



# Bayesian network-based methodology for selecting a cost-effective sewer asset management model

Julián Guzmán-Fierro, Sharel Charry, Ivan González, Felipe Peña-Heredia, Nathalie Hernández , Andrea Luna-Acosta and Andrés Torres 

## ABSTRACT

This paper presents a methodology based on Bayesian networks (BN) to prioritize and select the minimal number of variables that allows predicting the structural condition of sewer assets to support the strategies in proactive management. The integration of BN models, statistical measures of agreement (Cohen's Kappa coefficient) and a statistical test (Wilcoxon test) were useful for a robust and straightforward selection of a minimum number of variables (qualitative and quantitative) that ensure a suitable prediction level of the structural conditions of sewer pipes. According to the application of the methodology to a specific case study (Bogotá's sewer network, Colombia), it found that with only two variables (age and diameter) the model could achieve the same capacity of prediction (Cohen's Kappa coefficient = 0.43) as a model considering several variables. Furthermore, the methodology allows finding the calibration and validation percentage subsets that best fit (80% for calibration and 20% for validation data in the case study) in the model to increase the capacity of prediction with low variations. Furthermore, it found that a model, considering only pipes in critical and excellent conditions, increases the capacity of successful predictions (Cohen's Kappa coefficient from 0.2 to 0.43) for the proposed case study.

**Key words** | Bayesian networks for feature selection, influencing factors, prediction model, sewer asset management tool, structural condition assessment



## HIGHLIGHTS

- Bayesian network-based methodology for selecting a cost-effective sewer asset management model.
- Integration of BN models and statistical measures of agreement and tests were useful for a robust and straightforward selection of a minimum number of variables (qualitative and quantitative) that ensure a suitable prediction level of the structural conditions of the sewer network.
- Identification of the cluster that gives higher predictability.
- Identification of the calibration portion that gives higher predictability.
- Exploration of variables from different natures that could influence the deterioration of sewer assets.

## INTRODUCTION

Wastewater infrastructures, including collection pipes and treatment facilities, represent an enormous investment in physical assets. In the last 30 years, most municipalities have

invested in sewer system expansion to meet growth and treatment plant upgrades, but they allocated a relatively small proportion of the budget to sewer rehabilitation (AWWA

Julián Guzmán-Fierro  
Sharel Charry  
Ivan González  
Felipe Peña-Heredia  
Nathalie Hernández  (corresponding author)  
Andrés Torres   
Engineering Faculty,  
Pontificia Universidad Javeriana,  
Bogotá,  
Colombia  
E-mail: [nathalie\\_hernandez@javeriana.edu.co](mailto:nathalie_hernandez@javeriana.edu.co)

Andrea Luna-Acosta  
Rural and Environmental Studies Faculty,  
Pontificia Universidad Javeriana,  
Bogotá,  
Colombia

2012). As a result, most cities face the problem of ageing infrastructure in need of extensive and ongoing repair, rehabilitation or renewal (Caradot *et al.* 2017). Traditionally, it has been economically feasible to apply reactive management strategies, repairing when failures occur; however, this strategy will become less viable as the systems age and the funding gap increases (Rokstad & Ugarelli 2015).

The development of tools for proactive management of sewerage networks requires actions such as the collection and processing of a large amount of information and the construction of forecast models; those actions may require high economic costs. According to Caradot *et al.* (2013), the more information a model has about the variables that could influence the structural conditions, the higher is the performance of the prediction models. However, the number of variables differs depending on the case study and the strategies to prioritize the inclusion of some factors over others (Chornet 1994). Furthermore, the expert knowledge, the available information (Kabir *et al.* 2016; Angarita *et al.* 2017), and the ease of collection (costs and time) (Angkasuwansiri & Sinha 2013) are some of the strategies for choosing the variables that could influence the sewer condition. For example, Angkasuwansiri & Sinha (2013) determined that almost 60 variables could affect the performance of the pipes, and other studies have included between five and 16 variables that contribute directly to the deterioration of the sewerage network (Ariaratnam *et al.* 2001; Baah *et al.* 2015; Rokstad & Ugarelli 2015; Kabir *et al.* 2016; Laakso *et al.* 2018).

Regarding methodological procedures described in the literature for predicting the structural condition of sewer pipes, a wide variety of tools have been used such as: simple logistics (Ariaratnam *et al.* 2001), regression techniques (Sousa *et al.* 2014; Angarita *et al.* 2017), Markov statistical models (Scheidegger *et al.* 2015), artificial intelligence models (e.g. neural networks, Jafar *et al.* 2010) and random forest (Caradot *et al.* 2018). Most of these models can estimate the structural condition of sewer pipes with greater or lesser success depending on: (i) the number of analysed variables (Angkasuwansiri & Sinha 2013; Hernández *et al.* 2017); (ii) the quantity of data required for training (Angarita *et al.* 2017); and (iii) the assumptions used for the construction of models (Rokstad & Ugarelli 2015). The last conditions lead to the development of predictive models with extensive information and high capacity to capture the nonlinear relationships between the variables and the structural conditions (Kabir *et al.* 2016), especially when nominal variables are considered (Ariaratnam *et al.* 2001). However, the quantity of sufficient information to achieve a successful predictive model is unknown,

considering the larger amount of information, and higher costs for the utilities due to their collection.

In recent years, Bayesian networks (BN) have become a promising tool for cause-effect analyses, which allows representing uncertain knowledge in probabilistic systems such as risk analysis, and has proven to be effective in capturing and integrating both qualitative and quantitative information from various sources (Kabir *et al.* 2016; Li *et al.* 2016). For example, España (2007) developed a model for prioritizing the pipes to construct inspection plans based on BN, GIS (geographical information systems) and survival functions. In this model, the BN allowed incorporating the information in an organized way and limited zones where the information's cost was readily available. Furthermore, this identified the relevant variables to the failure mechanisms and their consequences, as well as to establish conceptual relationships among them.

Thanks to the properties of the Bayesian networks displayed in other experiences (España 2007; Jung & Hobert 2014), it could be a suitable tool to determine the quantity of variables that are enough to obtain satisfactory prediction quality for the structural condition of sewer pipes, with the purpose of reducing the collection costs of the variables that could influence the structural condition. Therefore, this paper presents a methodology based on BN to prioritize and select a minimal number of variables that allows predicting the structural condition of assets of sewerage infrastructure.

## MATERIALS AND METHODS

### Methodology

According to Figure 1, the methodology consists of five steps.

*Step 1* Merging in one database the CCTV reports (assessment of the structural condition of sewer assets) and geo-referenced information of physical characteristics of sewer networks, environmental and operational characteristics surrounding sewer assets. It is important that all numerical variables become categorical ones assuring more than one factor on each variable.

*Step 2* Creation of different structural condition aggrupation scenarios (SCs), in which the structural conditions are grouped (three SCs were considered: all the conditions depending on the local standard, critical conditions and others – 2 categories, only critical and excellent conditions – 2 categories) with the purpose of identifying in step 3 which aggrupation gives more predictive capacity to the model.

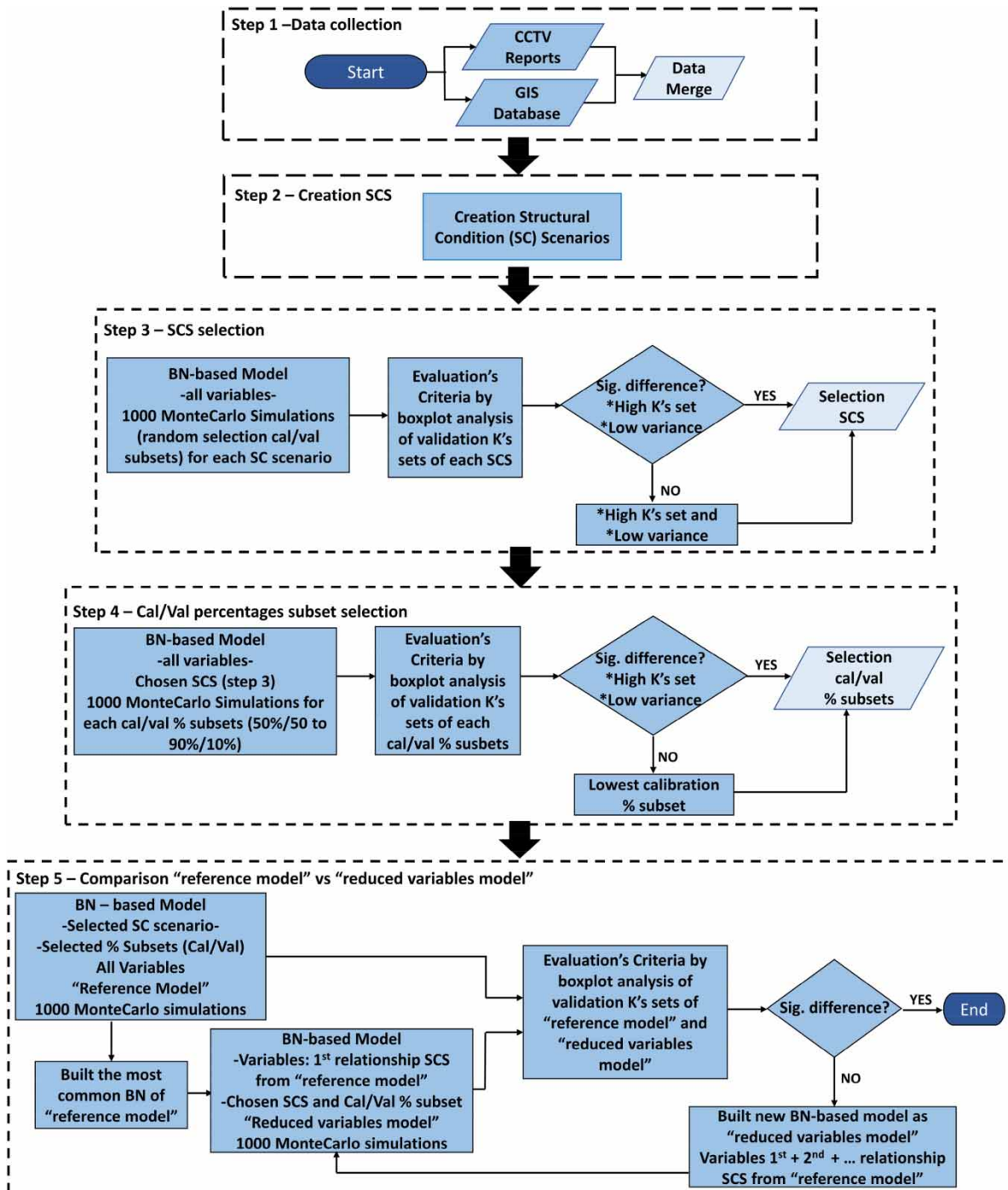


Figure 1 | Flowchart of the methodology.

*Step 3* Choosing the structural condition scenario (SCs), from the training of 1000 Bayesian networks (BN) – based models using the Hill-Climbing algorithm, Cohen’s Kappa coefficient (K) (Vieira *et al.* 2010) and Wilcoxon test. For each scenario defined in

*step 2*, BN-based models are trained considering 1000 random selections (Monte-Carlo simulations) of different calibration and validation subsets (percentages from 50%/50% to 90%/10%) and all available variables to predict the structural

condition range values for each validation set. K is used to evaluate the prediction performance of each model. Then, the K's sets of each scenario are compared to each other by Wilcoxon test to choose the structural condition scenario with significant highest performance prediction.

*Step 4* Choosing the percentage of data for calibration and validation subsets (from 50%/50% to 90%/10%) which gives more predictive capacity to the model. Considering the chosen SCs of step 3, for each calibration/validation percentage subset, it trains 1000 BN-models using the Hill-Climbing algorithm and all available variables to predict the structural condition range values for each validation set. As well as step 3, it calculates K to measure the model performance and the Wilcoxon test to choose the calibration/validation percentage subset that gives more predictive capacity. The chosen calibration/validation percentage subset is the one that shows the highest K's set, a low variance, and shows significant difference with K's sets of other calibration/validation percentage subsets. In the case that K's sets do not show a significant difference, the calibration/validation percentage subset that needs less data for training the model is chosen.

*Step 5* Comparing the prediction performance of reference model vs. reduced variables model. A BN model 'reference model' is constructed, considering the chosen SCS (step 3), calibration/validation percentage subsets (step 4) and all the available variables. From the reference model, a BN is built to extract the variables with a direct relationship (first parenting relationship) with the structural condition. Then, it builds a new model using only these variables (first parenting relationship). Anew, 1000 random selections (Monte Carlo simulations) of the chosen calibration/validation percentage subset are carried out in both models ('reference model' and the new 'reduced variables model') and these are evaluated by K, creating two K's sets related to the 'reference model' and the 'reduced variable model' respectively. The Wilcoxon test was applied for determining if both K's sets are significantly different or not. If there is not a difference significantly, it means that it is possible to build a model with few variables that achieves the same capacity prediction as a model considering all

available variables; conversely, if there is a difference significantly between both models, it builds a new model considering the variables with a direct relationship to the structural condition (first parenting relationship) and the variables with a direct relationship to the first relationship variables (second parenting relationship with the structural condition variable), and it carries out again the same comparison procedure. If the comparison still shows differences significantly between both the 'reference model' and 'reduced variables model', it considers the third, fourth and so on parenting relationship variables to build a new 'reduced variables models' until it finds a model that could achieve the same capacity prediction as the 'reference model'.

## Case study

The case study was Bogotá's sewer network. After data clean-up, 8349 consistent inspections (representing 430 km) were linked to 7,968 pipes (around 3% of the total sewer system). The sewer inspections were collected by the Empresa de Acueducto y Alcantarillado de Bogotá (EAAB), between 2007 and 2017. From CCTV inspections, structural and operational failures were found in the interior of the sewers, which allowed assessment of the state of the sewer assets according to the evaluation criteria of the EAAB standard NS-058 'Technical Aspects for Inspection of Sewer Networks and Structures' (EAAB 2001). The standard gives a score to each structural failure found in CCTV inspection; the magnitude of the score depends on its severity. Then, it sums all the scores of the structural failures found in the inspected asset. According to the sum of the scores, the standard provides a categorical classification of the structural condition in five grades (from 1 to 5). Table 1 shows the diagnosis and recommendations that correspond to the classification of the structural condition of the sewer assets given by this standard.

Besides, EAAB provided the geo-referenced EAAB databases updated to 2017 that contain the physical characteristics of 212,382 sewer system pipes in Bogotá in addition to the date of installation of each one (between 1939 and 2014) and the database of the water level in Bogotá in 2010.

Furthermore, for the development of work, databases of other public institutions were also considered, such as: (i) database of geotechnical zoning in 2010, collected by FOPAE (Bogotá's fund for prevention and emergency

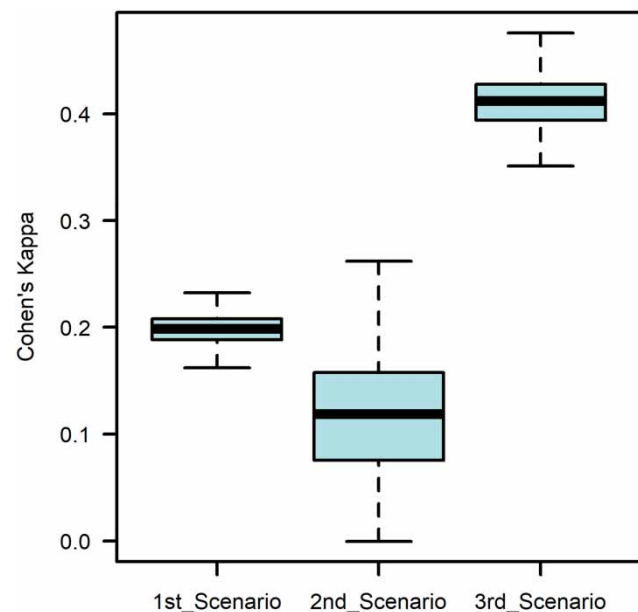
**Table 1** | Classification of structural grades, diagnosis and recommendations by NS-058

Sum of structural scores	Grade	Diagnosis	Recommendations
<10	1	With little structural defects which are not important for the structural stability of the system	New inspection within 4–5 years to verify the structural state of the sewer asset
10–39	2	The defects found are minor and do not compromise the stability of the network in the short term	Maintenance actions to correct the damage found and make a new inspection within 3–4 years to analyse the structural risk
40–79	3	Defects that could generate structural problems	Maintenance actions to correct defects by prioritizing them according to their severity or qualification. Make a new inspection within 2–3 years to verify the result, the actions carried out and that the structural risk has not increased
80–164	4	Defects with high importance and could generate structural problems	Rehabilitation actions that prevent the damage propagation by prioritizing defects according to the severity or qualification. Make a new inspection within 1–2 years to analyse the performed actions
165+	5	Collapsed sewer asset	Replacement actions of an urgent nature

care), which describes the type of soil and geomorphology; (ii) database of trees in Bogotá in the year 2017 provided by JBB (Bogotá's Botanical Garden); (v) land use database of the year 2016; (vi) road database that contains information of about 308,527 roads classified according to the type of traffic and its surface finish in the year 2016; (vii) database of the 20 districts of Bogotá in the year 2017 collected by IDECA (Bogotá's integrated infrastructure of spatial data).

## RESULTS AND DISCUSSION

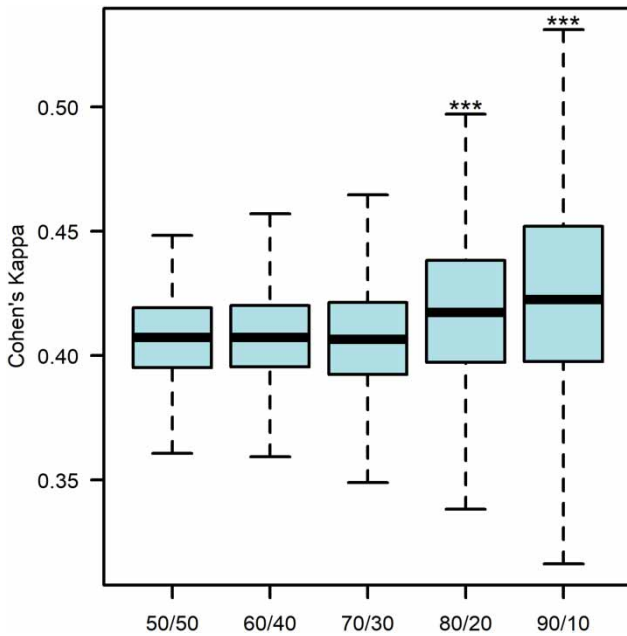
According to the results of the proposed methodology for the case study, the structural condition scenario with the highest prediction performance (K) was the scenario considering only sewer assets in critical and excellent conditions with a median of 0.41, which is significantly higher (Wilcoxon test,  $p$ -value < 0.001) than the first and second scenarios (0.19 and 0.12 respectively) (Figure 2). The percentage of calibration and validation subsets with the significant highest K-values (for the validation data) compared to the other percentage subsets was 90%/10% with a median of 0.42. (Figure 3). However, the variation of this group is the largest of all the percentage subsets, with values that vary between 0.32 and 0.53, which makes it less favourable to predict the structural condition of sewer pipes. On the other hand, K's median of 80%/20% had values around



**Figure 2** | Boxplots of validation of K's values obtained from the three SCSS: 1st scenario – five structural grades; 2nd scenario – critical and other conditions; and 3rd scenario – excellent and critical conditions.

0.417 with lower variability than the 90%/10% subset. Therefore, it chooses the 80%/20% percentage subset.

With the chosen scenario (only critical and excellent conditions) and calibration/validation percentage subsets (80%/20%), a Bayesian network was built considering all the available variables of the case study and the structural condition (CC\_ST) of sewer pipes. The available variables



**Figure 3** | Boxplots of validation K's values obtained from models that considered calibration/validation percentages subsets between 50%/50% and 90%/10% (\*\*\*)*p*-value < 0.001).

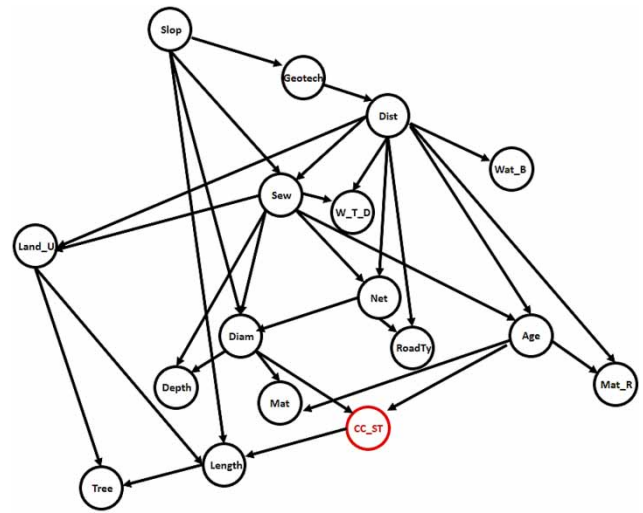
of the case studies are: Pipe Slope (Slop), Geotechnical Zone (Geotech), District (Dist), Water Body Closeness (Wat-B), Sewer Type (Sew), Water Level Depth (W\_T\_D), Land Use (Land\_U), Pipe Diameter (Diam), Pipe Type (Net), Road Type (RoadTy), Pipe Age (Age), Pipe Installation Depth (Depth), Pipe Material (Mat), Road Material (Mat\_R), Pipe Length (Length), and Intrusive Tree presence (Tree).

Figure 4 shows the most representative BN model with 1000 random selections considering all available variables, the chosen scenario (only pipes in critical and excellent conditions) and calibration/validation percentage subsets (80% for training data and 20% for validation data).

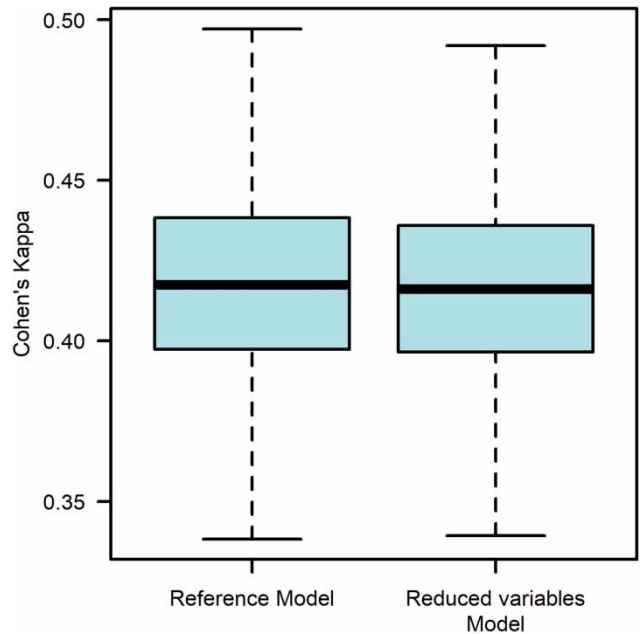
According to Figure 4, the pipe's age and pipe's diameter are the variables called as father variables (first-grade parenting variables) over the structural condition (CC\_ST) related to critical and excellent conditions. Therefore, according to the presented methodology, 1,000 BN-based models were built using father variables of the structural conditions found in Figure 4.

In Figure 5, the prediction performance (Ks) of the validation data is presented, comparing the two types of models: reference model (16\_var); and (ii) reduced variables model (only first parenting relationship -2\_var).

According to Figure 5, the K's set related to the 'reference model' shows a median of 0.438 that is not significantly higher (Wilcoxon test, *p*-value = 0.08) than the



**Figure 4** | The most representative Bayesian network considering all available variables and the structural condition scenario that only takes into account critical and excellent conditions (CC\_ST), training the model with 80% of data and validating with 20% of data.



**Figure 5** | Boxplot of validation K's sets obtained by the 'reference model' (16\_var) and the 'reduced variables model' considering only the first-parenting variables with the structural condition (2\_var): age and diameter of sewers.

median of the K's set related to the 'reduced variables model' (K's median of 0.432). This result shows that only considering the diameter and the age of the sewer assets; it is possible to build a prediction model, using BN, that achieves the same prediction capacity (K's > 0.04 – moderate

association agreement between predicted and observed structural conditions) than a model based on BN that considers more variables.

From the chosen model, it predicts the structural condition of sewer assets for validation data and to the entire sewer system of Bogotá. These results are shown in Figures 6 and 7. Figure 6 shows a map comparing the predicted and observed sewer assets for the Chapinero district, which is in the northeast of Bogotá near the hills. Figure 7 shows a map with the prediction of the whole of Bogotá's sewer system. Figure 6 shows that in general, the structural conditions of observed and predicted sewer pipes have similar distributions. The southeastern zone presents a higher concentration of pipes in structural condition 5 (critical structural conditions), which can be explained by the

presence of smaller diameters and old neighbourhoods such as Pardo Rubio and Chapinero Central, both consolidated in the first half of the 20th century (Mejía 2007). On the other hand, the northwest zone presents more pipes in structural condition 1 (excellent condition) since it has larger diameters and neighbourhoods such as Rosales and Chico, built more recently.

Figure 7 shows that there is a more significant number of pipes in condition 5 (critical conditions) in the eastern part of the city. This could be because of having the smallest pipes (diameters less than 400 mm) and the presence of old districts such as Chapinero, Martires and Usaquen with pipes over 40 years old. The situation described previously contrasts with the density of pipes in structural condition 1 in the central and northwestern zone, which

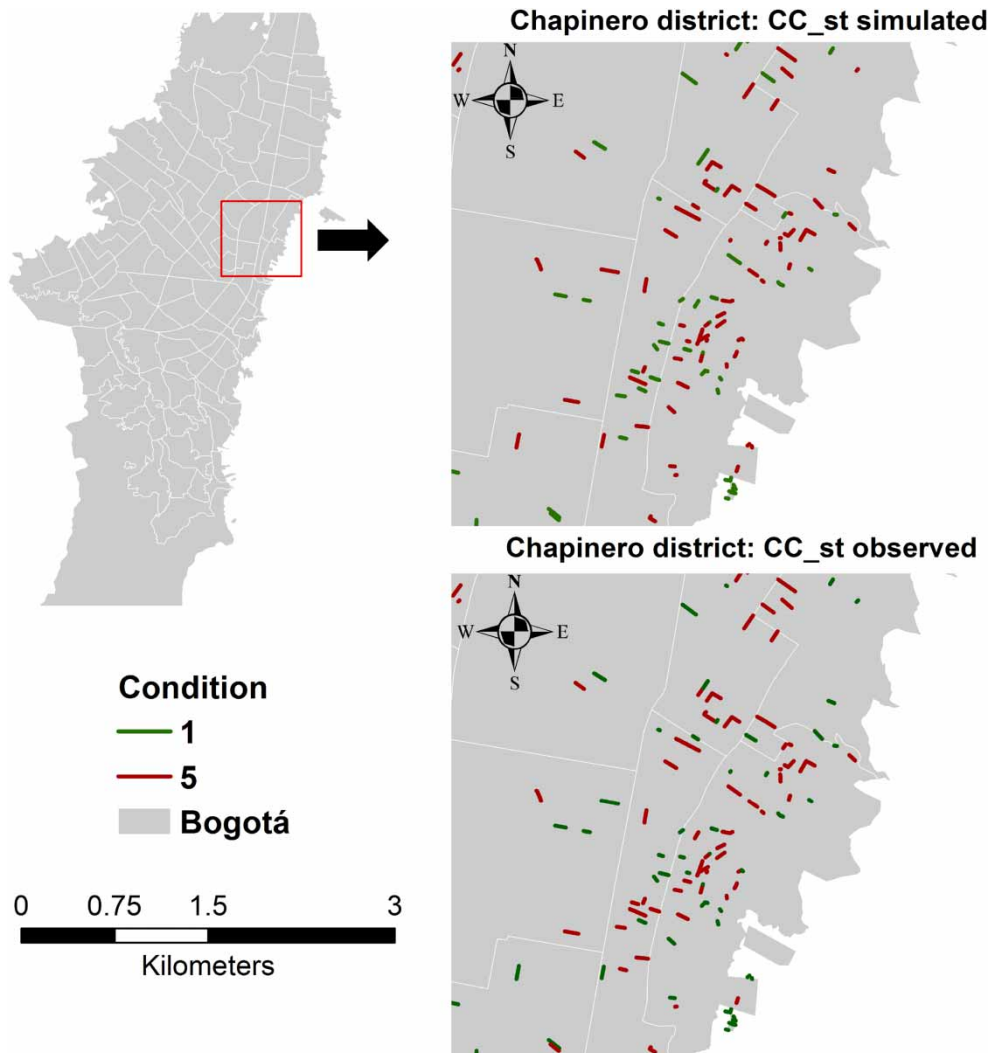
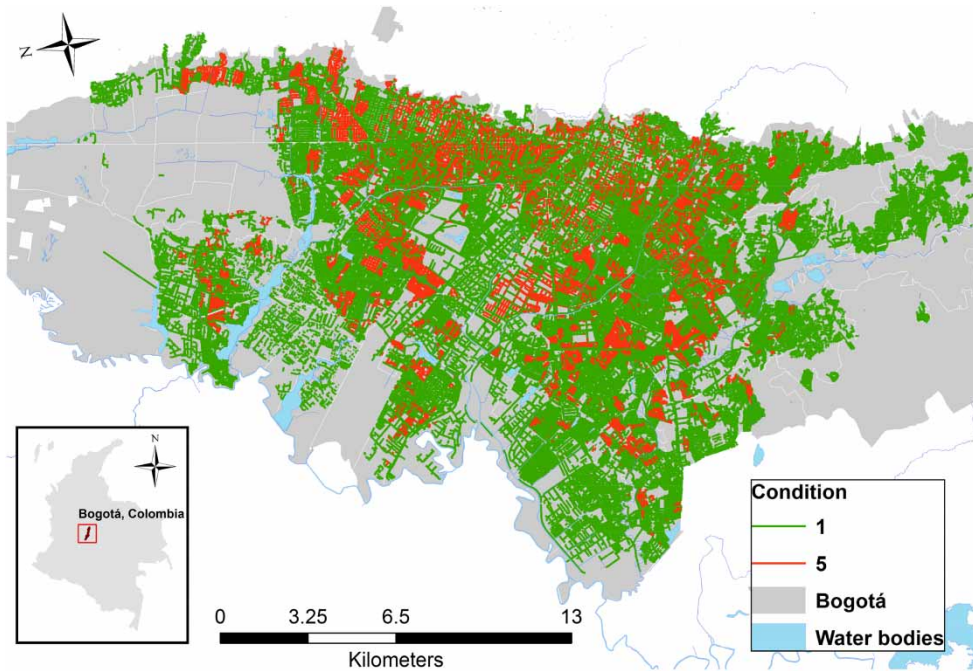


Figure 6 | Inspected and predicted sewer pipes in Chapinero sector.



**Figure 7** | Bogotá's predicted sewer conditions.

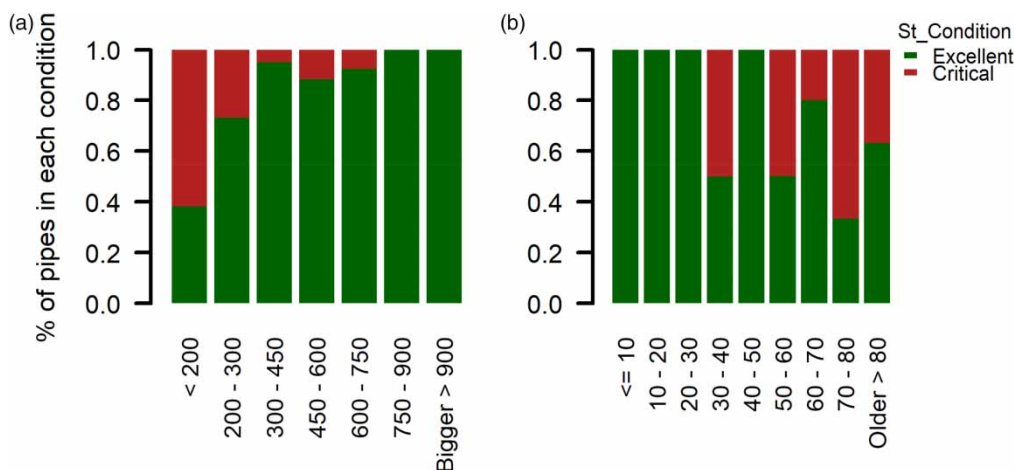
has larger diameter pipes (more than 600 mm) and newer districts such as Barrios Unidos and Suba.

According to the above results (Figure 7), Figure 8(a) and 8(b) show histograms that analyze the prediction of the structural condition (excellent and critical) of the whole sewer system regarding the diameter and age variables.

According to Figure 8(a), the histogram shows that the percentage of pipes in critical structural condition is higher for small diameters (less than 600 mm) than for

larger ones (greater than 600 mm). This trend agrees with the results found by Baur & Herz (2002), Kulandaivel (2004), Angarita *et al.* (2017) and O'Rourke & Vargas (2018), who obtained higher rates of pipes in critical structural conditions when the sewers' diameter decreased.

On the other hand, Figure 8(b) shows that the percentage of pipes in critical structural condition increases when the sewer pipes are older. This result confirms the findings of Ariaratnam *et al.* (2001), Kulandaivel (2004), El-Housni *et al.* (2017), Xu *et al.* (2018), who suggest that



**Figure 8** | (a) Histogram of Bogotá's sewer pipe diameters vs predicted structural conditions (%). (b) Histogram of Bogotá's sewer pipe ages vs predicted structural conditions (%).



the critical structural conditions in sewer pipes are related to the age of the pipes and their life cycle.

## CONCLUSIONS

The present study proposed a methodology for selecting effective sewer asset management models based on Bayesian networks to prioritize and select the minimal and enough variables that allow prediction of the structural condition of sewer pipes with the same accuracy as there would be with a model with a large number of variables. Selecting and reducing the number of variables diminishes the quantity of information to collect and therefore, requires a lower amount of investment resources for data collection.

Within the structural condition scenarios, the scenario that considers only the structural grade 1 (excellent conditions) and structural grade 5 (critical conditions) was the one that obtained a higher prediction capacity. This confirms the findings of Ariaratnam *et al.* (2001) and Lopez-Kleine *et al.* (2016), in which the prediction improves considering only two states: excellent status (without failures) and status with failures. Besides the proposal of only considering the pipes in critical and excellent conditions, leaving aside the pipes in intermediate conditions (grades 2, 3 and 4), reduces the uncertainty grade generated by wrong assessing due to the quality of inspection technology or wrong assessment of the sewers' structural conditions by the operators. Moreover, this proposal could support prioritization plans with the identification of those pipes that need a replacement urgently, and inspection plans with those pipes that were not in critical nor in excellent condition (intermediate condition).

Furthermore, the methodology also contemplates an analysis to select the calibration/validation percentages subset to identify which percentages' subset is appropriate to train a suitable prediction model. According to the case study, the subsets of percentages 80%/20 and 90%/10% give higher predictions in the validation results (Kappa around 0.42); however, the 90%/10% subset shows higher variability than 80%/20% subset. Therefore, the chosen subset was 80% of data for calibration and 20% of data for validation.

The use and integration in the methodology of Bayesian networks, Cohen's Kappa coefficient, Wilcoxon test, and Monte-Carlo simulations allow measurement in a simple, direct and robust way of prediction assessment of the model, the existence of significant differences between one model or another, and the decreasing of random effects. It

means a model more trustworthy for being used for the utilities in the sewer system of any city.

According to the results of the case study (Sewer Systems of Bogotá city, Colombia), within the 16 variables studied, only two variables (diameter and age) are enough to predict which pipes are in critical or excellent structural conditions. These two variables allow decreasing of the database size and the recollection costs, since physical characteristics are easy to obtain. This supports the findings of Baur & Herz (2002), Kulandaivel (2004), Angarita *et al.* (2017) and O'Rourke & Vargas (2018), which gives importance to the age and diameter of sewer pipes as factors that influence the structural condition: older and smaller (diameters less than 600 mm) sewer pipes tends to be in critical conditions.

## ACKNOWLEDGEMENTS

The authors would like to thank COLCIENCIAS and PUJ for supporting one of the authors in her Ph.D's studies ('Convocatoria 727 del 2015- Apoyo doctorados nacionales').

A special acknowledgement to EAB ('Empresa de Acueducto de Bogotá'), FOPAE ('Fondo de Prevención y Atención de Emergencias'), IDECA ('Infraestructura Integrada de Datos Especiales para el Distrito Capital'), and JBB ('Jardín Botánico de Bogotá') for supplying the databases information used in this research.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Angarita, H., Niño, P., Vargas, D., Hernández, N. & Torres, A. 2017 Identifying explanatory variables of structural state for optimum asset management of urban drainage networks: a pilot study for the city of Bogota. *Ingeniería e Investigación* 37 (2), 6–16.
- Angkasuwansiri, T. & Sinha, S. K. 2013 Comprehensive list of parameters affecting wastewater pipe performance. *Technol. Interface Int. J.* 13 (2), 68–79.
- 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.

- AWWA 2012 *Buried no Longer: Confronting America's Water Infrastructure Challenge*. AWWA's infrastructure financing report, American Water Works Association, Boulder, CO, USA
- Baah, K., Dubey, B. & Harvey, R. 2015 *A risk-based approach to sanitary sewer pipe asset management*. *Science of the Total Environment* **505**, 1011–1017. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2014.10.040>.
- Baur, R. & Herz, R. 2002 *Selective inspection planning with ageing forecast for sewer types*. *Water Science and Technology* **46** (6–7), 389–396.
- Caradot, N., Sonnenberg, H., Kropp, I., Schmidt, T., Ringe, A., Denhez, S., Hartmann, A. & Rouault, P. 2013 Sewer deterioration modelling for asset management strategies – state-of-the-art and perspective. In: *5th LESAM – Leading-Edge Strategic Asset Management 2013*, 10–23 September, Sydney, Australia.
- Caradot, N., Sonnenberg, H., Kropp, I., Ringe, A., Denhez, S., Hartmann, A. & Rouault, P. 2017 *The relevance of sewer deterioration modelling to support asset management strategies*. *Urban Water Journal* **14** (10), 1–9.
- Caradot, N., Riechel, M., Fesneau, M., Hernandez, N., Torres, A., Sonnenberg, H. & Rouault, P. 2018 *Practical benchmarking of statistical and machine learning models for predicting the condition of sewer pipes in Berlin, Germany*. *Journal of Hydroinformatics* **20** (5), 1131–1147.
- Chornet, A. P. 1994 *LA URBANIZACIÓN Y LA GESTION MUNICIPAL EN AMÉRICA LATINA*. *Las ciudades de América Latina: problemas y oportunidades* **33**, 9.
- El-Housni, H., Ouellet, M. & Duchesne, S. 2017 *Identification of most significant factors for modeling deterioration of sewer pipes*. *Canadian Journal of Civil Engineering* **45** (3), 215–226.
- Empresa de Acueducto y Alcantarillado de Bogotá EAAB 2001 *NS – 058. Aspectos Técnicos para inspección y mantenimiento de redes y estructuras de alcantarillado*. EAAB-E.S.P, Bogotá, Colombia.
- España, J. J. P. 2007 *Modelo para priorizar la inspección de redes de alcantarillado*. Doctoral dissertation, Uniandes, Bogotá, Colombia.
- Hernández, N., Caradot, N., Sonnerberg, H., Rouault, P. & Torres, A. 2017 Support tools to predict the critical structural condition of uninspected sewer pipes in Bogota D.C. In: *The Leading Edge Sustainable Asset Management of Water and Wastewater Infrastructure Conference*, 20–22 June 2017, Trondheim, Norway.
- Jafar, R., Shahrour, I. & Juran, I. 2010 *Application of Artificial Neural Networks (ANN) to model the failure of urban water mains*. *Mathematical and Computer Modelling* **51**, 1170–1180. <https://doi.org/10.1016/J.MCM.2009.12.033>.
- Jung, Y. J. & Hobert, J. P. 2014 *Spectral properties of MCMC algorithms for Bayesian linear regression with generalized hyperbolic errors*. *Statistics & Probability Letters* **95**, 92–100.
- Kabir, G., Sadiq, R. & Tesfamariam, S. 2016 *A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines*. *Structure and Infrastructure Engineering* **12** (8), 874–889.
- Kulandaivel, G. 2004 *Sewer Pipeline Condition Prediction Using Neural Network Models*. Thesis (PhD), Michigan State Univ, East Lansing, MI.
- Laakso, T., Ahopelto, S., Lampola, T., Kokkonen, T. & Vahala, R. 2018 *Estimating water and wastewater pipe failure consequences and the most detrimental failure modes*. *Water Science and Technology: Water Supply* **18**, 901–909. <https://doi.org/https://doi.org/10.2166/ws.2017.164>.
- Li, X., Chen, G. & Zhu, H. 2016 *Quantitative risk analysis on leakage failure of submarine oil and gas pipelines using Bayesian network*. *Process Safety and Environmental Protection* **103**, 163–173.
- López-Kleine, L., Hernández, N. & Torres, A. 2016 *Physical characteristics of pipes as indicators of structural state for decision-making considerations in sewer asset management*. *Ingeniería e Investigación* **36** (3), 15–21.
- Mejía, F. A. C. 2007 *Historia del desarrollo urbano del centro de Bogotá*. Instituto de Cultura Recreación y Deporte de Bogotá, Bogotá, Colombia.
- O'Rourke, M. & Vargas-Londono, T. 2018 *Influence of diameter on seismic response of buried segmented pipelines*. *Soil Dynamics and Earthquake Engineering* **107**, 332–338.
- Rokstad, M. M. & Ugarelli, R. M. 2015 *Evaluating the role of deterioration models for condition assessment of sewers*. *Journal of Hydroinformatics* **17** (5), 789–804.
- Scheidegger, A., Leitão, J. P. & Scholten, L. 2015 *Statistical failure models for water distribution pipes – a review from a unified perspective*. *Water Research* **83**, 237–247. <https://doi.org/10.1016/J.WATRES.2015.06.027>.
- Sousa, V., Matos, J. P. & Matias, N. 2014 *Evaluation of artificial intelligence tool performance and uncertainty for predicting sewer structural condition*. *Automation in Construction* **44**, 84–91. <https://doi.org/10.1016/J.AUTCON.2014.04.004>.
- Vieira, S. M., Kaymak, U. & Sousa, J. M. 2010 *Cohen's kappa coefficient as a performance measure for feature selection*. In: *International Conference on Fuzzy Systems*. IEEE, New York, NY, pp. 1–8.
- Xu, Q., Qiang, Z., Chen, Q., Liu, K. & Cao, N. 2018 *A superposed model for the pipe failure assessment of water distribution networks and uncertainty analysis: a case study*. *Water Resources Management* **32**, 1713–1723. <https://doi.org/10.1007/s11269-017-1899-8>.

First received 8 January 2020; accepted in revised form 8 June 2020. Available online 23 June 2020