Corrosion Science 90 (2015) 33–45

Contents lists available at ScienceDirect

Corrosion Science

journal homepage: www.elsevier.com/locate/corsci

Bayesian analysis of external corrosion data of non-piggable underground pipelines

F. Caleyo [a,*], A. Valor [a], L. Alfonso [b], J. Vidal [c], E. Perez-Baruch [d], J.M. Hallen [a]

[a] Departamento de Ingeniería Metalúrgica, ESIQIE, Instituto Politécnico Nacional, Zacatenco, México D.F. 07738, Mexico
[b] Universidad Autónoma de la Ciudad de México, Campus S.L. Tezonco, Iztapalapa, México D.F. 09790, Mexico
[c] Facultad de Física, Universidad de La Habana, San Lázaro y L, Vedado, La Habana 10400, Cuba
[d] Gerencia de Transporte y Distribución de Hidrocarburos, Pemex-PEP, Villahermosa, Tabasco 86038, Mexico

ARTICLE INFO

Article history:
Received 12 May 2014
Accepted 6 September 2014
Available online 2 October 2014

Keywords:
A. Steel
B. Modelling studies
C. Pitting corrosion

ABSTRACT

A new Bayesian methodology for the analysis of external corrosion data of non-piggable underground pipelines has been developed. It allows for the estimation of the statistical distributions of the density and size of external corrosion defects from corrosion data samples taken at excavation sites along the inspected pipeline and can incorporate the detection and measurement errors associated with field inspections. Corrosion data obtained from field inspections of an upstream pipeline and from an in-line inspection of a transportation pipeline are used to illustrate and validate the proposed methodology.

1. Introduction

Bayesian Data Analysis (BDA) has been used in the last decade with varying degrees of success in the assessment of corrosion data for upstream and transportation pipeline systems [1–25]. Previous applications of BDA include identification of risk factors in corroding pipeline systems [1–3], characterization of corrosion defect depth growth [4] and estimation of corrosion rate in operating pipelines [5–12], determination of the sample size required to estimate extreme pit depth in pipelines [13,14], degradation quantification through External Corrosion Direct Assessment (ECDA) [15–19], calibration of in-line inspection (ILI) tools [20,21], identification of failure type in corroded pipelines [22], updating of long-term corrosion estimates of corrosion-fatigue degradation [23,24], and modelling of high pH stress corrosion cracking in underground pipelines [25]. Other structural reliability fields have also profited from the application of Bayesian corrosion data analysis [26].

The main advantage of BDA with regard to corrosion data analysis is that, from a prior belief in the parameters that describe the distributions of corrosion defect size and density, and a relatively small amount of field data, reasonably accurate predictions can be made about the actual distributions of these corrosion parameters. This unique feature is of great interest in the evaluation of the damage caused by external corrosion in underground, non-piggable pipelines, for which the prediction of the size and spatial distribution of active pits remains a very complex task; commonly carried out using small corrosion data samples that feed statistical models such as Extreme Value Statistics [27–29].

The application of BDA in the evaluation of degradation caused by corrosion in non-piggable, underground pipelines has been commonly incorporated into ECDA frameworks [15–19]. The role of BDA in this synergy has traditionally been the estimation of the probability of detection (POD) of the inspection tools, and the estimation of the density (defects per unit length) and depth of active corrosion defects. The main drawbacks of these approaches, which continue limiting the extended application of BDA to corrosion analysis, are the relative complexity of the employed mathematical frameworks and the lack of a thorough description of the implementation details of these schemes (see, for example, [14,15]). There is also a lack of BDA tools for the analysis of field-gathered corrosion data obtained through (random) sampling of non-piggable, underground pipelines.

In this paper, a new BDA methodology is proposed, illustrated and validated for the assessment of external corrosion data obtained from field sampling inspections of non-piggable, underground upstream pipelines. The goal of this methodology is the estimation of the statistical distributions of the density and size of external corrosion defects from a relatively small number of corrosion data randomly taken at excavation sites along the pipeline. The results of a previous field study of external corrosion in different upstream pipeline systems in Southern Mexico [30] are used to suggest the prior, likelihood, and predictive models of the Bayesian analysis. The Bayes rule is used to determine the posterior distributions of the parameters defining the distributions of the density,

* Corresponding author. Tel.: +52 55 57296000x54205; fax: +52 55 57296000x55270.
E-mail address: fcaleyo@gmail.com (F. Caleyo).

http://dx.doi.org/10.1016/j.corsci.2014.09.012
0010-938X/

depth and length of the corrosion defects in the pipeline. The predictive distributions of these corrosion descriptors for the unobserved defects in the entire pipeline are obtained by averaging out the uncertainty in the estimated parameters. The proposed methodology has been validated using corrosion data obtained through ILI and also from data obtained by field inspection of corroding pipelines operating in the same region.

2. Theoretical foundations

2.1. Bayes' theorem

Bayes' theorem lies in the core of any BDA[1] [31]. In it, the strength of belief in parameter values $\boldsymbol{\theta}$ before any data is observed (prior distribution $\pi(\boldsymbol{\theta})$) is combined with the joint probability that the observed data ($\boldsymbol{X}$) follows the chosen model with parameter values $\boldsymbol{\theta}$ (sampling distribution or likelihood function $L\ (\boldsymbol{X}|\boldsymbol{\theta})$) to produce the strength of belief in parameter values $\boldsymbol{\theta}$ when the observed data $\boldsymbol{X}$ have been taken into account (posterior distribution $P_o(\boldsymbol{\theta}|\boldsymbol{X})$). It is important to underline that $\boldsymbol{\theta}$ and $\boldsymbol{X}$ represent a vector of parameter values $\{\boldsymbol{\theta}_i\}$ and a vector of measured data points $\{\boldsymbol{x}_j\}$, respectively.

In its continuous form, Bayes' theorem is written as [31]

$$\mathcal{P}_o(\boldsymbol{\theta}|\boldsymbol{X}) = \frac{\mathcal{L}(\boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \mathcal{L}(\boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{1}$$

where $\boldsymbol{\Theta}$ represents the space for $\boldsymbol{\theta}$, while the marginal likelihood $\int_{\boldsymbol{\Theta}} \mathcal{L}(X|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}$ (also known as evidence) denotes the probability that the data follow the chosen model under marginalization over all parameter values.

If the evidence is thought as a normalization factor, Eq. (1) can be written as [31]

$$\mathcal{P}_o(\boldsymbol{\theta}|\boldsymbol{X}) \propto \mathcal{L}(\boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{1a}$$

Therefore, $\mathcal{P}_o(\boldsymbol{\theta}|\boldsymbol{X})$ can be computed using expression (1a) and then normalized under the requirement that it is a probability density function (pdf). This approach, used throughout this work, considerably reduces the computational workload associated with BDA.

If the prior distribution of a given parameter θ_i is described by the vector of parameters $\boldsymbol{\alpha}$, then it is said that $\{\boldsymbol{\alpha}_k\}$ are the hyperparameters of $\boldsymbol{\theta}_i$.[2] For the general case where hyperparameters are considered, expression (1a) is written as

$$\mathcal{P}_o(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{\alpha}) \propto \mathcal{L}(\boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\alpha}) \tag{1b}$$

2.2. Bayesian prediction

Once the posterior distribution of $\boldsymbol{\theta}$ is estimated using expression (1b), it is possible to make a prediction of the probability of new unobserved data values, conditional on the observed data $\boldsymbol{X}$ and hyperparameters $\boldsymbol{\alpha}$. If the data is assumed to have a distribution M (note that M is also used to construct the sampling distribution or likelihood function), then it is possible to predict the predictive distribution $\mathcal{P}_p(\hat{x}|\boldsymbol{X},\boldsymbol{\alpha})$ of the unobserved data points by averaging out the uncertainty in $\boldsymbol{\theta}$. This is achieved by marginalizing $\mathcal{P}_o(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{\alpha})$ over $\boldsymbol{\theta}$ [31]:

$$\mathcal{P}_p(\hat{x}|\boldsymbol{X},\boldsymbol{\alpha}) = \int_{\boldsymbol{\Theta}} \boldsymbol{M}_X(x|\theta)P_o(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\theta} \tag{2}$$

[1] Bayesian Data Analysis is used in the text to encompass other commons terms such as Bayesian inference, Bayesian updating, Bayesian probability and Bayesian statistics [31].

[2] For example, if the failure rate (λ) of an exponentially distributed variable has a Gamma prior distribution with scale σ and shape ζ parameters, then the hyperparameters of $\boldsymbol{\theta} = \{\lambda\}$ are $\boldsymbol{\alpha} = \{\sigma, \zeta\}$.

3. BDA framework

3.1. Variables of interest

In order to apply BDA to corrosion data, the generic formulation used in the preceding section must be translated into a practical, corrosion-specific formulation involving the corrosion variables, distributions, and parameters to be investigated.

The variables of interest in this study are the depth (d), length (ℓ), and density (n) of the corrosion defects; the latter in defects per excavation site. The models for the data (M), sampling (L), and prior (π) distributions of these variables were proposed from their empirical distributions, which were obtained in an extensive field survey conducted in Southern Mexico over a 7-yr period from 2005 to 2012 [30]. During the field work, corrosion data were gathered at randomly selected ditch sites in five gathering/upstream pipeline systems, totalling 964 km and 620 pipelines. In each one of the 16,636 excavated sites, the depth, length, and number (per site) of the observed external corrosion-caused metal losses were recorded; these variables were obtained for a total of 13,286 external corrosion defects. The field reports also included the trench length and the age, coating type and condition, diameter, wall thickness (pwt), steel grade, and operating pressure of the inspected pipeline [30].

The empirical distributions of the depth and length of the observed external corrosion defects were found to be better described by the Generalized Extreme Value (GEV) distribution, whose pdf is given by the expression [32]:

$$f_{\mathrm{GEV}}(x) = \begin{cases} \frac{1}{\sigma}\exp\left\{-\left[1+\zeta\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\zeta}\right\}\left[1+\zeta\left(\frac{x-\mu}{\sigma}\right)\right]^{-1-1/\zeta}, & \zeta \neq 0 \\ \frac{1}{\sigma}\exp\left\{-\left(\frac{x-\mu}{\sigma}\right) - \exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, & \zeta = 0 \end{cases} \tag{3}$$

where ζ, σ, and μ are the shape, scale, and location parameters, respectively.

The location (μ), scale (σ), and shape (ζ) of the GEV distributions fitted to the measured vectors of data points for the depth, $\boldsymbol{D} = \{\boldsymbol{d}_i\}$, and length, $\Lambda = \{\ell_i\}$, of the observed defects are given in Table 1.

The number n of defects per (2.44 m-long) excavation site was fitted to a Negative Binomial (NegBin) distribution with parameters p and η. The probability mass function (pmf) of n, a nonnegative integer, is [32]

$$f_{\mathrm{NB}}(n) = \binom{\eta+n-1}{\eta-1}p^{\eta}(1-p)^n \tag{3a}$$

It is worth noting that, in this study, η is a positive real-valued number. This kind of generalization is known as Gamma–Poisson mixture or Pólya process [33]. The reasons behind the choice of this form of the NegBin are given and justified in a separate paper under preparation. The parameters of the NegBin distribution fitted to the measured vector of data points of defect density, $\boldsymbol{N} = \{\boldsymbol{n}_i\}$, are also given in Table 1.

3.2. Posterior distributions

Although a certain degree of physical dependence is to be expected to occur between the depth, length, and density of corrosion defects [34], the mathematical burden associated with considering such dependence in a BDA could render it intractable. A key point to make the present BDA as conceptually simple and easy to implement as possible is that these variables can be treated as statistically independent. This approach is not new and has been used by other authors [35,36]. The independence assumption can also be made for the parameters of the corrosion data distributions without incurring in significant errors. Under such assumption,