Construction and Building Materials 116 (2016) 226–234

Contents lists available at ScienceDirect

Construction and Building Materials

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

A statistical based framework for predicting field cracking performance of asphalt pavements: Application to top-down cracking prediction

Shihui Shen [a], Weiguang Zhang [b,*], Lin Shen [c], Hai Huang [a]

[a] Division of Business and Engineering, Pennsylvania State University, Altoona, PA 16601, United States
[b] Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611, United States
[c] Department of Civil and Environmental Engineering, University of Hawaii, Honolulu, HI 96822, United States

HIGHLIGHTS

- A statistical based framework is introduced to develop cracking model for pavement.
- The framework is capable of taking into account confounding variables in the field.
- A binary logistic probability model allows predicting the cracking initiation potential.
- The initiation and propagation of top-down cracking may follow different mechanisms.

ARTICLE INFO

Article history:
Received 12 December 2015
Received in revised form 23 April 2016
Accepted 28 April 2016
Available online 4 May 2016

Keywords:
Statistical models
Top-down cracking
Pavement performance
Partial least squares regression
Binary logistic regression

ABSTRACT

Most existing cracking performance models of asphalt pavements, such as top-down cracking models, are mechanistic or mechanistic-empirical based. These models usually focus on a specific type of cracking mechanism. The prediction quality of these models can also vary when field cracking conditions are complicated and clear identification of the distress type is difficult. Literature suggests that a statistical based method can account for variability and a large number of influencing factors, and could be a promising alternative. This paper aims to introduce a statistical based framework for performance prediction using top-down cracking as an example. Such a framework can be modified and implemented by local agencies for a variety of cracking distresses based on specific needs and requirements.

Detailed steps of the statistical framework are presented through the development of top-down cracking models. Results indicate that the framework works effectively by integrating pavement performance concepts with several statistical methods including Partial Least Squares (PLS) Regression, Binary Logistic Regression, and Leave one out cross validation (LOOCV). Using the developed statistical based framework, critical factors that may affect the initiation and propagation of top-down cracking are identified, and their sensitivities to the field top-down cracking are further discussed.

1. Introduction

Most existing cracking performance models for asphalt pavements are mechanistic or mechanistic-empirical based. Using top-down cracking as an example, one of the most widely used prediction models is based on the cumulative damage concept given by Miner's. The damage is calculated as the ratio of the predicted number of traffic repetitions to the allowable number of load repetitions. The model was calibrated using field pavement performance data collected by Long-Term Pavement Performance (LTPP) program and was integrated into AASHTOWare Pavement ME Design program [1]. Lytton et al. [2] used fracture mechanics based upon the Paris law to model the crack propagation stage in the theoretical Superpave Model. Roque and his co-authors [3,4] used a viscoelastic continuum damage (VECD) model to predict crack initiation and an HMA fracture mechanics (HMA-FM) model to predict crack propagation. Several important material property sub-models, including aging, healing and moisture damage were developed and incorporated into the models [5]. These models have provided important insight into pavement failure mechanisms, and have shown various application values under different scenarios.

However, literatures also reported that the majority of current top-down cracking models do not align well with field pavement

* Corresponding author.
E-mail addresses: szs20@psu.edu (S. Shen), weiguang.zhang@essie.ufl.edu (W. Zhang), linshen@hawaii.edu (L. Shen), huh16@psu.edu (H. Huang).

http://dx.doi.org/10.1016/j.conbuildmat.2016.04.148
0950-0618/

performance [6–9]. Similar findings have also been reported for other cracking performance models such as transverse cracking models and bottom-up fatigue cracking models [7,10,11]. These discrepancies could be the result of common issues for these prediction models. For example: (1) a typical performance model only addresses a specific type of cracking mechanism, (2) a mechanistic based model usually has limited capacity for variations in construction/climate in the field, and (3) use of a single model may yield poor prediction quality for field performance due to the complexity of distress conditions. Isolating a single damage mechanism is very difficult.

Statistical methods have been used in pavement engineering fields to develop performance prediction models and identify influencing factors. Multiple Linear Regression (MLR) is one of the most widely used statistical methodologies for the development of prediction models of transverse cracking, rut depth and other field distresses [12–14]. Other statistical methodologies include artificial neural networks [15], Spatial Statistics analysis [16], and Joint Estimation [17]. Each method may have specific advantages. For instance, Spatial Statistics analysis can consider pavement performance difference at different points, while Joint Estimation can be used to address problems in which different field or laboratory data sources have different levels of precision.

Statistical methods also provide several techniques to allow users to identify the effect of variables on responses based on varied scenarios [18] such as large/small sample size, with/without collinear data and different types of variable. As an example, Analysis of Variance (ANOVA) has been widely used to identify the significant input parameters for prediction models of transverse cracking, top-down cracking, bottom-up cracking and rut depth [13,19,20].

Through meaningful and well-controlled data collection, organization, analysis, and interpretation, statistics can provide prediction models that work well for both the selected response, and the new introduced response.

Therefore, this paper has two major objectives:

(a) Introduce a statistical based framework for cracking performance prediction that can be modified and implemented by local agencies based on specific needs and requirements, and
(b) Use the top-down cracking model as an example to explain the application of the statistical framework for the development of a prediction model.

2. Methodology and general framework

A number of statistical methods can be used to determine the relationship between predictor variables and responses, among which Multiple Linear Regression (MLR) is widely used in pavement engineering. However, if collinearity exists between variables, MLR could result in an over-fitting model that cannot accurately predict new responses [21]. Even worse, MLR could give incorrect signs for parameters [22]. Here, the collinearity is defined as a high level of correlation between two predictor variables. Although collineated data can manually be removed from the database, it is not ideal because some meaningful data might be unintentionally discarded [23]. Alternatively, Partial Least Squares (PLS) method could be used. The PLS method is able to solve collinearity problems, and also ensures that only highly contributing parameters of the prediction model are used. In addition, the PLS method is suitable for relatively small datasets. This is particularly useful for pavement performance prediction since obtaining large field performance and laboratory data is usually very time consuming and costly. Based on the above considerations, plus other advantages like simplicity, reliability and versatility, Zhang et al. [24] applied the PLS method for field transverse cracking prediction and results appeared to be reasonable.

Validation of the effectiveness of the prediction model is another critical component in the model development. There are many ways to validate the prediction model statistically, such as k-fold Cross Validation (CV) and Leave One out Cross Validation (LOOCV). LOOCV is used in this study since it is always complete, and every single data point can be used. Using the PLS method in conjunction with LOOCV, the model simultaneously decides upon the optimum variable number of independents, determine the coefficients of model parameters, and validate the model's capability of predicting new performance.

As suggested by literatures [4,24], the initiation and propagation of pavement cracking could follow different mechanisms. A pavement might develop its first crack at quite an early stage while the severity of the crack (length, width, and quantity of crack) does not increase significantly for a long time. On the other hand, it may also be possible that a pavement does not crack for an extended period of service time; but once cracks begin to propagate, they do so at a very fast pace. Therefore, separate models for the initiation and propagation of cracking models are more ideal.

For crack initiation, a probabilistic based model is more suitable than a deterministic based model [25]. Instead of providing a "yes" or "no" answer to the prediction of crack initiation, the probability for a pavement to crack under a particular condition (material, environment, structure, etc.) can be determined. Depending on the probability results, engineers and/or local agencies can set up different threshold values (for example, greater than 70% probability threshold for interstate highways or greater than 85% for local roads) as a limit of crack initiation to aid decision-making.

In this paper, the Binary Logistic (BL) Regression method is used together with the PLS Regression method to develop a probabilistic based crack initiation model for top-down cracking.

For crack propagation, conventional deterministic based models that predict the quantity of cracks (length of cracks) is developed and calibrated by using the PLS and the LOOCV methods. The PLS regression attempts to extract only the variables that account for most of the variation in the response. Considering the nature of the PLS regression to solve collinearity problems, it helps users to find significant variables from different types such as pavement structure, material properties and climate, instead of focusing on one type of variable. A linear regression model can be constructed based on the selected variables and responses.

The development of statistical based cracking prediction models mainly involves four steps:

- Data collection. Potential data that can be included into the analysis include but not limited to:
  - Dependents: pavement performance data such as crack area and crack length.
  - Independents: potential influencing factors for particular cracking distress such as climate, material properties, traffic, pavement structures, etc. These data should be included based on engineering experience and could vary due to data availability.
- Data preprocessing. The existence of collinearity, the randomness of data distribution, and other assumptions for the statistical analysis must be firstly checked to determine the type of regression method to be used, the needs of data transformation, and other data preprocessing needs.
- Model development. Statistical analysis is performed to select optimum number of independents, develop model parameters, and calibrate the models. This step usually can be performed using a commercial statistical software such as SPSS and Minitab. Crack initiation model and crack propagation model are constructed respectively.