Structural Safety 56 (2015) 18–29

Contents lists available at ScienceDirect

Structural Safety

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

A clustering approach for assessing external corrosion in a buried pipeline based on hidden Markov random field model

Hui Wang [a], Ayako Yajima [a], Robert Y. Liang [a,*], Homero Castaneda [b,*]

[a] Department of Civil Engineering, The University of Akron, Akron, OH 44325-3905, USA
[b] Department of Material Science and Engineering, National Corrosion Center, Texas A&M University, College Station, TX 77845, USA

ARTICLE INFO

Article history:
Received 9 May 2014
Received in revised form 18 May 2015
Accepted 20 May 2015

Keywords:
Pipeline
External corrosion
Finite mixture model
Hidden Markov random field
Clustering
In-line inspection

ABSTRACT

This paper describes the use of a clustering approach based on hidden Markov random field to extract potential homogeneous segments from a large length right-of-way of a pipeline structure with heterogeneous soil properties. This approach extends the conventional finite mixture model so that the spatial correlation of external corrosion sites can be taken into consideration. An algorithm is established for classifying corrosion defects using soil properties from an *in-situ* survey and location information from in-line inspection reports. The categorized corrosion defects reveal the hidden patterns of corrosion degradation in different segments along a pipeline structure. Stochastic simulation is employed to test this clustering approach. An example involving a 110-km pipeline interval is employed to illustrate the implementation of the clustering approach. The results indicate that the process of external corrosion propagation in a buried pipeline is position-dependent and is highly related to the soil environment. In addition, the results show that this phenomenon can be interpreted by segmentation using the proposed clustering method. A clustering-based inspection strategy is discussed as a way to apply the present approach.

1. Introduction

An underground pipeline is a superior choice for the long-distance transportation of oil, gas and other products for downstream operations in the petroleum industry. However, soil corrosivity can decrease the performance of pipeline protection systems (i.e. direct protection such as coatings, as well as protection from an external source such as cathodic protection) to varying degrees and can produce external corrosion defects that may lead to failure conditions [1].

Metal deterioration in corrosive environments has been well studied for specific types of corrosion and materials during the last few decades. Several approaches (e.g., deterministic approach [2,3], probabilistic models [4–7] and data mining approach [8,9]) are applied to illustrate corrosion propagation on metal surfaces. For the progression of pitting corrosion in pipeline structures, a comprehensive review is available in [10]. It is well known that the corrosivity of the surrounding soil as well as the physicochemical characteristics of materials will greatly affect the degradation rate [11,12]. In the context of a buried pipeline structure, because a pipeline is typically very long, the soil properties along the pipeline right-of-way could vary to a significant degree. Hence, the external corrosion propagation is position-dependent. In terms of probability and stochastic process for assessing the corrosion propagation, a notable limitation of some previous models [13–15] is the lack of consideration of the spatial variability of soil properties.

For considering the spatial variation of the soil corrosivity along a pipeline infrastructure, a practical approach is to divide the pipeline into equal segments named *zones*. Each zone has a set of physical and chemical properties of the soil surrounding the pipeline structure. The corrosion depth can be measured and located with an in-line inspection (ILI) device; hence, each sized defect is linked with a zone and its soil properties. All the soil properties form a high-dimensional feature space, and the defects are mapped into the feature space as vectors. Clustering techniques should be employed to discover the intrinsic structure or hidden pattern of data points in the feature space. As the soil properties within each cluster will have high similarity, it is relatively easy to estimate the probability distribution of the corrosion depth within each cluster. Several clustering techniques have been successfully applied to extreme wind speed analysis [16], structural vulnerability analysis [17,18] in civil engineering and external corrosion in pipeline structures without considering spatial correlation [19]. However,

* Corresponding authors. Tel.: +1 (330) 972 7190; fax: +1 (330) 972 6020 (R.Y. Liang).

E-mail addresses: hw34@zips.uakron.edu (H. Wang), ay11@zips.uakron.edu (A. Yajima), rliang@uakron.edu (R.Y. Liang), hcastaneda@tamu.edu (H. Castaneda).

http://dx.doi.org/10.1016/j.strusafe.2015.05.002
0167-4730/

it is novel to go one step further and apply a clustering approach that considers spatial correlation in assessing external corrosion in a pipeline structure.

Since the corrosion growth of defects that are close to each other may be similar, spatial correlation could apply a certain constraint to the corrosion propagation. The hidden Markov random field (HMRF) is introduced here to handle the spatial constrained problem because it provides a convenient way of formulating an extension of a finite mixture model for spatially dependent data [20]. The merit of the HMRF derives from Markov random field (MRF) theory, in which the spatial information is encoded through the constraints of the neighborhood configuration, as we expect neighboring defects to have the same class label and a similar corrosion degradation process. The configuration of the state sequence (i.e. the sequence of cluster labels assigned to the defects) of HMRF is unobservable, but it is able to be inferred through a random field of features (i.e., the soil properties) [21]. HMRF is a powerful tool for considering spatial correlation in image segmentation [22–25], and in this work, we apply it in the context of assessing the external corrosion of a pipeline.

In the present work, hidden Markov random field theory and the finite Gaussian mixture (FGM) model are employed to develop a clustering approach named the HMRF-FGM model, which can be used to identify and extract segments with homogeneous soil properties from a pipeline region with a large length and a heterogeneous soil environment. This approach is compared to the conventional FGM model [20] to illustrate its accuracy and robustness. By applying HMRF-FGM model, defect depth records reported by two consecutive in-line inspections are classified into groups corresponding to the homogeneous segments. The evolution of corrosion propagation within different segments is studied. A cluster-based inspection strategy for improving the existing maintenance policy is discussed as one of the potential applications of the present approach.

The present paper is organized into six sections. In Section 2, we describe the framework of the clustering approach. A parameter study and robustness assessments of HMRF-FGM are performed in Section 3. The HMRF-FGM is applied to a real life pipeline structure in Section 4. A clustering-based inspection strategy and some potential issues are discussed in Section 5, and conclusions are presented in Section 6.

2. Framework of the clustering approach

In the framework for the clustering approach, we consider $L = \{1, 2, 3, \ldots, l\}$ as the set of states (i.e. the cluster ID of a defect). Let $S = \{1, 2, 3, \ldots, s\}$ be a finite index set, and we shall refer to the set S as the set of locations or sites of all the defects from the start point to the end point along the interval of interest in the pipeline. Now let $X_j = \{x_j | x_j \in L\}$ be a state space of a particular site $j \in S$. Then the product space $\mathbf{X} = \prod_{j=1}^{s} X_j = L^s$ is defined as the configuration space. The state values of all the sites $\mathbf{x} = (x_1, x_2 x_3, \ldots, x_s)$ are a configuration in the product space $\mathbf{X}$. The configuration is considered as a random field $p(\mathbf{x})$. Then, let us consider a second random field $p(\mathbf{y})$, the state space of which is also a product space and is denoted as $\mathbf{Y} = \prod_{j=1}^{s} Y_j, Y_j = \{y_j | y_j \in \mathbf{R}^d\}$, and $\mathbf{y} = (y_1, y_2 y_3, \ldots, y_s)$ is a realization of $p(\mathbf{y})$. In the context of pipeline corrosion, the random variable Y is a random vector of the soil features and the d-dimension space $\mathbf{R}^d$ is the feature space formed by all soil properties.

Suppose a defect at site j is assigned with a certain label l (i.e. defect j belongs to cluster l). The probability of the observation of local soil properties y_j will follow a conditional probability distribution:

$$p(y_j | x_j = l) = f(y_j | \theta_l), l \in L \tag{1}$$

where θ_l is the set of distribution parameters. It is worth noting that for all $l \in L$, the distribution family $f(\cdot; \theta_l)$ has the same known analytical form.

In this work, we employ the finite Gaussian mixture (FGM) model as the conditional probability distribution. Before the derivation of HMRF-FGM model, we must first briefly introduce the FGM model.

2.1. Finite Gaussian mixture model

We suppose that all the defects can be classified into l clusters. The density $f(y_j)$ of Y_j can be written in the form

$$f(y_j | \Phi) = \sum_{i=1}^{l} \pi_i f_i(y_j | \theta_i) \tag{2}$$

where π_i are nonnegative quantities that sum to one and are referred to as the *mixing proportions* or *weights*. π_i are independent of the individual defect $j \in S$. The $f_i(y_j | \theta_i)$ are referred to as the *component densities*. Here, we use multivariable Gaussian distribution as the *component density* i.e. $\theta_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), i \in L$ where $\boldsymbol{\mu}$ is a vector that contains the mean of every feature and $\boldsymbol{\Sigma}$ is the covariance structure of all the features, which is defined in [26]. We take Φ as the parameter set of the mixture model $\Phi = \{\pi_i, \theta_i | i \in L\}$. This is so-called finite Gaussian mixture model.

The standard finite Gaussian mixture model presented in Eq. (2) is widely used in unsupervised clustering [20]. However, it is not considered to be a complete model in spatially-constrained problem, as it only describes the data statistically and there is no spatial information involved in this framework. In Eq. (2), we notice that the *mixing proportions* are spatially independent, which means the probability of a defect's belonging to a certain cluster is homogenous along the entire pipeline interval. To overcome this drawback of the FGM model, a clustering model that is adaptive to spatial information based on hidden Markov random field model will be presented.

2.2. Hidden Markov random field – Finite Gaussian mixture model

A typical HMRF has the following three characteristics [25]:

(1) The configuration of all sites $\mathbf{x} = (x_1, x_2 x_3, \ldots, x_s)$, i.e. the clustering result, is unobservable.
(2) The observable random field $p(\mathbf{y}|\mathbf{x})$ is called the observed or emitted random field given any particular configuration $\mathbf{x} = (x_1, x_2 x_3, \ldots, x_s)$.
(3) A typical conditionally independent assumption is adopted that has the expression:

$$p(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{s} p(y_j | x_j) \tag{3}$$

The conditionally independent assumption still guarantees the *Markovianity* of the random field $p(\mathbf{x}, \mathbf{y})$ [21]. More complicated assumptions might also be considered [20].

Based on the above, the joint probability of (X, Y), i.e. a clustering configuration and the corresponding observations of soil properties, has the form

$$P(\mathbf{x}, \mathbf{y}) = P(\mathbf{x}) P(\mathbf{y}|\mathbf{x}) = P(\mathbf{x}) \prod_{j=1}^{s} p(y_j | x_j) \tag{4}$$

According to Gibbs models [22], the probability distribution function of a certain configuration of sites $\mathbf{x} = (x_1, x_2 x_3, \ldots, x_s)$ has the form: