processes need to be discretized, which significantly increases the number of uncertain parameters that have to be inferred. This is not a problem for aBUS, as the performance of the method is not influenced by the total number of uncertain model parameters.

All other features were thereafter added sequentially to the above described stochastic model. For each added feature, a suitable adaptive parametric model was identified that also captures potential dependencies with other features already accounted for in the stochastic model.

References

[1] Ambühl, S. and Sørensen, J. D. (2017). "On different maintenance strategies for casted components of offshore wind turbines." *Report No. DCE No. 222*, Aalborg University, Aalborg, Denmark.

[2] Asgarpoor, S. and Doghman, M. (1999). "A maintenance optimization program for utilities' transmission and distribution systems." *Proceedings of the 1999 North American Power Symposium*, 454–459.

[3] Au, S.-K. and Beck, J. L. (2001). "Estimation of small failure probabilities in high dimensions by Subset Simulation." *Probabilistic Engineering Mechanics*, 16(4), 263–277.

[4] Bao, Z., Gui, C., and Guo, X. (2018). "Short-term line maintenance scheduling of distribution network with pv penetration considering uncertainties." *IEEE Access*, 6, 33621–33630.

[5] Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento Jr, C., Prendinger, H., and Henriques, E. M. (2018). "Forecasting fault events for predictive maintenance using data-driven techniques and arma modeling." *Computers & Industrial Engineering*, 115, 41–53.

[6] Betz, W., Beck, J. L., Papaioannou, I., and Straub, D. (2018a). "Bayesian inference with reliability methods without knowing the maximum of the likelihood function." *Probabilistic Engineering Mechanics*, 53, 14–22.

[7] Betz, W., Papaioannou, I., Beck, J. L., and Straub, D. (2018b). "Bayesian inference with subset simulation: strategies and improvements." *Computer Methods in Applied Mechanics and Engineering*, 331, 72–93.

[8] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[9] Breiman, L. (2001). "Random forests." *Machine learning*, 45(1), 5–32.

[10] Butler, K. L. (1996). "An expert system based framework for an incipient failure detection and predictive maintenance system." *Proceedings of International Conference on Intelligent System Application to Power Systems*, IEEE, 321–326.

[11] Chamoso, P., De Paz, J. F., Bajo, J., and Villarrubia, G. (2018). "Agent-based tool to reduce the maintenance cost of energy distribution networks." *Knowledge and Information Systems*, 54(3), 659–675.

[12] Chen, T. and Guestrin, C. (2016). "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

[13] Cline, B., Niculescu, R. S., Huffman, D., and Deckel, B. (2017). "Predictive maintenance applications for machine learning." *2017 annual Reliability and Maintainability Symposium (RAMS)*, IEEE, 1–7.

[14] Coleman, C., Damodaran, S., Chandramouli, M., and Deuel, E. (2017). "Making maintenance smarter: Predictive maintenance and the digital supply network." *Deloitte University Press*.

[15] Daily, J. and Peterson, J. (2017). "Predictive maintenance: How big data analysis can improve maintenance." *Supply Chain Integration Challenges in Commercial Aerospace*, Springer, 267–278.

[16] Daniela, L. et al. (2011). "Decision-support tools for municipal infrastructure maintenance management." *Procedia Computer Science*, 3, 36–41.

[17] de Faria Jr, H., Costa, J. G. S., and Olivas, J. L. M. (2015). "A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis." *Renewable and Sustainable Energy Reviews*, 46, 201–209.

[18] Friedman, J. H. (2001). "Greedy function approximation: a gradient boosting machine." *Annals of Statistics*, 1189–1232.

[19] Friedman, J. H. (2002). "Stochastic gradient boosting." *Computational Statistics & Data Analysis*, 38(4), 367–378.

[20] Ghazvini, M. F., Morais, H., and Vale, Z. (2012). "Coordination between mid-term maintenance outage decisions and short-term security-constrained scheduling in smart distribution systems." *Applied Energy*, 96, 281–291.

[21] Haladuick, S. and Dann, M. R. (2018). "Decision making for long-term pipeline system repair or replacement." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 4(2), 04018009.

[22] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

[23] Hastings, N. A. (2021). *Physical Asset Management*. Springer, 3 edition.

[24] Kanakoudis, V. (2004). "Vulnerability based management of water resources systems." *Journal of Hydroinformatics*, 6(2), 133–156.

[25] Kanakoudis, V. and Tolikas, D. (2001). "The role of leaks and breaks in water networks: technical and economical solutions." *Journal of Water Supply: Research and Technology—AQUA*, 50(5), 301–311.

[26] Kanakoudis, V. and Tolikas, D. (2004). "Assessing the performance level of a water system." *Water, Air and Soil Pollution: Focus*, 4(4), 307–318.

[27] Kanakoudis, V. and Tsitsifli, S. (2011). "Water pipe network reliability assessment using the dac method." *Desalination and Water Treatment*, 33(1-3), 97–106.

[28] Kaźmierczak, J., Loska, A., and Dabrowski, M. (2012). "Use of geospatial information for supporting maintenance management in a technical network system." *21th European Congress on Maintenance and Asset Management Euromaintenance*, 287–297.

[29] Kelly, A. (2006). *Strategic Maintenance Planning*, Vol. 1. Elsevier.

[30] Khan, F. I. and Haddara, M. M. (2003). "Risk-based maintenance (rbm): a quantitative approach for maintenance/inspection scheduling and planning." *Journal of Loss Prevention in the Process Industries*, 16(6), 561–573.

[31] Korvesis, P. (2017). "Machine learning for predictive maintenance in aviation." Ph.D. thesis, Université Paris-Saclay (ComUE), Paris.

[32] Langroudi, P. P. and Weidlich, I. (2020). "Applicable predictive maintenance diagnosis methods in service-life prediction of district heating pipes." *Environmental and Climate Technologies*, 24(3), 294–304.

[33] Leoni, L., BahooToroody, A., Abaei, M. M., De Carlo, F., Paltrinieri, N., and Sgarbossa, F. (2021). "On hierarchical bayesian based predictive maintenance of autonomous natural gas regulating operations." *Process Safety and Environmental Protection*, 147, 115–124.

[34] Levitt, J. (2003). *Complete Guide to Preventive and Predictive Maintenance*. Industrial Press Inc.

[35] Li, C. and Mahadevan, S. (2016). "An efficient modularized sample-based method to estimate the first-order Sobol' index." *Reliability Engineering & System Safety*, 153, 110–121.

[36] Li, G., Huang, Y., Bie, Z., and Ding, T. (2020). "Machine-learning-based reliability evaluation framework for power distribution networks." *IET Generation, Transmission & Distribution*, 14(12), 2282–2291.

[37] Mobley, R. K. (2002). *An Introduction to Predictive Maintenance*. Elsevier.

[38] Montiel, H., Vílchez, J. A., Arnaldos, J., and Casal, J. (1996). "Historical analysis of accidents in the transportation of natural gas." *Journal of Hazardous Materials*, 51(1-3), 77–92.

[39] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT press.

[40] Narkhede, S. (2018). "Understanding auc-roc curve." *Towards Data Science*, 26, 220–227.

[41] NARUC (2020). "Artificial intelligence for natural gas utilities: A primer." *Report no.*, National Association of Regulatory Utility Commissioners (October). https://pubs.naruc.org/pub/F74D4EC2-155D-0A36-310F-F5B22DC3E286.

[42] Niculescu-Mizil, A. and Caruana, R. (2005). "Predicting good probabilities with supervised learning." *Proceedings of the 22nd International Conference on Machine learning*, 625–632.

[43] Palmer, R. D. (2013). *Maintenance Planning and Scheduling Handbook*. McGraw-Hill Education.

[44] Papadakis, G. A. (1999). "Major hazard pipelines: a comparative study of onshore transmission accidents." *Journal of Loss Prevention in the Process Industries*, 12(1), 91–107.

[45] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*, 12, 2825–2830.

[46] Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). "Coda: convergence diagnosis and output analysis for mcmc." *R news*, 6(1), 7–11.

[47] Rausand, M. (2013). *Risk Assessment: Theory, Methods, and Applications*, Vol. 115. John Wiley & Sons.

[48] Sacco, T., Compare, M., Zio, E., and Sansavini, G. (2019). "Portfolio decision analysis for risk-based maintenance of gas networks." *Journal of Loss Prevention in the Process Industries*, 60, 269–281.

[49] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008a). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons.

[50] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008b). "Introduction to sensitivity analysis." *Global sensitivity analysis. The primer*, 1–51.

[51] Selcuk, S. (2017). "Predictive maintenance, its implementation and latest trends." *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 231(9), 1670–1679.

[52] Serradilla, O., Zugasti, E., and Zurutuza, U. (2020). "Deep learning models for predictive maintenance: a survey, comparison, challenges and prospect." *arXiv preprint arXiv:2010.03207*.

[53] Sobol, I. M. (1993). "Sensitivity estimates for nonlinear mathematical models." *Mathematical Modelling and Computational Experiment*, 1(4), 407–414.

[54] Sobol, I. M. (2001). "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates." *Mathematics and Computers in Simulation*, 55(1-3), 271–280.

[55] Straub, D. and Papaioannou, I. (2015). "Bayesian updating with structural reliability methods." *Journal of Engineering Mechanics*, 141(3), 04014134.

[56] Su, H., Zio, E., Zhang, J., Yang, Z., Li, X., and Zhang, Z. (2018). "A systematic hybrid method for real-time prediction of system conditions in natural gas pipeline networks." *Journal of Natural Gas Science and Engineering*, 57, 31–44.

[57] Tsitsifli, S. and Kanakoudis, V. (2010). "Predicting the behavior of a pipe network using the "critical z-score" as its performance indicator." *Desalination*, 250(1), 258–265.

[58] Tsitsifli, S., Kanakoudis, V., and Bakouros, I. (2011). "Pipe networks risk assessment based on survival analysis." *Water Resources Management*, 25(14), 3729–3746.

[59] Tsui, K. L., Chen, N., Zhou, Q., Hai, Y., and Wang, W. (2015). "Prognostics and health management: A review on data driven approaches." *Mathematical Problems in Engineering*, 2015.

[60] Uribe, F., Papaioannou, I., Betz, W., and Straub, D. (2020). "Bayesian inference of random fields represented with the karhunen–loève expansion." *Computer Methods in Applied Mechanics and Engineering*, 358, 112632.

[61] Vanmarcke, E. (2010). *Random Fields: Analysis and Synthesis*. World Scientific Publishing, Singapore, 2 edition.

[62] Wijs, R., Nane, G., Leontaris, G., Van Manen, T., and Wolfert, A. (2020). "Improving subsurface asset failure predictions for utility operators: A unique case study on cable and pipe failures resulting from excavation work." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 6(2), 05020002.

[63] Wiranowski, M. I. and Bergles, S. B. (2015). "Big data: Opportunities and constraints for gas utilities." *Report no.*, Wilkinson Barker Knauer LLP (July).

[64] Yuhua, D. and Datao, Y. (2005). "Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis." *Journal of Loss Prevention in the Process Industries*, 18(2), 83–88.

[65] Zemenkova, M. Y. (2019). "Neural network monitoring and predictive control of the reliability and safety of gas distribution networks using deep learning algorithms." *IOP Conference Series: Materials Science and Engineering*, Vol. 663, IOP Publishing.