Evaluating the role of deterioration models for condition assessment of sewers

Marius Møller Rokstad and Rita Maria Ugarelli

ABSTRACT

Ensuring reliable structural condition of sewers is an important criterion for sewer rehabilitation decisions. Deterioration models applied to sewer pipes support the rehabilitation planning by means of prioritising pipes according to their current and predicted structural status. There is a benefit in applying such models if sufficient inspection data for calibration, an appropriate deterioration model, and adequate covariates to explain the variability in the conditions are available. In this paper it is discussed up to what level the application of sewer deterioration models can be beneficial under limited data availability. The findings show that the indirect nature of the explanatory covariates which are commonly used in sewer deterioration models makes it difficult to harness any benefit from modelling sewer conditions at a network level, but that the deterioration model application still may be beneficial for prioritising inspection candidates. The prediction power of the current sewer deterioration models is limited by the adequacy of the explanatory variables available, and by the fact that different failure modes are mixed in the aggregated condition class, and not modelled explicitly.

Key words | CCTV inspection, decision support, infrastructure asset management, sensitivity analysis, sewer condition assessment, sewer deterioration

Marius Møller Rokstad (corresponding author)
Rita Maria Ugarelli
Department of Hydraulic and Environmental Engineering,
Norwegian University of Science and Technology,
SP. Andersens veg 5,
7491 Trondheim,
Norway
E-mail: *marius.rokstad@ntnu.no*

Rita Maria Ugarelli
SINTEF Building and Infrastructure,
Pb. 124 Blindern,
0314 Oslo,
Norway

LIST OF ABBREVIATIONS

AI:	Artificial Intelligence
ANN:	Artificial Neural Network
CARE-S:	Computer Aided REhabilitation of Sewer Networks (European research project)
CC:	Condition Class
CCTV:	Closed-circuit television
IAM:	Infrastructure Asset Management
NHMC:	Non-homogeneous Markov Chain
RF:	Random Forest
SMS:	Secured and Monitored Service from Oslo VAV

LIST OF NOTATIONS

i, j, k and m	are indexing variables
c	is the number of condition labels in a condition classification system
$t_{i,\text{inst}}$	is the year of installation for pipe i
$t_{i,\text{insp}}$	is the year of inspection for inspection j
$t_i = t_{i,\text{insp}} - t_{i,\text{inst}}$	is the age of pipe at a given inspection
y_j	is the observed condition class in inspection j
Z_j	is a set of explanatory covariates for the pipe in inspection j
S	is the full set of pipes with explanatory covariates in a sewer network or stratum, containing N pipes
O_j	is a condition observation of a pipe in S, which contains the identification of the pipe which has been observed i, a vector of explanatory covariates to be used in a deterioration model Z_j, the year of inspection, and the observed condition class y_j
R	is a subset of observations O_j from S, containing $n < N$ pipes

$\hat{y}_i(T)$	is the estimated condition class of pipe i in year T
$\Pr(\hat{y}_i(T))$	is the estimated probability for pipe i to be in condition class y at time T
$\mathbf{Z}_{0i}$	is a vector of time-independent covariates for pipe i
$\mathbf{Z}_{1i}$	is a vector of time-dependent covariates for pipe i
S_{ik}	is the survival function for pipe i used in GompitZ, i.e. the probability that pipe i is in condition class k or better
$\boldsymbol{\theta} = [\alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1]$	is the vector of model parameters in GompitZ
$u_i \sim N(0, \sigma^2)$	is the individual frailty factor in the GompitZ survival function
$\boldsymbol{x} = [x_1, \ x_2, \ldots, \ x_c]$, where $\boldsymbol{x_k}$	is the number of pipes observed in condition class k for a set of observations R
$\boldsymbol{m} = [m_1, m_2, \ldots, m_c]$, where m_k	is the total number of pipes in CC k in a set S
p_k	is the proportion of pipes in condition class k
$\text{var}(p_k \mid n)$	is the selection variance estimate of p_k given a subset size n
L_i	is the length of pipe i
L_0	is the length of inspected sewers
L_{c0}	is the length of inspected sewers in a critical condition

INTRODUCTION

Well-functioning sewer infrastructure is a prerequisite for the prosperity of any modern society. However, the sewers of the industrialised world are currently ageing, as the peak of capital investment has passed (Alegre & Matos 2009). Traditionally it has been economically feasible to apply a reactive management strategy, and repair whenever a failure occurs; the reactive strategy is however expected to become less viable as the systems age.

To ensure that sewer systems provide long-term sustainable levels of service, at a sustainable level of cost, one may employ infrastructure asset management (IAM) as a successor to the reactive strategy. IAM may be understood as '... a set of management, financial, economic, engineering activities, systematic and coordinated, to optimally manage the physical assets and their associated performance, risks and expenditures over their life cycle with the objective of ensuring level of service in the most cost-effective manner' (Ugarelli et al. 2010). To implement IAM one must define performance, cost and risk objectives, and diagnose one's system by assessing the gap between these objectives and the actual status of the system.

More specifically, to prioritise sewer pipes for rehabilitation due to structural condition, one must first collect data for the diagnosis. There exist many methods for assessing the structural condition of sewer pipes (Kley et al. 2013), and the most common method is to perform closed-circuit television (CCTV) inspections. Defects and dysfunctions (Pollert et al. 2005; Le Gauffre et al. 2007) which are registered from the inspection footage, are used to classify pipes in so called condition classes (CC), ranging from 1 to 4 or 5 according to the standard applied (in general 1 stands for 'as good as new' and 4 or 5 for 'close to collapse'). The classification process is called condition assessment. Since the resulting CC is based on a standardised protocol, and therefore in principle an objective term, it can be used as a response variable in a deterioration model. The combination of inspection data and model predictions form the analysis, which will help in identifying rehabilitation needs and support IAM decisions.

There is thus a link between data from the field (inspections), models and IAM decisions, when the objectives of a utility are given. Given that one has an appropriate model, one would expect that the predictive accuracy would increase as the amount of available calibration data increases. However, more data is associated with higher costs, and the benefits of improved decision support should be balanced with the costs of its data needs. It is therefore of importance to evaluate sewer deterioration models with respect to the quality of predictions.

There exist numerous different sewer deterioration models which can utilise CC as input data to predict the condition distribution of a sewer network. However, most condition-based sewer deterioration models have some traits in common. Consider the situation where a sewer system has a set, S, containing N sewer pipes, for which a

subset R, containing n pipes, has been condition assessed. The time of installation for pipe i is $t_{i,\text{inst}}$. Each pipe j in R was observed at a time $t_{j,\text{insp}}$ with resulting CC y_j, where y_j can be in any one out of c CC labels. Lastly it is assumed that the vector Z_j contains information about the variability in the sewer CCs, and that Z_j can be used as explanatory factors in the model. A sewer deterioration model f will then utilise the observations R to predict the condition of any pipe in S at a time T, either in deterministic or probabilistic terms. This can be expressed in mathematical terms as:

$$S = \{Z_i, i = 1, 2 \ldots N\},$$
$$R = \{O_j, j = 1, 2, \ldots n\}, R \subset S$$
$$O_j = \{i, Z_j, t_{j,\text{insp}}, y_j\}, y_j \in \{1, 2, \ldots c\}$$
$$\Downarrow$$
$$\hat{y}_i(T) = f(Z_i, T|R) \vee \Pr(\hat{y}_i(T)) = f(Z_i, T|R)$$
$$\forall i \in S, \hat{y}_i \in \{1, 2, \ldots c\} \tag{1}$$

In Equation (1) each observation j is represented in O_j with the pipe identification i, the explanatory covariates of the pipe Z_j, the time of inspection $t_{j,\text{insp}}$ and the inspection result y_j. $\hat{y}_i(T)$ is the model's predicted CC for pipe i at time T.

Infrastructure deterioration models are commonly classified as *physical*, *statistical* or *artificial intelligence* models (Yang 2004). Physical models are based on mathematical relationships which give physical meaning (Savic *et al.* 2009), and should as such be a good source of information about the deterioration process. However, physical models have not been widely applied on sewer deterioration phenomena mainly due to the complexity of the physical factors that affect pipe deterioration (Rajani & Kleiner 2001; Tran 2007; Ana & Bauwens 2010), although examples like *ExtCorr* and *WATS internal corrosion model* (Vollertsen & König 2005) exist.

Statistical sewer deterioration models are based on relationships between factors that influence the deterioration process, where one or more of these factors are treated as random variables. Statistical inference and probabilistic terms can be used to interpret the results from statistical sewer deterioration models. A wide variety of statistical sewer deterioration models are described in the literature, from simple logistic (Ariaratnam *et al.* 2001) or multiple (Chughtai & Zayed 2008) regression techniques,

cohort survival models (Baur & Herz 2002), to more complex models.

A popular class of statistical models are the Markov chain models. The Markov chain has been an appealing choice for deterioration modelling because one can treat each CC as a Markov state, and easily apply restrictions to the Markov chain transition probabilities which are characteristic for the real deterioration process (Wirahadikusumah *et al.* 2001), and have therefore been developed in many different forms to model sewer deterioration (Micevski *et al.* 2002; Kleiner *et al.* 2004; 2006; Baik *et al.* 2006; Le Gat 2008; Scheidegger *et al.* 2011; Egger *et al.* 2013).

Artificial intelligence (AI) is '… the science and engineering of making intelligent machines, especially intelligent computer programs' (McCarthy 2007), often through mimicking human intelligence (Ana & Bauwens 2010) and reasoning. AI methods are often, though not always, 'black boxes' (Ana & Bauwens 2010), non-inferential, and prone to over-fitting (Savic *et al.* 2009). Examples of sewer deterioration modelling with AI methods include *neural networks* (Najafi & Kulandaivel 2005; Tran *et al.* 2006; Tran *et al.* 2009), *Bayesian networks* (Jung *et al.* 2012) and *random forests* (RF) (Jung *et al.* 2012; Harvey & McBean 2014).

Regardless of the model type, the assumption is that the combination data and sewer deterioration model should provide improved knowledge about the diagnosis of the sewers ($\hat{y}_i(T)$), and consequently improve the quality of the decision support in the IAM process, either on an individual pipe or a network level. The quality of the predictions determines the quality of the decision support. The need to assess the quality of water infrastructure deterioration model predictions has therefore been pointed out in the literature. After a comprehensive review of models for water main breaks (Kleiner & Rajani 2001; Rajani & Kleiner 2001), Kleiner and Rajani concluded that prediction uncertainty often was unknown and that more research was needed to validate model results. Further, Ana & Bauwens (2010) reviewed nine different physical, statistical and AI based sewer deterioration models, and concluded that wider acceptance and application of such models are dependent on proof that the prediction results are reliable. Caradot *et al.* (2014) state that there is a knowledge gap in terms of how the availability of CCTV observations affects the quality

of sewer condition predictions, and how many observations are needed.

There exist examples of sewer deterioration prediction quality assessments (e.g. Scheidegger *et al.* 2011), although only applied on synthetic (virtually generated) datasets; and it is uncertain to what extent such assessments apply for real-world networks. There is thus still a need to evaluate what effect the quality of sewer deterioration prediction accuracy has in the decision context it is used. This paper attempts to fill this knowledge gap by demonstrating and discussing up to what level and what kind of IAM decisions can be supported by sewer deterioration models, given the current data availability and model predictions in a real dataset.

RESEARCH METHOD

In the introduction it is indicated that structural condition often is a criterion which influences IAM decisions. Naturally, there will be other criteria which affect IAM decisions; however, the aim of this paper is to discuss the benefits of applying sewer deterioration models, therefore one may consider a decision context in which sewer condition is evaluated independently from other decision-influencing factors (such as the consequence of sewer failure or flooding risk). When considering the sewer condition in an IAM context, the sewer manager will be faced with some questions:

1. What is the distribution of CCs in the sewer network?
2. Which sewer pipes should be prioritised for renewal (or inspection)?

Condition assessments to answer the first question will from hereon be referred to as *network level condition assessments*, while the second will be referred to as *individual pipe level condition assessments*. In order to assess the first question, the utility manager can use the collected condition data, and apply a sewer deterioration model to assess the condition of sewers which have not been observed. If the condition assessment shows that the distribution of CCs in the network is acceptable, the utility manager does not have to invest resources on improving the condition of the network. If the opposite is the case, then the utility manager

must decide where to invest in condition-improving measures. A third option is that there is uncertainty about the distribution of conditions; if this is the case, one will have to invest resources in reducing the uncertainty (e.g. more inspections). If it is decided that it is necessary to implement condition-improving measures, the utility owner can choose to prioritise the sewer pipes which have already been condition assessed and found to be in a critical condition. However, a complimentary strategy could be to use the deterioration model results to prioritise the inspection of unobserved pipes, i.e. pipes which are predicted to be in a critical condition are prioritised for 'targeted' inspection.

Condition assessment data and deterioration models will be instrumental for answering the two condition-related questions. With more inspections and well-performing models, the utility should be better able to answer the questions. In this paper a methodology is applied, where the ability of two widely different sewer deterioration models to answer the two aforementioned condition-related questions is assessed by Monte Carlo simulations. In the following subsections, the models which have been applied, the methods and the data they have been applied on are presented.

Deterioration models investigated

In this paper two different models are applied. The first one, *GompitZ*, is a statistical model which is specifically developed to model the deterioration of sewers. The second one, RF, is a general-purpose machine-learning algorithm. These models will now be explained in more detail.

GompitZ

GompitZ is a sewer deterioration model which was developed under the research project CARE-S (Computer Aided REhabilitation of Sewer Networks) (Sægrov 2006), and is a Non-homogeneous Markov Chain model (Le Gat 2008). In a Markov Chain there are a finite number of states in which an element can be located, and the probability of making a transition from one state to another is only dependent on the current state (Grinstead & Snell 1997); the CC of a pipe is modelled as a Markov Chain state in GompitZ.

The Markov transition probabilities are calculated using Gompertz' distribution, where a pipe's decay rate is the sum of a time-dependent and a time-independent deterioration rate component (Makeham 1860). Both the time-dependent and the time-independent components in GompitZ may be modified with vectors of explanatory covariates ($\mathbf{Z_1}$ and $\mathbf{Z_0}$ in Equation (2), respectively). The user can therefore select covariates which affect both the initial condition as well as the deterioration rate (Vollertsen & König 2005; Le Gat 2008). The survival function S_{ik} for a pipe i, which expresses the probability that a pipe is in CC k or better (out of c possible classes) when its age is t_i, is written in Equation (2). From this expression, one can derive the conditional state probabilities and the marginal likelihood functions for each pipe. The regression parameters of GompitZ (θ) are determined by maximising the likelihood function. It is referred to Le Gat (2008) or Rokstad $et\ al.$ (2014) for more details about GompitZ calibration.

$$S_{ik}(t|u_i) = P\{Y(t) \leq k\}$$
$$= \exp\left[-\exp\left[\alpha_k + \mathbf{Z}_{0i}^T\boldsymbol{\beta}_0 + t_i \cdot \exp(\mathbf{Z}_{1i}^T\boldsymbol{\beta}_1 + u_i)\right]\right]$$
$$k = \{1, 2, \ldots, c - 1\}$$
$$S_{ic}(t|u_i) = 1$$
$$u_i \sim N(0, \sigma^2),\ \text{individual frailty factor}$$
$$\boldsymbol{\theta} = [\alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \boldsymbol{\beta}_0, \boldsymbol{\beta}_1] \tag{2}$$

The calibration parameters (θ) in GompitZ are then used to predict CC probabilities for each pipe and each year in a user-defined prediction period. If one describes GompitZ in the same manner as in Equation (1), one obtains Equation (3), where $t_{i,\text{inst}}$ is the year of installation of pipe i, and T is a time in the user-defined prediction horizon of GompitZ. One may note that the vector of explanatory factors ($\mathbf{Z}_i$), has been divided into a time-dependent ($\mathbf{Z_{1i}}$), and time-independent ($\mathbf{Z_{0i}}$) component.

$$S = \{\mathbf{Z}_{0i}, \mathbf{Z}_{1i}, t_{i,\text{inst}}, i = 1, 2 \ldots N\}$$
$$R = \{O_j, j = 1, 2, \ldots n\}, R \subset S$$
$$O_j = \{i, \mathbf{Z}_{0j}, \mathbf{Z}_{1j}, t_{j,\text{inst}}, t_{j,\text{insp}}, y_j\}$$
$$y \in \{1, 2, \ldots c\}$$
$$\Downarrow$$
$$\Pr(y_i(T) = k) = f_{\text{GompitZ}}(\mathbf{Z}_{0i}, \mathbf{Z}_{1i}, t_{i,\text{inst}}, T|R)$$
$$\forall i \in S, k \in \{1, 2, \ldots c\} \tag{3}$$

Random forest

A RF is an ensemble learning classification and regression algorithm which generates a number of decision trees (Safavian & Landgrebe 1991) with randomly selected covariates, which together make an aggregated classification prediction (Breiman 2001). Each tree is grown with a $bootstrapped$ sample of the calibration data, and $aggregated$ by letting the decision trees vote for the most popular predicted class for each instance. The $bootstrap$ $aggregation$ method reduces the variance of the predictions (Hastie $et\ al.$ 2009), and therefore increases the accuracy of the model. Each tree in the RF predicts a class membership for each evaluated instance, and the trees vote for the most popular class; the class which wins the election is the final prediction of RF. The votes can also be used to assess class membership probabilities. For instance, if 42 out of 500 trees voted that a sewer pipe was in CC 4, one could assign a CC 4 probability of 8.4%. Using the terminology from Equation (1) to express RF for sewer deterioration modelling purposes one obtains:

$$S = \{\mathbf{Z}_i, t_{i,\text{inst}}, i = 1, 2 \ldots N\}$$
$$R = \{O_j, j = 1, 2, \ldots n\}, R \subset S$$
$$O_j = \{i, \mathbf{Z}_j, y_j\}$$
$$y \in \{1, 2, \ldots c\}$$
$$\Downarrow$$
$$[\hat{y}_i,\ \Pr(y_i = k)] = f_{RF}(\mathbf{Z}_i|R) \forall i \in S, k \in \{1, 2, \ldots c\} \tag{4}$$

RF is not suitable for forecasting sewer conditions into the future, because it makes no assumptions about the deterioration process (as opposed to GompitZ). The time of inspection, time of installation, or age of inspection are therefore not explicitly considered in Equation (4), but can be included as an element in $\mathbf{Z}_i$, either explicitly as $t_{i,\text{inst}}$ and $t_{i,\text{insp}}$, or as age at inspection $t_{i,\text{insp}} - t_{i,\text{inst}}$. Unobserved pipes must then be assigned a reasonable value for $t_{i,\text{insp}}$, so that the assigned 'age at inspection' falls within the bulk of inspection ages for the observed pipes.

RF is not a problem-specific method, and has as such been applied to a multitude of classification and regression problems. However, some examples of RF being used in water infrastructure can be mentioned; Harvey & McBean (2014) used RF to predict the condition of individual sewer pipes, as did Jung $et\ al.$ (2012), while Wu $et\ al.$ (2013) used RF to classify defects from CCTV inspection footage.

Accuracy assessment methodology

A Monte Carlo approach has been applied to assess the accuracy of the model predictions of GompitZ and RF. First consider a dataset (S) in which all pipes have been observed once in the near past, see Equation (5). Based on this full set of observations, one can randomly draw a subset R_m of size n, see Equation (6).

$$S = \{O_i, i = 1, 2\ldots, N\}, O_i = \{\mathbf{Z}_i, t_{i,\text{inst}}, t_i, y_i\} \tag{5}$$

$$R_m = \{O_j, j = 1, 2\ldots, n\}, \quad R_m \subset S \tag{6}$$

The subset R_m can be used to calibrate the models, and predict the condition states for the full set of observations S. GompitZ yields predictions according to Equation (7), while RF yields the output in Equation (8) (RF returns both CC and votes).

$$\Pr_m(y_i(T) = k) = f_{\text{GompitZ}}\left(\mathbf{Z}_{0i}, \mathbf{Z}_{1i}, t_{i,\text{inst}}, T | R_m\right)$$
$$\forall i, k \in \{1, 2, \ldots c\} \tag{7}$$

$$[\hat{y}_i, \Pr_m(y_i = k)] = f_{RF}\left(\mathbf{Z}_i | R_m\right) \quad \forall i, k \in \{1, 2, \ldots c\} \tag{8}$$

By repeating the random subset selection, model calibration and prediction many times, one can record the variability in the model predictions, for a given size of n. Further, by performing this procedure for different sizes of n, one can assess how the variability is dependent on the amount of inspection data available.

The Matlab translation (Jaiantilal 2009) of the R implementation (Liaw & Wiener 2002) of RF and the GompitZ v2.08 application were used to complete the work presented in this paper.

Calculating the selection variance

If one has inspected all the pipes in a sewer system there will be no uncertainty about the distribution of CCs. However, if only a subset has been inspected, one can still estimate the proportion of each CC, but the estimate will be uncertain because the condition of every individual pipe is not known. If one has very reliable information about the factors which affect the CC probability distributions for the unobserved pipes (good covariates), then one will be able to predict the distribution of CC proportion accurately. If one has no additional (or unreliable) information, then the estimate will be less accurate. A 'baseline' accuracy level is the accuracy one has in the predicted distribution of conditions given that one has no additional information than the conditions one has observed. This level is denoted as the *selection variance*.

To calculate the selection variance when n out of N pipes have been inspected, let $\mathbf{x} = [x_1, x_2, \ldots, x_c]$, where x_k is the number of pipes observed in CC k, and $\mathbf{m} = [m_1, m_2, \ldots, m_c]$, where m_k is the total number of pipes in CC k in the full set of N pipes. An estimate for $\mathbf{m}$, and the corresponding proportion of the CCs $\mathbf{p}$, can then be calculated according to Equation (9)

$$\hat{m}_k = \frac{N}{n} x_k \Rightarrow \hat{p}_k = \frac{\hat{m}_k}{N} \tag{9}$$

By recording the value of $\mathbf{p}$ for each Monte Carlo repetition, one can assess the selection variance of the condition proportions for each subset size, $\text{var}(p_k \mid n)$. It is interesting to compare the selection variance with the accuracy assessment results from GompitZ and RF, because the comparison provides an indication of the performance of the model. If the accuracy is equivalent to the selection variance, then the model predictions are no better than randomly drawing samples and estimating the distribution of conditions based on these (like predicting the distribution of the colours of the remaining balls in an urn based on the colours of a subset of balls drawn from the urn). However, if there are predictors in the model which can account for the variability, then one would expect that the prediction variance would be low compared to the selection variance; the greater the discriminatory power of the covariates, the lower variance one would expect.

It is possible to estimate the selection variance when one has only observed a subset of CCs by performing Markov Chain Monte Carlo simulations on a multi-variate hyper-geometric distribution. An outline of how this can be done is presented in Rokstad *et al.* (2014).

Calculating the inspection efficiency

Given that one has inspected a subset of pipes in a network, R_m, one can use this subset of observations to calibrate a deterioration model (Equation (8)) and obtain predicted CC probabilities for each unobserved sewer pipe. Sorting these pipes according to their probability of being in a critical condition, and ranking forthcoming inspections by this probability would be the most efficient utilisation of the observations in R_m, given the objective to identify as many sewers in a critical CC as possible with the least possible inspections. The efficiency of this process can be assessed by comparing the predicted critical CC probability with the observed conditions of the sewer pipes in the following way:

1. Start with a randomised subset R_m, calibrate the deterioration model, and predict CC probabilities for all pipes which are not in the subset R_m. Record the length inspected, L_0, and the length of critical sewers detected, L_{c0}.
2. Consider the sewer pipe i with the highest critical CC probability, and record its length L_i (pipe i should not be member of R_m).
3. If the utility owner inspects this pipe, one will have to add its length to the total inspected length ($L_0 \leftarrow L_0 + L_i$).
4. If the sewer pipe actually is in the critical CC, then the amount of critical sewer identified is increased ($L_{c0} \leftarrow L_{c0} + L_i$) or it is not increased.
5. Repeat steps 2 to 4 consecutively for all sewer pipes from highest to lowest probability of belonging to the most critical CC.

By repeating this process for several randomised calibration subsets, one can consider the average efficiency of the model predictions for a given calibration subset size. By plotting the amount of critical sewers identified as a function of the total length inspected, one can visually assess the efficiency of the process. A steep curve indicates that the model predictions are good, and that the deterioration model is capable of identifying critical sewers, while a less steep curve indicates that the deterioration model has a lower performance.

When prioritising sewers for inspection in a real planning situation, one would typically not only use the predicted CC probabilities to determine which sewers to inspect, but also additional criteria, such as sewer criticality and potential for coordination with other activities. In a sewer inspection programme, one will often inspect several segments which are in proximity to each other, in order to reduce equipment and crew mobilisation costs. In order to make a practical inspection plan one will need to combine the CC probability predictions with other criteria. In this paper however, only the CC probabilities are used, as the scope of the paper is to investigate the performance of deterioration models.

Description of data

The data that have been considered in this case study are from Oslo VAV (Oslo municipality, Norway). In total 12,003 CCTV condition assessments were considered in the models, amounting to a total network length of 499 km or 27% of the complete Oslo VAV sewer network. The median year of installation for these pipes is 1956. The dataset was divided into four strata, as indicated in Table 1. All CCTV inspections in this dataset were conducted between 2002 and 2012, and approximately 85% of the inspections were carried out in the period 2008–2012. The CCTV inspections have been evaluated and coded into condition grades from 1 to 5 (best to worst), according to the standardised Norwegian sewer condition classification system in NVR 150/2007 (Bernhus et al. 2007).

The four strata were calibrated in the research project *Secured and Monitored Service from Oslo VAV (SMS)*,

Table 1 │ Dataset properties

Stratum description	# pipes	CC1 (%)	CC2 (%)	CC3 (%)	CC4 (%)	CC5 (%)
Concrete pipes (<Ø600 mm)	3,783	29	11	21	16	23
Concrete culverts (≥Ø600 mm)	564	71	9	9	3	7
Other pipe materials (<Ø600 mm)	6,989	45	9	16	11	18
Other culvert materials (≥Ø600 mm)	667	71	9	9	3	8
Total	12,003	43	10	17	12	19

where the goal was to use condition monitoring efforts as an aid for rehabilitation planning (Ugarelli *et al.* 2013). The following covariates were used in the calibration:

- Pipe diameter
- Type of effluent (storm water, foul water, combined)
- Construction period (1850–1929, 1930–1945, 1946–1969, 1970–2011)
- Road traffic
- Type of bedding soil
- Presence of trees

(Ugarelli *et al.* 2013). All of the aforementioned covariates have been known to show a significant impact on sewer deterioration in other studies in the literature (Davies *et al.* 2001; Chughtai & Zayed 2008; Ana *et al.* 2009; Ana & Bauwens 2010). When GompitZ was calibrated for the dataset (Ugarelli *et al.* 2013), only significant covariates (up to a significance level of 2.5%) were used (by χ^2-test, see Le Gat 2008). Only significant covariates have been used further in this paper.

RESULTS

The assessments were performed on a network level, where the ability to predict the proportions of sewers in each CC on the network as a whole was evaluated, and on an individual pipe level, where the ability to identify individual pipes in a critical CC was evaluated.

Network level condition assessment

The accuracy assessment method has been applied for subset sizes $n/N = 10, 20, \ldots, 90\%$ for the case study data, with 1,000 repetitions for each subset size. Descriptive statistics, such as expected values, variance etc., have been recorded for each subset size. Figure 1 shows the results for the two larger datasets. One may observe that there are considerable uncertainties in the predictions when a lesser extent of the sewer network has been observed. Even when 40% of the network has been inspected, there are
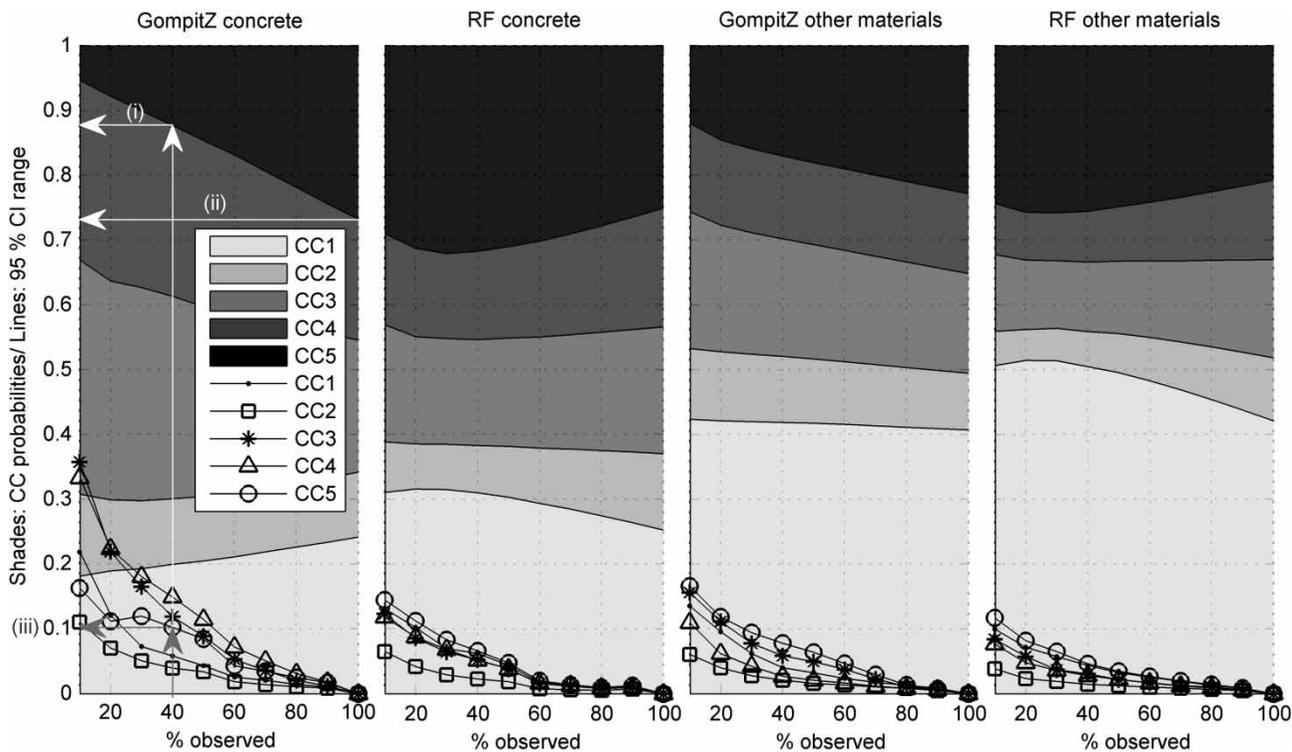

Figure 1 | CC probabilities as a function of inspection rate (shades) and 95% confidence interval range (lines). The arrows in the left plot (concrete stratum with GompitZ) shows how to read the plot for a pre-defined calibration subset size; from this one can see that if one inspects 40% of the concrete pipes and applies GompitZ, one will on average estimate that the percentage in CC5 is approximately 13% (value (i)), when it in reality is 26% (value (ii)). The 95% confidence interval for the 40% inspection will be around 10.3% (value (iii)). From the plots one can see that depending on the subset size and the model applied, one will have various degrees of prediction accuracy and bias.

still significant uncertainties in the CC proportions. One may also notice that the uncertainty of the larger dataset (other materials) is in general smaller than the smaller (concrete) dataset, and that RF in general performs better than GompitZ in terms of uncertainty. Both the GompitZ and RF predictions show clear biases as the amount of calibration data is reduced. The RF predictions show signs of underestimating the classes on few occasions (CC2 and CC4); this is a typical problem when applying tree-structured classifiers, and can in principle be solved by adjusting the cut-off values in the model (Harvey & McBean 2014). The GompitZ predictions display more severe bias tendencies, but in a different form – the bias in GompitZ is related to the definition of the individual frailty factor u_i, which is not symmetrical around zero in the survival function. Figure 2 shows the standard deviations from the concrete stratum simulations with GompitZ and RF, compared to the selection deviation. From these figures one can see that both GompitZ and RF generally have variances which exceed the selection variance. This means that the predictions from the models have greater uncertainty than the non-informed inference one could make based solely on the distribution of the raw CC observations.

All RF predictions presented in this paper have been based on the probabilistic output (votes) from the model, as it was discovered that the votes in general performed better in terms of uncertainty and bias. RF would generally be outperformed by GompitZ, if the absolute CC predictions were used.

The implications of the uncertainty assessments can be interpreted in the decision context in which they are used. For instance, if the goal for the concrete stratum were to keep the percentage of sewers in CC5 below 10%, one can see that if one had observed 40% of the network, and applied GompitZ, one would on average estimate the CC5 percentage to be 13%, when it in reality is over 26% (the values are read from the left plot of Figure 1). However, if one applied RF, one would overestimate this percentage. Further, if one calculated the 95% confidence interval of the selection variance, one would vastly underestimate the uncertainty of the predictions (see Figure 2). There is thus a great potential for making the wrong decisions based on the model predictions, and the benefit of inspection data is not enhanced by applying the deterioration model. Accurate network-level predictions with the current covariates can hence only be improved by collecting an adequate amount of inspection data.

For GompitZ one could have used the statistical significance levels for the covariates to evaluate the quality of the model. If the recorded p-values for a calibration subset size were found to be sufficiently and consistently low (e.g. $p < 0.05$), one could conclude that the model was of good quality, and if the recorded p-values were not sufficiently low or inconsistent, one could conclude that the model was not of satisfactory quality. However, the p-values do not reflect the quality of the model predictions with respect to the condition-related questions evaluated in this paper. This can be illustrated by considering Figure 2; as the
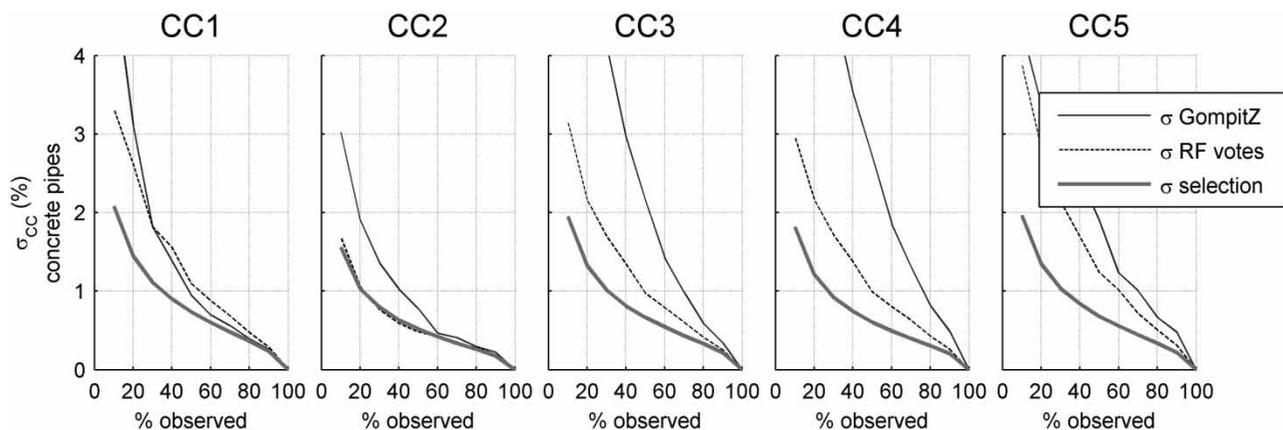

Figure 2 │ Standard deviations for the CC proportion predictions for GompitZ and RF for the *concrete stratum*, compared to the selection variance. The plots show that the prediction accuracies of the models in general are lower than the selection variances, which implies that CC distribution predictions will be more precise if based solely on raw CC observations, compared to the model predictions.

calibration subset size approaches 100%, the proportion of significant covariates will approach *1* (since the original dataset only contains significant covariates; see *Description of data* section), the variance in the predictions are nevertheless still consistently higher than the selection variance as the subset size approaches 100%, and the quality of the predictions are hence not reflected by the observation of consistently significant covariates.

One can investigate in more detail how the model performs for a certain calibration situation by considering the confusion matrix. Table 2 shows the confusion matrix for GompitZ for the 40% calibration subset example. From this table one can observe that the underestimation of CC5 is mostly caused by the fact that GompitZ predicts a proportion of the pipes as CC3 or CC4 when they in fact are CC5. The table further shows that there is a tendency towards the mean through overestimation of instances in CC3; many pipes from all other CCs are predicted as CC3. Considering the goal of identifying pipes in CC5, one can use the specificity and sensitivity measures to assess the predictive power of the model. GompitZ' specificity with respect to CC5 is in this case 40.5%, while the sensitivity is 97.6%. Thus, a pipe which in reality is in CC5 would therefore be classified as CC5 with 40.5% probability, and the probability of classifying a pipe which is not CC5 as CC5 is 2.4% (100 – 97.6%).

Individual pipe level condition assessment

In order to identify as many individual sewer pipes in a critical condition as possible, one could choose different inspection strategies, for instance *random*, *age-based*, or to inspect a part of the network, calibrate a deterioration model, and use the predictions to rank inspections by critical CC probability (*targeted*). The age-based inspection strategy is characterised by starting with the oldest pipe, and consecutively inspecting from oldest to newer. To evaluate the different strategies, the notion of *inspection efficiency* has been introduced and presented in *inspection efficiency diagrams*.

Figure 3(a) shows the inspection efficiency for GompitZ and RF for calibration subset sizes from 10 to 90%, compared with a random and an age-based inspection strategy (straight diagonal line). To interpret Figure 3(a), consider that the utility manager wants to identify, e.g. 50% of the concrete pipes which are in CC5. If a random inspection regime was chosen, one would then on average have to inspect 50% of all the concrete pipes. From the age-based inspection line, one may see that there is no strong ageing (purely as a function of time) phenomenon for the concrete pipes, and that one therefore would have to inspect almost as much as with a random strategy (47.9%). However, if one had inspected 10% of the network at random, and used this information to calibrate GompitZ or RF, one would only need to inspect 37.2% of the network in order to identify 50% of the pipes in CC5. Similar graphs are shown for the other strata in Oslo, and similar inspection efficiency numbers are summarised in Table 3. For the culvert datasets (Figure 3(c) and 3(d)) one may see that the inspection efficiency is better than for the concrete pipes stratum. For the *other material pipes* stratum (see Figure 3(b) and Table 3 (third column)), one may see that there is a strong ageing phenomenon, pipes which are older have thus a higher probability of being in CC5. With the same

Table 2 │ The table shows the averaged confusion matrix for GompitZ predictions for the concrete pipes based on a 40% calibration subset of the concrete pipe observations

| | | Predicted | | | | | |
		CC1	CC2	CC3	CC4	CC5	Total
Observed	CC1	610.1	113.3	233.4	120.5	13.7	1091
	CC2	67.2	148.3	131.8	71.0	7.7	426
	CC3	104.9	83.6	434.9	151.6	14.1	789
	CC4	56.2	36.8	176.9	317.8	33.3	621
	CC5	53.8	39.7	199.1	217.3	346.1	856
	Total	892.14	421.75	1176.00	878.16	414.95	3783

The matrix shows that pipes which are in CC5 very often are predicted as CC3 or CC4, causing a gross underestimation of the CC5 proportion. One may further observe that there is a tendency towards the mean; many of the pipes which are not in CC3 are predicted as CC3, and the proportion of pipes in CC3 is hence overestimated.
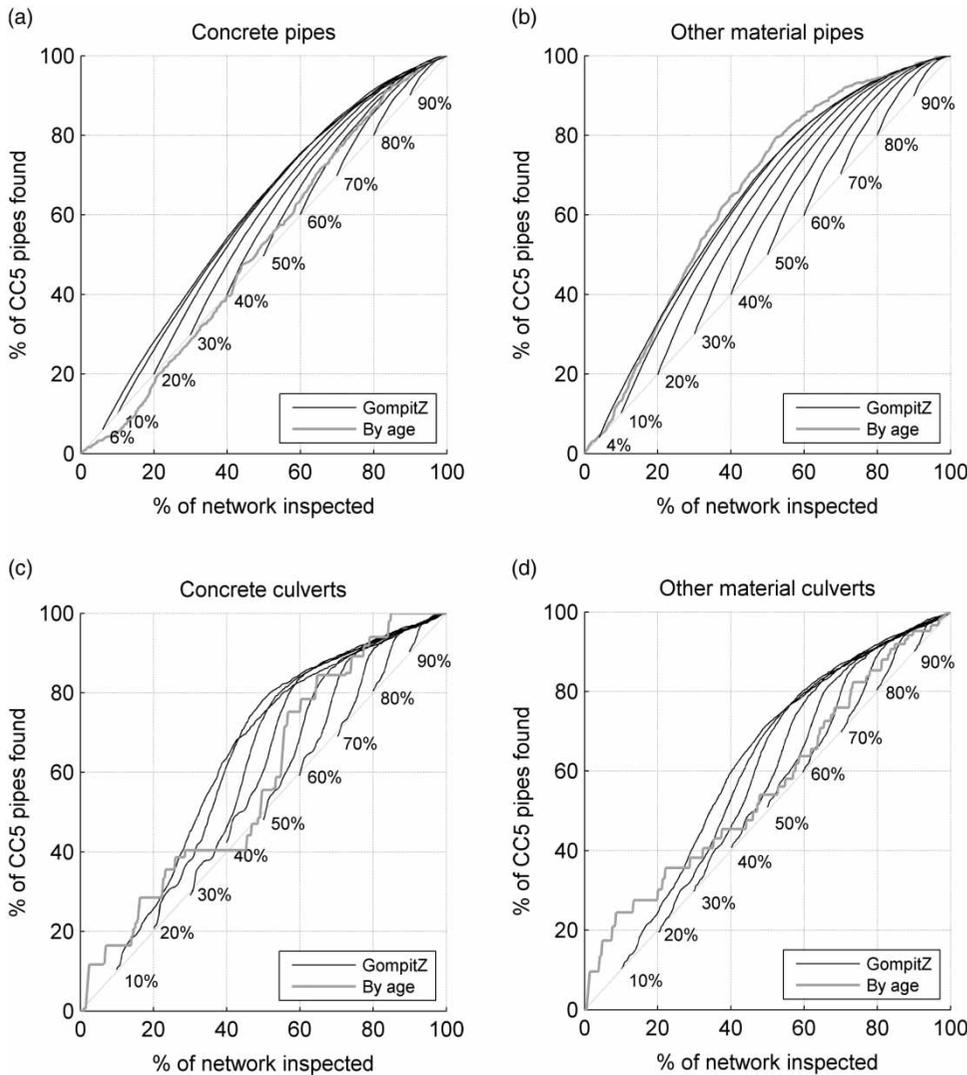
Figure 3 │ Inspection efficiencies for GompitZ as a function of calibration data subset size, for the four strata: concrete pipes (a), non-concrete pipes (b), concrete culverts (c), and non-concrete culverts (d). The graphs show how efficiently one could identify sewer pipes in CC5 by calibrating a model, and using the model predictions to target inspections. The higher the lift of the curve, the better the model is at predicting the occurrence of CC5 pipes. Only results for GompitZ are shown. For the two larger datasets ((a) and (b)), GompitZ and RF performed very similarly, while GompitZ outperformed RF for the two smaller datasets ((c) and 3(d)).

Table 3 │ Relative inspection length (% of stratum length) necessary to identify 50% of the pipes in CC5 for each stratum, either by ranking the pipes by age and inspecting from oldest to newest, or by prioritising inspections by GompitZ prediction probabilities of being in CC5 (using 10% subset size for calibration)

Stratum	Concrete pipes (%)	Other materials (%)	Concrete culverts (%)	Other material culverts (%)
Ranking by age	47.9	30.1	49.2	47.1
Ranking by GompitZ prediction	37.2	32.7	32.4	34.4

The table values are equivalent of reading Figure 3 at 50% on the vertical axes.

goal as for the concrete pipes, one may see that the most effective strategy would be to inspect by age, as the models are not able to adequately reproduce the strong correlation between pipe age and probability of being in CC5 which one observes when ranking the pipes by age.

To evaluate and compare the performance of each calibration subset size, the area under the inspection efficiency curve has been calculated. With a completely random inspection regime, one would have an area of 50% under the curve. If one were omniscient and knew a priori which

pipes were in CC5, one would have an area of 100% under the curve. The best-performing curve for the concrete pipe stratum was obtained by using a subset of 6% of the calibration data, and this yielded an area of 59.7% under the curve, whereas the area under the age-based inspection curve was only 51.7%. The results of this calculation are presented in Table 4. The only stratum which is not performing better than the age-based inspection strategy is the *other materials stratum*, where the area under the age-based curve is 65.5% (which is also the highest of the strata).

GompitZ and RF behaved very similarly in terms of the inspection efficiency performance for the two pipe strata, in fact so similarly that the lines almost perfectly overlapped. This indicates that the analytical form of the model is not of great importance for the purpose of ranking inspection candidate pipes, with the current data quality. However, GompitZ outperformed RF on individual pipe predictions for the two smaller (culvert) datasets.

DISCUSSION

The results section showed results from the evaluation of the performance of two sewer deterioration models, in terms of their ability to aid condition-based sewer IAM decision-making. The results from the network level predictions showed that the benefit of sewer inspection data is not enhanced by the application of a sewer deterioration model, given the current data in Oslo, and that the inaccuracy in the predictions is increased, rather than reduced. With the current data quality one would achieve predictions of greater accuracy by estimating the distribution of conditions solely based on the raw observations. The amount of inspections needed could be determined by assessing the selection variance, and choosing an inspection rate

which is in proportion with the accuracy one needs to assess the gap between condition distribution and target values. An important question to ask is why the network level model predictions are not better. The performance of sewer deterioration predictions is inhibited by several factors, among them are:

- The CCTV inspection CC is an aggregation of several classification codes, all of which are determined by the inspection operator, according to interpretation of the classification coding system. Some of the classification codes are easy to interpret based on the footage, and are consequently reported quite consistently, while others are more difficult to interpret, and are reported less consistently (Dirksen *et al.* 2011). The subjective nature of the classification process, and the potential for misinterpreting footage, makes CCTV inspection data prone to uncertainty and inconsistency.
- The different classification codes which are aggregated in the overall CC have different failure modes. The Norwegian classification system contains classification codes such as *corrosion*, *presence of sediments*, *root intrusion* and *product error*, to mention some. There are different failure modes behind these classification codes, and they are often explained by different covariates. For instance, *root intrusion* is exclusively dependent on *local conditions* (such as presence of trees, depth of sewer etc.), *product error* is solely dependent on *external conditions* (quality of sewer manufacturer), while *sediment* build-up is dependent both on pipe characteristics (slope, roughness) as well as *hydraulic conditions* in the sewer system. When one only accounts for some of the failure modes which can contribute to the CC, one will perceive some of the instances in the dataset as heterogeneous.
- The covariates which have been used for the Oslo case study do not directly explain the physical failure modes of the sewer pipes. Covariates such as *pipe diameter*, *type of effluent*, *traffic level*, and *presence of trees* are all at best indirect indicators (or risk factors) of sewer failure modes. However, when considering the failure modes such as *corrosion*, one would benefit from explanatory covariates such as *corrosive gas production*, while *cracks* and deformation would benefit from covariates which reflect *mechanical load*.

Table 4 │ Calculation of the area under the inspection efficiency curve

Stratum	Concrete pipes (%)	Other materials (%)	Concrete culverts (%)	Other material culverts (%)
Area under age-curve	51.7	65.5	58.5	56.3
Area under the best curve	59.7	64.3	63.0	60.9
@ subset size	6	4	10	10

- The different failure modes which aggregate to an overall CC are by nature *stochastic phenomena*, some of which are more predictable than others, depending on the explanatory factors available.

One could also ask if the model assumptions about the deterioration process are appropriate. In this paper two different models were applied, namely GompitZ and RF. GompitZ has a strict analytical form, which imposes an assumption about factors which account for initial condition and condition deterioration rate for the sewer pipes, while RF is a data-driven method with less assumptions. In the results from the network-level predictions, it was shown that RF outperformed GompitZ with respect to predictive accuracy. The assumptions about the deterioration process in GompitZ do therefore not seem to be appropriate for the Oslo data. However, one cannot conclude that the assumptions about the deterioration process are wrong. Given the quality of the input data, i.e. the adequacy of the explanatory factors, the aggregation of different failure modes, the subjective classification coding, and the inherent stochastic nature of some of the failure modes, it is likely that the vague trends in the dataset will be dominated by the analytical form of the model, and that the effect of the 'weak' covariates will be over- or underestimated, which results in an excessively high prediction variance. It may well be that if one applies the assumptions of, e.g. GompitZ on specific failure modes, with adequate covariates to explain each failure mode, one would be able to make predictions of higher quality, than with the current paradigm of predicting aggregated CCs.

Pollert *et al.* (2005) demonstrated by experiment and hydraulic simulations that the occurrence of different sewer pipe failure modes affects the hydraulic capacity of pipes in different manners, and suggested individual formulas for calculating the headloss resulting from three of the failure modes which are commonly observed in sewer pipes (*displaced pipes*, *obstacles* and *roots*). The application of failure mode-specific deterioration models may thus have benefits not only with respect to being able to more accurately predict the condition, but also for modelling the consequence of deterioration in terms of the operational performance (e.g. capacity or flooding probability) of the system.

Under the description of the data, it was mentioned that all covariates were found to be significant in the initial calibration of GompitZ. The findings in this paper show that even though one has significant covariates, it is not guaranteed that the model will provide predictions with an accuracy that is appropriate for the measures one wants to evaluate.

Despite the limited success of predicting CC distributions on a network level, it has been shown that the use of sewer deterioration models can be useful on an individual pipe level, even with the current data availability and quality in Oslo. The utilisation of inspection data of a subset from the sewers can aid the utility managers in ranking sewer pipes by failure likelihood, and identify sewers in a critical condition more effectively than a random or age-based inspection regime. The benefit of applying a sewer deterioration model is manifested in a higher detection rate for critical sewer pipes.

The predictions made for the purpose of scheduling uninspected sewers for inspections are less sensitive to the data quality, because they are based on ranking the predictions by probability of being in the worst CC, and in spite of the fact that the effect of the explanatory factors can be over- or underestimated in the model, the ranking will still consistently rank pipes which are more likely to be in a critical condition higher. Thus, although neither GompitZ nor RF are able to accurately predict the condition of unobserved sewers, both models are capable of detecting trends in the dataset which can be used to differentiate the likelihood of being in a critical condition.

The inspection efficiency considerations for the Oslo datasets showed that the predictions from GompitZ and RF are not necessarily improved by adding more observations, with respect to the inspection efficiency; this fact reflects that the covariates used are *indirect risk factors* rather than factors which account for *actual failure modes*. By removing the explanatory covariates, one can show that the performance of the models will be lower, and improving the sewer deterioration model performance for the purpose of prioritising inspections should therefore benefit more from obtaining covariates with better explanatory power, compared to obtaining more CC observations. Better understanding of which factors can be used to describe the physical deterioration phenomena (both with

respect to deterioration phenomena in general as well as local conditions), and understanding how these can be assessed (directly or indirectly) in a cost-effective manner should improve the efficiency to a greater extent than obtaining more inspections. It may be necessary to move away from the paradigm of predicting aggregated CCs based on CCs as input data, and rather make predictions based on the actual observation codes (such as *corrosion*, *cracks*, *roots* etc.) in the inspection reports. There exist examples of models which use sewer failure modes to predict sewer failures, for instance sediment and blockage (Rodríguez *et al.* 2012), exfiltration/infiltration (Desilva *et al.* 2005), and corrosion (Vollertsen & König 2005) prediction models, which could be considered as alternatives to the current application of GompitZ or RF. There also exist studies showing correlations between failure modes and explanatory factors, such as tree types and root intrusion (Östberg *et al.* 2012), which in principle could be used to develop new models for sewer condition prediction.

The example of quantity versus quality is well illustrated by the *other material pipes* stratum, which did not perform better than an age-based inspection strategy. Not only have different failure modes been mixed in this dataset, but also different materials, which potentially react differently to the covariates, have been mixed (stratified); the act of aggregating data to obtain big datasets only makes sense when the explanatory covariates are adequate, and the stratification of the data does not result in an non-homogeneous dataset. Finer stratification (by material, soil type etc.) could lead to better predictions from the *other material pipes* stratum.

GompitZ has, as opposed to RF, the possibility of forecasting the distribution of conditions into the future, since GompitZ assumes a time-dependent form in the survival function. Based on the findings about the ability to predict the CC distribution in the present, one should be careful when drawing conclusions about the deterioration rate of the system from GompitZ forecasts. When there is lack of evidence that the analytical form of a sewer deterioration model fits the deterioration rate phenomenon, one should at the very least substantiate the deterioration forecasts with repeated inspections on the subset which was used for calibrating the forecasting model. RF performed better than GompitZ in terms of network level prediction accuracy, and unless there is evidence that the assumptions

about deterioration rate in GompitZ are true, one should apply the model with the predictive accuracy and least severe bias.

CONCLUSIONS

This paper has demonstrated an assessment of the benefits of applying sewer deterioration models for IAM planning, as a function of the proportion of the network which has been inspected. The assessment has been performed by considering the prediction accuracy and efficiency with respect to questions related to condition assessment which frequently arise in the IAM planning process at network and single pipe level. At a network level it was considered how the accuracy of predicting the CC proportions, and at an individual pipe level it was considered how the efficiency of identifying critical condition sewers, is affected by applying a deterioration model. The assessments have been based on inspection data from Oslo VAV in Norway.

The results showed that the network level predictions in most cases had less accuracy than the selection variance baseline, which implies that the model predications have lower accuracy than an uninformed estimate of the distribution of conditions (the selection variance). The benefit of applying a sewer deterioration model did hence not manifest itself as improved accuracy. The low quality of the predictions can be ascribed to the classification aggregation and the lack of explicitly considering specific failure modes, the lack of appropriate covariates, and possible inadequacy of the model assumptions.

The results further showed that although the predictions in the overall condition distribution of the sewer network were of low quality, they can still be useful in identifying individual sewer pipes of a certain CC. The benefit of applying a sewer deterioration model does in this context manifest itself in improved understanding about which sewers are in a critical CC, and the ability to detect these with fewer inspections.

The current paradigm, in which CCs are predicted based on CC calibration data, does not utilise knowledge or data to their full extent. The CC is an aggregation of classification codes, which all reflect different failure modes. By attempting to mix all the different classification codes, one

assumes that the physical phenomena can be expressed by the explanatory covariates, through the aggregated CC. Although this may be partly true, it is likely that heterogeneity and random effects will dominate the observations when the explanatory power of the covariates is weak. In order to improve predictions, one must more carefully consider the different failure modes which contribute to the CC, and collect the data which is believed to affect each of them.

ACKNOWLEDGEMENTS

REFERENCES

Alegre, H. & Matos, R. 2009 Paving the way for a sustainable asset management of urban water infrastructures: outcomes from the 5th World Water Forum. *Water Asset Management International* **5** (2), 2.

Ana, E. V. & Bauwens, W. 2010 Modeling the structural deterioration of urban drainage pipes: the state-of-the-art in statistical methods. *Urban Water Journal* **7** (1), 47–59.

Ana, E., Bauwens, W., Pessemier, M., Thoeye, C., Smolders, S., Boonen, I. & De Gueldre, G. 2009 An investigation of the factors influencing sewer structural deterioration. *Urban Water Journal* **6** (4), 303–312.

Ariaratnam, S. T., El-Assaly, A. & Yang, Y. 2001 Assessment of infrastructure inspection needs using logistic models. *Journal of Infrastructure Systems* **7** (4), 160–165.

Baik, H.-S., Jeong, H. S. D. & Abraham, D. M. 2006 Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems. *Journal of Water Resources Planning and Management* **132** (1), 15–24.

Baur, R. & Herz, R. 2002 Selective inspection planning with ageing forecast for sewer types. *Water Science & Technology* **46** (6–7), 389–396.

Bernhus, C., Nilssen, O., Mosevoll, G. & Hansen, A. 2007 *R150 – Dataflyt – Klassifisering av Avløpsledninger*. Report of Norsk Vann. Norsk Vann, Hamar.

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

Caradot, N., Sonnenberg, H., Kropp, I., Schmidt, T., Ringe, A., Denhez, S., Hartmann, A. & Rouault, P. 2014 Sewer deterioration modeling for asset management strategies. *Water Asset Management International* **10** (3), 3–8.

Chughtai, F. Z. & Zayed, T. 2008 Infrastructure condition prediction models for sustainable sewer pipelines. *Journal of Performance of Constructed Facilities* **22** (5), 333–341.

Davies, J. P., Clarke, B. A., Whiter, J. T. & Cunningham, R. J. 2001 Factors influencing the structural deterioration and collapse of rigid pipes. *Urban Water* **3** (1–2), 73–89.

Desilva, D., Burn, S., Tjandraatmadja, G., Moglia, M., Davis, P., Wolf, L., Held, I., Vollertsen, J., Williams, W. & Hafskjold, L. 2005 Sustainable management of leakage from wastewater pipelines. *Water Science & Technology* **52** (12), 189–198.

Dirksen, J., Clemens, F. H. L. R., Korving, H., Cherqui, F., Le Gauffre, P., Ertl, T., Plihal, H., Müller, K. & Snaterse, C. T. M. 2011 The consistency of visual sewer inspection data. *Structure and Infrastructure Engineering* **9** (3), 214–228.

Egger, C., Scheidegger, A., Reichert, P. & Maurer, M. 2013 Sewer deterioration modeling with condition data lacking historical records. *Water Research* **47** (17), 6762–6779.

Grinstead, C. C. M. & Snell, J. L. 1997 *Introduction to Probability*, 2 edn. American Mathematical Society, Providence, Rhode Island.

Harvey, R. R. & Mcbean, E. A. 2014 Predicting the structural condition of individual sanitary sewer pipes with random forests. *Canadian Journal of Civil Engineering* **41** (4), 294–303.

Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J. & Tibshirani, R. 2009 *The Elements of Statistical Learning*. Springer, New York.

Jaiantilal, A. 2009 Random Forest (Regression, Classification and Clustering) Implementation for Matlab (and Standalone). Google Code.

Jung, I.-S., Garrett, J. H. J., Soibelman, L. & Lipkin, K. 2012 Application of classification models and spatial clustering analysis to a sewage collection system of a mid-sized city. R. Issa & I. Flood (eds) *Computing in Civil Engineering: Proceedings of the 2012 ASCE International Conference on Computing in Civil Engineering*, June 17–20, 2012, Clearwater Beach, Florida, pp. 537–544.

Kleiner, Y. & Rajani, B. 2001 Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* **3** (3), 131–150.

Kleiner, Y., Sadiq, R. & Rajani, B. 2004 Modeling failure risk in buried pipes using fuzzy Markov deterioration process. In: *Pipeline Engineering and Construction* (J. Galleher & M. Stift Jr.). American Society of Civil Engineers, Reston, Virginia, pp. 1–12.

Kleiner, Y., Sadiq, R. & Rajani, B. 2006 Modelling the deterioration of buried infrastructure as a fuzzy Markov process. *Journal of Water Supply: Research and Technology – AQUA* **55** (2), 67–80.

Kley, G., Kropp, I., Schmidt, T. & Caradot, N. 2013 *Review of Available Technologies and Methodologies for Sewer*

Condition Evaluation. Report of Kompetenzzentrum Wasser Berlin gGmbH (KWB). KWB, Berlin.

Le Gat, Y. 2008 Modelling the deterioration process of drainage pipelines. *Urban Water Journal* **5** (2), 10.

Le Gauffre, P. J., Claude, J., Vasconcelos, E., Breysse, D., Gibello, C. & Desmulliez, J.-J. 2007 Performance indicators and multicriteria decision support for sewer asset management. *Journal of Infrastructure Systems* **13** (2), 105–114.

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

Makeham, W. M. 1860 On the law of mortality and the construction of annuity tables. *The Assurance Magazine, and Journal of the Institute of Actuaries*, **8** (6), 301–310.

Mccarthy, J. 2007 What is Artificial Intelligence. [Online] URL: http://www.formal.stanford. edu/jmc/whatisai.html (accessed 10 October 2014).

Micevski, T., Kuczera, G. & Coombes, P. 2002 Markov model for storm water pipe deterioration. *Journal of Infrastructure Systems* **8** (2), 49–56.

Najafi, M. & Kulandaivel, G. 2005 Pipeline condition prediction using neural network models. In: *Pipelines 2005@ Optimizing Pipeline Design, Operations, and Maintenance in Today's Economy*. ASCE, *Houston, Texas, USA*, pp. 767–781.

Östberg, J., Martinsson, M., Stål, Ö. & Fransson, A.-M. 2012 Risk of root intrusion by tree and shrub species into sewer pipes in Swedish urban areas. *Urban Forestry & Urban Greening* **11** (1), 65–71.

Pollert, J., Ugarelli, R., Saegrov, S., Schilling, W. & Federico, V. 2005 The hydraulic capacity of deteriorating sewer systems. *Water Science & Technology* **52** (12), 207–214.

Rajani, B. & Kleiner, Y. 2001 Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water* **3** (3), 151–164.

Rodríguez, J. P., Mcintyre, N., Díaz-Granados, M. & Maksimović, Č. 2012 A database and model to support proactive management of sediment-related sewer blockages. *Water Research* **46** (15), 4571–4586.

Rokstad, M. M., Le Gat, Y. & Ugarelli, R. M. 2014 On the uncertainty of probabilistic sewer deterioration forecast modelling. In: *11th International Conference on Hydroinformatics*, *New York City, USA*.

Sægrov, S. 2006 *CARE-S Computer Aided Rehabilitation for Water and Storm Water Networks*. IWA Publishing, London, UK.

Safavian, S. R. & Landgrebe, D. 1991 A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics* **21** (3), 660–674.

Savic, D. A., Giustolisi, O. & Laucelli, D. 2009 Asset deterioration analysis using multi-utility data and multi-objective data mining. *Journal of Hydroinformatics* **11** (3), 14.

Scheidegger, A., Hug, T., Rieckermann, J. & Maurer, M. 2011 Network condition simulator for benchmarking sewer deterioration models. *Water Research* **45** (16), 4983–4994.

Tran, H. D. 2007 Investigation of Deterioration Models for Stormwater Pipe Systems. PhD Thesis, Victoria University.

Tran, D., Ng, A., Perera, B., Burn, S. & Davis, P. 2006 Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. *Urban Water Journal* **3** (3), 175–184.

Tran, H., Perera, B. & Ng, A. 2009 Predicting structural deterioration condition of individual storm-water pipes using probabilistic neural networks and multiple logistic regression models. *Journal of Water Resources Planning and Management* **135** (6), 553–557.

Ugarelli, R., Venkatesh, G., Brattebo, H., Federico, V. D. & Saegrov, S. 2010 Asset management for urban wastewater pipeline networks. *Journal of Infrastructure Systems* **16** (2), 112–121.

Ugarelli, R. M., Selseth, I., Le Gat, Y., Røstum, J. & Krogh, A. H. 2013 Wastewater pipes in Oslo: from condition monitoring to rehabilitation planning. *Water Practice & Technology* **8** (3–4), 487–494.

Vollertsen, J. & König, A. 2005 *Report D6: Model Testing and Evaluation*. Report of SINTEF, SINTEF, Trondheim.

Wirahadikusumah, R., Abraham, D. & Iseley, T. 2001 Challenging issues in modeling deterioration of combined sewers. *Journal of Infrastructure Systems* **7** (2), 77–84.

Wu, W., Liu, Z. & He, Y. 2013 Classification of defects with ensemble methods in the automated visual inspection of sewer pipes. *Pattern Analysis and Applications* **16** (3), 417–430.

Yang, J. 2004 *Road Crack Condition Performance Modeling Using Recurrent Markov Chains and Artificial Neural Networks*. PhD Thesis, University of South Florida.