

Geoadditive Bayesian regression models for water mains failure rate prediction

Ngandu Balekelayi

Graduate Student, School of Engineering, University of British Columbia, Kelowna, Canada

Solomon Tesfamariam

Professor, School of Engineering, University of British Columbia, Kelowna, Canada

ABSTRACT: Application of pure linear deterioration models for Water Distribution Networks (WDNs) is not effective in the representation of the physical degradation of water pipes because of the theoretical approach of water pipes deterioration or simply the uncertainty related to the specific form of effects that a covariate has on the response variable. Polynomial approaches are convenient to represent the complexity of the physical phenomena. However, even high degree polynomials wiggly estimate the relationships and are unsatisfactory in some regions where they fail to fit the observed data. Flexible regression techniques that enable automatic data-driven estimation of nonlinear relations between covariates and response constitute an alternative approach that is able to represent the physical deterioration process. In this study, a Geoadditive Bayesian regression model with smooth nonlinear splines functions for the continuous covariates and spatially distributed effects for the geospatial information of the pipes is applied to predict the failure rate of metallic water mains. The results highlight nonlinear dependency between continuous covariates and the response variable. A map representing the effect of the covariates and the geospatial location of the pipes on the response variable is produced. This map can be used as an early indicator to localize areas where the effect of covariates on the failure rate is high and prioritize them for inspections and maintenance.

1. INTRODUCTION

Water distribution network (WDN) is an essential part of buried urban water infrastructures and its failure has high economic, social and environmental costs (Wu and Liu 2017; Wilson et al. 2015). Individual pipe performance declines until failure occurs due to cumulative effects of pipes intrinsic characteristics, operational and environmental factors (Kakoudakis et al. 2017). Increased inspection activities of water pipes infrastructure due to the adoption of proactive management made available extensive pipes database. The availability of these database poses significant challenges in the development of deterioration models that are important for the continuity and quality of services provided (Scheidegger et al. 2015).

Wilson et al. (2017) classified water mains deterioration models into two categories: (1) physical models that are more precise and do not require extensive historical data (e.g. Rajani and

Tesfamariam 2004; Rajani and Kleiner 2001) and (2) statistical models that learn the failure causes in the historical record of failure of pipes (e.g. Kabir et al. 2015; Kleiner and Rajani 2001). The complexity and the lack of understanding of the physical mechanism leading to the failure of buried pipes coupled to the high cost of acquisition of the required data do not allow the use of physical models in large networked infrastructure (Kleiner and Rajani 2001). Statistical models are the most used models to predict water pipes failure because they can be applied with various levels of data and there is no need of understanding the physical mechanism that lead to the failure of pipes (Wilson et al. 2017). Several drawbacks have been identified in statistical deterioration models for water mains (e.g. inaccurate prediction due to low data quality, applicable to specific region and limited number of available variables: lack of data). Recent development of statistical techniques can enhance existing deterioration models through the

inclusion of nonlinearity and capture the heterogeneities embedded in the geospatial location information. Geoadditive regression models overcome the above limitations. These models use Penalized splines (P-splines) with a reduced degree to capture the nonlinear effects of covariates and the unobserved covariates are represented by a surrogate geospatial variable. Balekelayi and Tesfamariam (under review) applied geoadditive regression model to study the deterioration of sewer pipes. Other applications are found in social sciences (März et al. 2016; Scheipl et al. 2013) and recently an application to tree falling during storms has been presented by Kabir et al. (2018).

This study aims at improving existing statistical water pipe deterioration models through the inclusion of nonlinearity characterized by polynomial P-splines in the physical deterioration process and the geospatial information in terms of location variable as a surrogate for unobserved spatial variables not included in the data. A Gaussian Markov Random Field captures the spatial autocorrelation while random effects better represent observed local heterogeneities.

2. POLYNOMIAL P-SPLINES

Assume the observations $(Y_i, x_i, z_i), i = 1, \dots, n$ are elements in the inspection database where Y_i (continuous variables) represents the pipe breakage rate, $x_i = (x_{i1}, \dots, x_{iq})$ and $z_i = (z_{i1}, \dots, z_{ip})$ are vectors of categorical (e.g. material, cathodic protection) and continuous covariates (e.g. age, length), respectively. Linear model is reasonable and simple to interpret approach of simulating the effect of covariates on an output response. However, in natural processes (e.g. water pipes deterioration), there is some continuous covariates whose effects cannot, at least, a priori be described with simple linear functions. Generalized Additive Models where the linear predictors of continuous variables are replaced by nonlinear smooth functions define semi parametric predictor η_i as represented in the following equation:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_q x_{iq} + f_1(z_{i1}) + \dots + f_p(z_{ip}) \quad (1)$$

where η_i is the predictor of the observed quantity Y_i , f_1, \dots, f_p = nonlinear smooth functions of the continuous covariates and the regression coefficients $\beta_0, \beta_1, \dots, \beta_n$ are identical to linear regression coefficients. Eq.(1) is composed with two components the parametric linear part representing the categorical variables linear effects and the nonlinear part capturing the nonlinearity in the deterioration process. Eq.1 can be rewritten as follow: Eq.(2)

$$\eta_i = \eta_i^{linear} + f_1(z_{i1}) + \dots + f_p(z_{ip}) \quad (2)$$

Uncertainty has become an integral part of modeling and data acquisition (Kabir et al. 2015). Piecewise polynomial fit of the univariate functions f_j defined in Eq.1-2 with additional smoothness restrictions at the boundaries adds flexibility in the model to capture these uncertainties. The smooth functions f_j defined for continuous covariate z_i are approximated by polynomial splines of degree l_j . Each continuous covariate z_i domain is divided into small ranges defined at the knots $\xi_1 < \xi_2 < \dots < \xi_{h_j}$. In each range, truncated polynomials are defined to correct the first part of the function as shown in Eq.(3). This latter gives a combined representation of a polynomial of degree l_j and locals polynomials defined in interval determined by the knots that fulfill both requirements of global representation and smoothness characterized by the truncated functions $(z_{i,j} - \xi_{j,2})_+^{l_j}$.

$$f_j(z_{i,j}) = \gamma_{j,1} + \gamma_{j,2} z_{i,j} + \dots + \gamma_{j,l_j+1} z_{i,j}^{l_j} + \gamma_{j,l_j+2} (z_{i,j} - \xi_{j,2})_+^{l_j} + \dots + \gamma_{j,l_j+h_j-1} (z_{i,j} - \xi_{j,h_j-1})_+^{l_j} \quad (3)$$

where $(z_{i,j} - \xi_{j,k})_+^{l_j} = \begin{cases} (z_{i,j} - \xi_{j,k})^{l_j} & \text{if } z_{i,j} \geq \xi_{j,k} \\ 0 & \text{otherwise} \end{cases}$

Basic splines (B-splines) constitute a better alternative to solve the collinearity of truncated polynomials (nearly linear dependent) when the

knots are very close to each other. B-splines in Eq. (4) are constructed from piecewise polynomials that are fused smoothly at knots to fit the required smoothness constraints.

$$\begin{aligned} B_{j,1}(z_{i,j}) &= 1, B_{j,2}(z_{i,j}) = z_{i,j}, \dots, B_{j,l_j+1}(z_{i,j}) \\ &= z_{i,j}^{l_j}, B_{j,l_j+1}(z_{i,j}) \\ &= (z_{i,j} - \xi_{j,2})_+^{l_j}, \dots, B_{j,l_j+h_j-} \\ &= (z_{i,j} - \xi_{j,h_j-1})_+^{l_j} \end{aligned} \quad (4)$$

The univariate function f_j will be the integration of $(l+1)$ polynomial pieces of degree l_j joined in a $(l-1)$ continuously differentiable way. Eq. (3) is rewritten as given in Eq. (5) giving the basic flexible form of the polynomial representation:

$$\begin{aligned} f_j(z_{i,j}) &= \sum_{k=1}^{m_j} \gamma_{j,k} B_{j,k}(z_{i,j}), \\ i &= 1, \dots, n; \\ j &= 1, \dots, p \end{aligned} \quad (5)$$

where $\gamma_{j,k}$ = amplitude (regression coefficient) that accordingly scale the B-Splines $B_{j,k}$ to fit the observed data. Each univariate spline f_j is a linear combination of $m_j = h_j + l_j - 1$ B-Splines (basis functions) $B_{j,k}$ evaluates at specified knots $z_{j,min} = \xi_{j,1} < \xi_{j,2} < \dots < \xi_{j,h_j} = z_{j,max}$.

Good fit of observed data requires a high number of equidistant knots h_j and an imposing a roughness penalty $\lambda_j \sum_{k=d}^{m_j} (\Delta^d \gamma_{j,k})^2$ on adjacent B-Splines coefficients $\gamma_{j,k}$ to smoothen the function f_j at the knots. The roughness penalty term will reduce the occurrence of high coefficients $\gamma_{j,k}$ associated with B-splines defined in Eq.(5). A trade-off between the polynomial degree l_j and the number of knots h_j is sought. Generally, 20 to 40 equidistant knots are recommended to ensure enough flexibility for polynomials degrees varying between 1 and 3 (März et al. 2016). The penalized least square criterion is estimated using the following equation:

The d -th (Δ^d) order operator represents the difference between two regression coefficients of consecutives B-splines, i.e. $\Delta^1 = \gamma_{j,k} - \gamma_{j,k-1}$ for $d = 1$.

In matrix representation, the smooth function f_j Eq. (5) will be:

$$f_j = (f_j(z_{1,j}) \dots f_j(z_{n,j}))' = \mathbf{Z}_j \boldsymbol{\gamma}_j \quad (6)$$

Therefore, the semi-parametric representation in Eq.(2) becomes:

$$\boldsymbol{\eta} = \boldsymbol{\eta}^{linear} + \mathbf{Z}_1 \boldsymbol{\gamma}_1 + \dots + \mathbf{Z}_p \boldsymbol{\gamma}_p \quad (7)$$

where $\mathbf{Z}_j = \mathbf{B}_{j,k}(\mathbf{z}_{i,j})$ is a $(n \times m_j)$ design matrix and $\boldsymbol{\gamma}_j = (\gamma_{j,1} \dots \gamma_{j,m_j})'$ is a vector of regression coefficients. The penalized least squared criterion (PLS) is given by the following Equation (März et al. 2016):

$$PLS(\boldsymbol{\lambda}) = (\mathbf{y} - \boldsymbol{\eta})'(\mathbf{y} - \boldsymbol{\eta}) + \sum_{j=1}^p \lambda_j \boldsymbol{\gamma}_j' \mathbf{K}_d \boldsymbol{\gamma}_j \quad (8)$$

where \mathbf{K}_d = penalty matrix based on the d -th order differences. However, this approach does not consider the uncertainty related to the data and the model itself. Furthermore, the minimization of the PLS in Eq. (8), often fail in practice because no optimal solution for the λ_j exists or the computational effort become intractable as the number of smooth function in the model increases (Lang and Brezger 2004). The development of computers and MCMC simulations made the Bayesian inference, an attractive way to analyze complex statistical models (Fahrmeir et al. 2013)

3. BAYESIAN STRUCTURED ADDITIVE MODEL

The developed structured additive model in Eq. (7) is flexible enough to capture the uncertainty related to inspection data and modeling process. The Bayesian framework that is convenient to capture the stochastic nature of the deterioration process allows the vectors of regression coefficients $\boldsymbol{\gamma}_j$ and $\boldsymbol{\beta}$ in the developed model to be considered as random variables and appropriate prior distributions are assigned. Non-informative priors are assumed for the parametric coefficients $\boldsymbol{\beta}$, i.e., $p(\beta_r) \propto const, r = 1 \dots q$. The first or second order Gaussian random walks

priors for the semi parametric part γ_j are defined. Random walks RM priors are represented as:

$$RW_1: \gamma_{j,k} = \gamma_{j,k-1} + u_{j,k} \quad k = 2 \dots m_j \quad (9)$$

$$RW_2: \gamma_{j,k} = 2\gamma_{j,k-1} - \gamma_{j,k-2} + u_{j,k} \quad k = 3 \dots m_j \quad (10)$$

where $u_{j,k} \sim N(0, \tau_j^2)$ = Gaussian error terms.

The quadratic penalty $\lambda_j \gamma_j' K_d \gamma_j$, see Eq. (8), is replaced by a joint multivariate Gaussian smoothing prior distribution for the regression coefficients γ_j :

$$p(\gamma_j | \tau_j^2) = \frac{1}{(\tau_j^2)^{\text{rank}(K_j)/2}} \exp\left(-\frac{1}{2\tau_j^2} \gamma_j' K_j \gamma_j\right), \quad j = 1 \dots p \quad (11)$$

where K_j = penalty matrix based on the d -th order difference and τ_j^2 = variance parameter that control the smoothness of f_j and corresponds to the inverse of the smoothing parameter λ_j defined in the penalty term in frequentist approach.

4. GEOADDITIVE MODEL

The geospatial information such as the region where the pipe is buried can be a potential source of information that embed unknown covariates affecting the pipes failure. This information in terms of geospatial location covariates acts as a surrogate of unobserved covariates. It captured the possible autocorrelation and local heterogeneities. The Geoadditive model resulting from the addition of this valuable information to the structured additive model in Eq. (7) is represented in Eq. (12).

For water mains applications, the geospatial information represents unobserved data from various human and natural activities (e.g. *load transfer, surface use, surface type, other urban water (wastewater, storm water) failure, ground disturbance, groundwater level, ground failures, soil backfill type, direct environment pH*)

$$\eta_i = x_i' \beta + f_1(z_{i1}) + \dots + f_p(z_{ip}) + f_{geo}(s_i) \quad (12)$$

The geospatial effect f_{geo} in Eq. (12) on the pipe's failure has two components: the structured correlated effect f_{struct} and the unstructured specific location effect (local heterogeneities), i.e. $f_{geo} = f_{struct} + f_{unstruct}$. The correlated spatial effect $f_{struct} = (f_{struct}(S_1) \dots f_{struct}(S_n))' = \mathbf{Z}_{struct} \gamma_{struct}$ is represented as follows: for each district $s \in \{1, \dots, S\}$ in the region of interest (e.g. in a city, different communities may represent districts), a separate regression coefficient is estimated. The vector of regression coefficients $\gamma_{struct} = (\gamma_{struct}(1) \dots \gamma_{struct}(S))'$ collects all distinct spatial effects. The $(n \times S)$ designed matrix \mathbf{Z}_{struct} connects an observation i with the corresponding spatial effects. Gaussian Markov Random Field (GMRF) priors as represented in Eq. (13) are assigned to the spatial regression coefficients γ_{struct} . These priors allow the districts in the neighborhood to have similar effects.

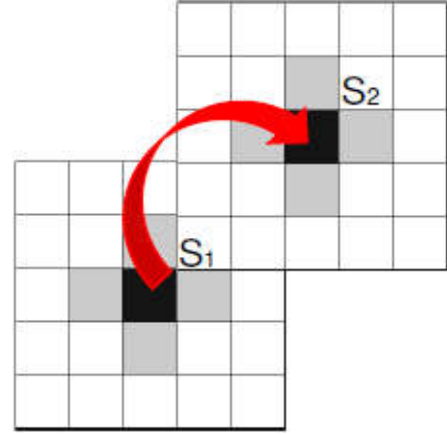


Figure 1: First order neighborhood on regular grid

$$\gamma_{struct}(s) | \gamma_{struct}(-s) \sim N\left(\frac{1}{|N(s)|} \sum_{r \in N(s)} \gamma_{struct}(r), \frac{\tau_{struct}^2}{|N(s)|}\right); \quad s = 1 \dots S \quad (13)$$

where $\gamma_{struct}(-s)$ = vector containing all the spatial effects except the one for district s , and $|N(s)|$ denotes the total number of neighbors that share a common boundary with district s . In Figure 1, the district S_1 has 4 neighbors that will

be give $|N(s)| = 4$. For irregular regional data, the number of neighbors may vary.

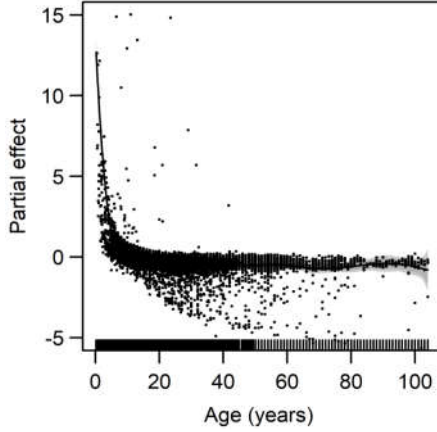


Figure 2: nonlinear partial effect of the Age covariate to the pipe breakage rate

5. GEOADDITIVE PIPE FAILURE MODEL FOR WATER MAINS

The above mathematical framework is applied to water main failure. Several covariates affect water main failure. In this study 13 covariates have been considered and the way they affect the pipe failure is examined (see Table 1). The geospatial location of pipes (here the district in which the pipe is buried is considered as surrogate covariate representing the unobserved and unknown covariates that impact the failure of pipes.

5.1. Methodology

5.1.1. Build-up

The availability of data from pipes inspections led to the search of more reliable models to predict the failure of pipes. Recent advances in statistical models such as the geoadditive approaches addressed two important issues in water pipes failure models' development. First, it allows the inclusion of nonlinearity of covariates and next, they allow the analysis of the effects of unobserved covariates in terms of geospatial location variables that are included in models.

5.1.2. Developments

All the continuous covariates in the database are assumed to have a nonlinear effect to the Pipes

Breakage Rate (*PBR*). Categorical covariates are modeled linearly, and their coefficients estimated following the least squared error techniques. Continuous covariates are modeled nonlinearly, and the coefficients are estimated as described in section 3.

Table 1: Factors affecting the pipes breakage rate

Covariates	Description
<i>Material</i>	Designed <i>material</i> of pipes (categorical: 1= cast iron, 2= ductile iron)
<i>Age</i>	The difference between the reported failure date and the installation date (continuous)
<i>Length</i>	The manhole to manhole distance (continuous)
<i>Diameter</i>	size of the pipes (continuous)
<i>NOPF</i>	Number of previous failures
<i>Rservs</i>	Number of residential connections to the pipe (continuous)
<i>Cservs</i>	Number of commercial buildings connected to the pipe (continuous)
<i>Soil corrosivity index</i>	The nature of soil representing its aggressiveness to metallic pipes (continuous)
<i>Cathodic Protection (CP)</i>	Is the cathodic protection in place(categorical)
<i>Thawing index</i>	Magnitude of Thawing season (continuous)
<i>Freezing index</i>	Severity of freezing period (continuous)
<i>Rain Deficit</i>	Difference between received and evaporated precipitation (continuous)
<i>Geospatial location</i>	The geographic location of the water pipe.
<i>PBR</i>	Pipe breakage rate is the number of breakage per year/100km (continuous)

The proposed model is as follow:

$PBR \sim$

$$a_1 \times (Material) + a_2 \times (CP) + f(Age) + f(R_{servs}) + f(C_{servs}) + f(Length) + f(Diameter) + f(NOPF) + f(FI) + f(TI) + f(RD) + f(Vintage) + f(SCI) + f(TI, RD) + map(id, bnd = clgr) + rd(id, bnd = clgr) \quad (14)$$

5.1.3. Advantages

The proposed model is a realistic approach to mimic a physical deterioration process that is not well understood and some factors that may have

effects on the output response (*PBR*) are unknown or missing in the database.

5.1.4. Contrast

The detailed geospatial location of pipes is important but not necessarily required to run the simulation. The membership of a pipe to a given region is enough to develop the model. Different models have been developed to predict the *PBR* and estimate the remaining useful life. All these models highlighted the complexity of the phenomena. The expensive parameters turning to adjust the data driven techniques internal architectures are not required.

5.1.5. Disadvantages

The approach presented here is “supervised” since the developer decide to model the continuous covariates with the single nonlinear smooth functions. However, continuous covariates may have both types of effects leading to an aggregated effect on the output covariates.

5.2. Application

5.2.1. Case selection

The proposed methodology is applied on the water distribution network of the City of Calgary in Alberta, Canada. The City has 1.2 million population and is situated at approximately 1,048m above the mean sea level. A humid continental climate with mean daytime temperature ranging from 26°C in July to -3°C in mid-January is observed in the city.

Calgary's WDN is composed of 21.2% ductile iron (DI), 15.2% CI, 3.7% asbestos cement concrete (AC) and concrete (CON), 2.9% steel (ST), 0.8% copper (CU), and 56.3% plastic pipes (PVC, PE, and FPVC). The record database shows that CI and DI pipes experienced around 95% of the total breaks, whereas 5% of the breaks are attributed to AC, CON, CU, ST, and plastic pipes. After a period of high deterioration observed between 1956 to 1980, the municipality implemented the cathodic protection coupled with an extensive replacement of corroded pipes. In this study, the deterioration of CI (cast iron) and DI (ductile iron) pipes is analyzed because they

present a high percentage of breaks and they are more susceptibility to corrosion.

5.2.2. Results

The result in Figure 2 represents the partial nonlinear effect of the “Age” covariate on the *PBR*. The literature presents the “Age” having a bathtub shape. That shape is captured by the proposed modeling approach. In Figure 2, young pipes are the most prone to failure. This is due to the installation techniques and transportation and conservation of new pipes. Once installed the pipes mature and do not have significant effect to the failure. However, old pipes are more susceptible to failure and the “Age” effect should increase. The increase is not visible on the figure because the gaussian prediction of the expected mean do not capture the large uncertain area. It is required to beyond the mean and explore extremes effects to see the real patterns of the effects and reduce uncertainty. All the other continuous

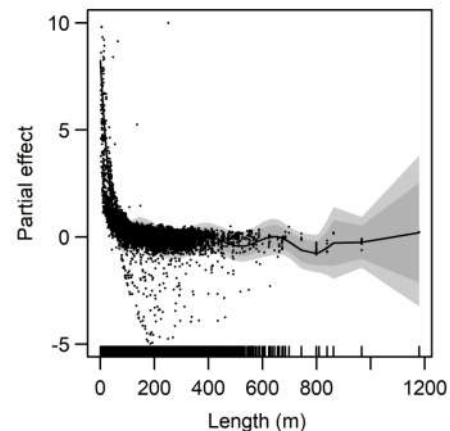


Figure 3: nonlinear partial effect of the length covariate to the pipe breakage rate

Short pipes (see Figure 3) have high nonlinear effect on the *PBR* because, no particular treatment is required to install them compared to long pipes that are installed by highly qualified workers.

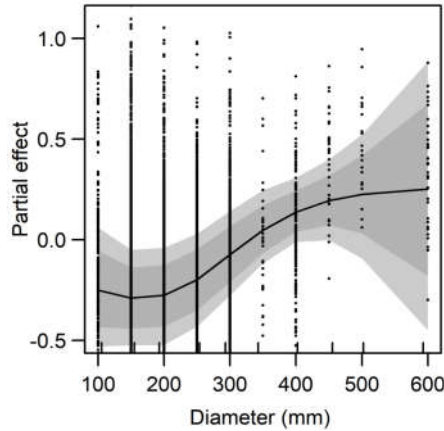


Figure 4: nonlinear partial effect of the diameter covariate to the pipe breakage rate

The number of previous failure (see Figure 4) on the same pipe is a particular covariate. Once a pipe fails, it is repaired (mostly welded) and its physical characteristics are changed due to the hot processing. Thus, it will be more prone to failure as it ages

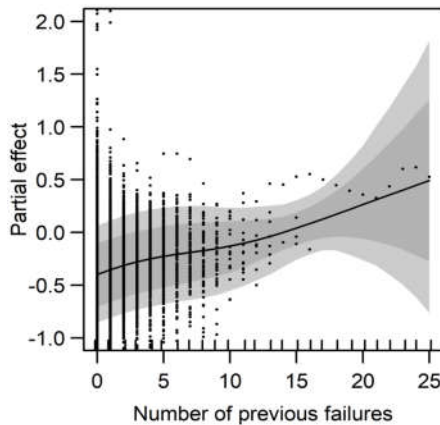


Figure 5: nonlinear partial effect of the number of previous failure covariate to the pipe breakage rate

Large diameter (Figure 4) has high effects on the pipes breakage rate. The large surface of metallic pipes exposed to the environment is the most probable reason for this observation.

Climatic parameters (Freezing Index, Thawing Index, Rain Deficit) are shown to have effects on the output. These parameters should be analyzed in spatiotemporal framework to highlight their contributions to the failure of pipes.

The map visualization (Figure 6) of areas where the structured geospatial effects are elevated is an important tool to orient future

inspections for identification of new factors that are responsible of fast or low deterioration rate of water pipes.

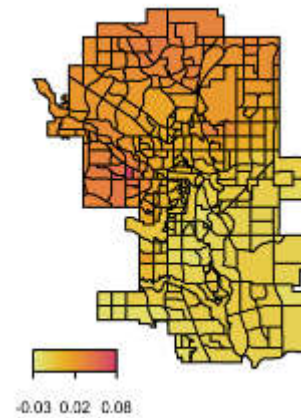


Figure 6: Structured geospatial effect of the location of pipes

5.2.3. Discussions

The application of geoadditive models to predict the failure rate of water pipes confirms the nonlinear effects of continuous covariates on the pipes breakage rate. The errors RMSE (1.74) and MAE (0.74) and the coefficient of determination (0.66) between predicted values and observed show a good agreement between observation and predictions. The range of the *PBR* is between 0.018-167.64). Figure 7 shows a linear correlation between the predictions and observations. Metallic pipes are installed in the interior city of Calgary. The replacement work undertaken led to the installation of PVC pipes on the border.

The challenge in the application of geoadditive models is the optimal definition of the number of knots to avoid overfitting (see Eq.(8)) and the degree of splines to reduce the computational burden. Furthermore, data spreading around the mean is an issue that should be addressed. The analysis of different univariates functions obtained show no particular pattern of a single factor to explain the variation of *PBR*.

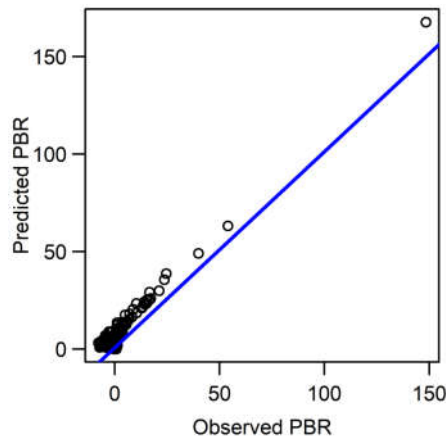


Figure 7: linear correlation between the observed and predicted PBR

6. CONCLUSION

The application of geoadditive regression approach to pipe failure models demonstrated the existence of nonlinear relationships between regressors and response value. Moreover, the geospatial information incorporated in the model enhances the regression model ($R^2 = 0.66$) and also gives visual effect on the areas where the unobserved covariates have high impact on the failure rate. This visual information will serve for further investigations (inspections) to understand and highlight the unknown factors that affect the pipes failure at each location. The predicted failure rate is spatially dependent. Future work will consist in the estimation of risk based critical pipes to assist in the prioritization of maintenance and or inspection.

7. REFERENCES

- Balekelayi, N., and Tesfamariam, S. (under review). Statistical inference of sewer pipes deterioration using Bayesian geoadditive regression model. *Jour. of Infrastr. Syst.*
- Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). Regression: models, methods and applications. London: Springer.
- Kabir, E., Guikema, S., and Kane, B. (2018). Statistical modeling of tree failures during storms. *Reliab. Eng. and Syst. Safety*, 177(May), 68–79.
- Kabir, G., Tesfamariam, S., and Sadiq, R. (2015). Predicting water main failures using Bayesian model averaging and survival modelling approach. *Reliab. Eng. and Syst. Safety*, 142, 498–514.
- Kakoudakis, K., Behzadian, K., Farmani, R., and Butler, D. (2017). Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering. *Urban Water Journal*, 14(7), 737–742.
- Kleiner, Y., and Rajani, B. (2001). Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water*, 3.
- Lang, S., and Brezger, A. (2004). Bayesian P-Splines. *Jour. of Comput. and Graph. Statistics* (Vol. 13).
- März, A., Klein, N., Kneib, T., and Musshoff, O. (2016). Analysing farmland rental rates using Bayesian geoadditive quantile regression. *Eur. Rev. of Agr. Economics*, 43(4), 663–698.
- Rajani, B., and Kleiner, Y. (2001). Comprehensive review of structural deterioration of water mains: Physically based models. *Urban Water*, 3(3), 151–164.
- Rajani, B., and Tesfamariam, S. (2004). Uncoupled axial, flexural, and circumferential pipe-soil interaction analyses of partially supported jointed water mains. *Can. Geotech. J.*, 41(6), 997–1010.
- Scheidegger, A., Leitão, J. P., and Scholten, L. (2015). Statistical failure models for water distribution pipes A review from a unified perspective. *Water Research*, 82, 237–247.
- Scheipl, F., Kneib, T., and Fahrmeir, L. (2013). Penalized likelihood and Bayesian function selection in regression models. *Adv. in Stat. Analysis*, 97(4), 349–385.
- Wilson, D., Filion, Y., and Moore, I. (2015). State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water Journal*, 9006(12), 1–12.
- Wilson, D., Filion, Y., and Moore, I. (2017). State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water Journal*, 14(2), 173–184.
- Wu, Y., and Liu, S. (2017). A review of data-driven approaches for burst detection in water distribution systems. *Urban Water Journal*, 14(9), 972–983.