



KNOWLEDGE DISCOVERY AND DATA MINING IN PAVEMENT INVERSE ANALYSIS

Kasthurirangan Gopalakrishnan¹, Ankit Agrawal², Halil Ceylan³,
Sunghwan Kim⁴, Alok Choudhary⁵

^{1,3,4}Iowa State University, Ames, IA, USA

^{2,5}Northwestern University, Evanston, IL, USA

E-mails: ¹rangan@iastate.edu (corresponding author); ²ankitag@eecs.northwestern.edu;

³hceylan@iastate.edu; ⁴sunghwan@iastate.edu; ⁵choudhar@eecs.northwestern.edu

Submitted 15 December 2011; accepted 29 March 2012

Abstract. This paper describes the use of data mining tools for predicting the non-linear layer moduli of asphalt road pavement structures based on the deflection profiles obtained from non-destructive deflection testing. The deflected shape of the pavement under vehicular loading is predominantly a function of the thickness of the pavement layers, the moduli of individual layers, and the magnitude of the load. The process of inverse analysis, more commonly referred to as 'backcalculation', is used to estimate the elastic (Young's) moduli of individual pavement layers based upon surface deflections. A comprehensive synthetic database of pavement response solutions was generated using an advanced non-linear pavement finite-element program. To overcome the limitations associated with conventional pavement moduli backcalculation, data mining tools such as support vector machines, neural networks, decision trees, and meta-algorithms like bagging were used to conduct asphalt pavement inverse analysis. The results successfully demonstrated the utility of such data mining tools for real-time non-destructive pavement analysis.

Keywords: road, transportation, artificial neural network (ANN), infrastructure, statistical analysis.

Introduction

Since the 1960s, nondestructive deflection testing has been used to assess the structural capacity and integrity of pavement sections. For the last two decades, the predominant form of deflection testing for both project-level and network-level pavement evaluation has been the falling weight deflectometer (FWD, Fig. 1). Typically, the FWD deflection measurements are used to estimate the in-situ elastic moduli of each pavement layer as material input parameters for rehabilitation and overlay design (Alavi *et al.* 2008).

A conventional Asphalt Concrete (AC) pavement is typically consists of three layers: a surface layer paved with AC mixture (known as surface course or wearing course), a granular base made up of relatively high-quality aggregates (base course), and a subgrade layer made up of existing soil. Sometimes, an optional sub-base layer comprised of relatively low-quality aggregates is also included. The deflection of a pavement represents the combined system response of the pavement layers to an applied load. Based on this mechanical concept, the *in situ* moduli of individual layers can be estimated from FWD measurements through appropriate analysis meth-

ods. This procedure is referred to as pavement modulus backcalculation. The backcalculation of layer modulus of asphalt pavement has been recognized as a complex problem (Sharma, Das 2008).



Fig. 1. Close-up of truck-mounted falling weight deflectometer (FWD) used in non-destructive testing (NDT) of pavements

In recognition of the limitations of the current American Association of State Highway and Transportation Officials (AASHTO) pavement design guide which are based on empirical regression techniques relating simple material characterizations, traffic characterization and measures of performance, mechanistic-empirical (M-E) pavement design and analysis approaches have been developed. For example, the new AASHTO pavement design guide is the new Mechanistic Empirical Pavement Design Guide (MEPDG) and its software developed through National Cooperative Highway Research Program (NCHRP) 1-37 A project (NCHRP 2004).

The mechanistic part of M-E design is the application of the engineering mechanics principles to calculate pavement responses (stresses, strains, and deflection) under loads for the prediction of the pavement performance history. The empirical nature of the M-E design stems from the fact that the laboratory-developed pavement performance models are adjusted or calibrated to the observed performance measurements (distress) from the actual pavements. With the evolution and adoption of mechanistic-empirical pavement design, the need to obtain reliable material properties has increased. Further, when new materials are being used in the rehabilitation design (such as for an asphalt concrete overlay), a combination of laboratory-measured properties for some layers and field-derived parameters for others may result. While the field-derived parameters may be valuable, in the sense of characterizing the damaged in-situ characteristics, the values may be seemingly in conflict with the laboratory values for new materials.

The interpretation of FWD data to characterize material properties in pavement structure are carried out using empirical equations or correlation and/or the use of mechanistic based approaches. The mechanistic-based approaches under the umbrella of ‘backcalculation’ refer to the calculation of the pavement layer properties which best describe the measured deflection in layered elastic or finite element models to represent the pavement system. The FWD backcalculation procedure involves two calculation directions, namely forward and inverse. In the forward direction of analysis, theoretical

deflections are computed under the applied load and the given pavement structure using assumed pavement layer moduli. In the inverse direction of analysis, these theoretical deflections are compared with measured deflections and the assumed moduli are then adjusted in an iterative or an optimization procedure until theoretical and measured deflection basins match acceptably well. The moduli derived in this way are considered representative of the pavement response to load, and can be used to calculate stresses or strains in the pavement structure for analysis purposes. This is an iterative or an optimization method to solve the inverse problem, and will not have a unique solution for most cases.

Although several traditional and non-traditional pavement backcalculation techniques have been proposed over the years (Gopalakrishnan et al. 2010), which are briefly reviewed later, researchers are always interested in exploring advanced techniques that have the potential of more accurately characterizing pavement system responses.

1. Objective and Scope

The primary objective of this paper is to introduce some of the advanced data mining tools to the pavement community and examine their usefulness in solving an inverse problem encountered in the non-destructive condition evaluation of existing pavements.

Conventional three-layered flexible (asphalt) pavements are considered in this paper although the overall methodology is applicable to other pavement types. A 2-D Finite Element (FE) flexible pavement response model is used to generate a comprehensive synthetic database of pavement surface deflections corresponding to a wide range of pavement layer moduli and thicknesses. Advanced data mining tools are used to develop pavement layer moduli prediction (backcalculation) models based on deflection and thickness inputs. The predictive models are then applied to actual FWD deflection data acquired in the field to demonstrate their validity and robustness for real-time non-destructive pavement structural evaluation. The overall proposed approach described in this paper is illustrated in Fig 2.

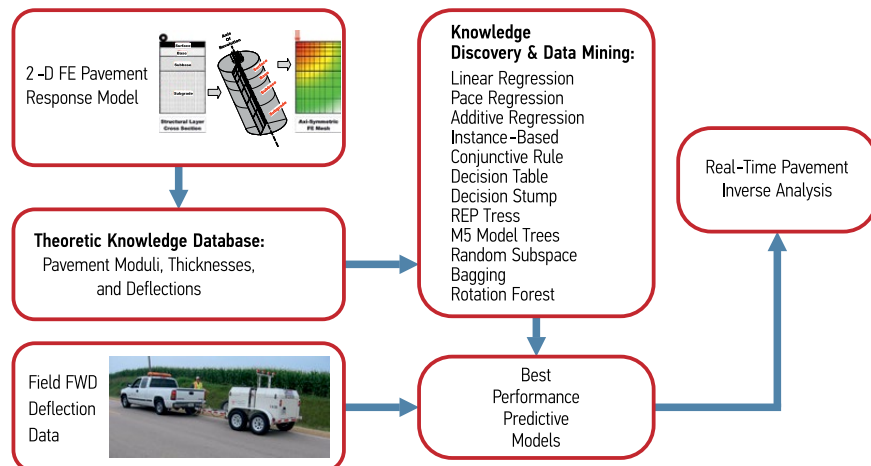


Fig. 2. Schematic of overall proposed approach to real-time flexible pavement inverse analysis with knowledge discovery and data mining

2. Pavement Inverse Analysis: Existing Traditional and Non-Traditional Approaches

2.1. Traditional Approaches

A number of the backcalculation approaches with software programs for flexible pavements have been developed over the years to backcalculate material properties from FWD data.

The traditional backcalculation approaches are briefly discussed as follows:

- backcalculation equations: regression equations have been developed to predict subgrade modulus using the deflection testing data (Newcomb 1987). The 1993 design guide (AASHTO 1993) presents a backcalculation equation for subgrade modulus;
- equivalent thickness concept;
- optimization and iterative methods.

The AREA method for flexible pavements (Hoffman, Thompson 1981), AREA method for rigid pavements (Ioannides *et al.* 1989; Ioannides 1990; Barenberg, Petros 1991), ILLI-SLAB (Foxworthy, Darter 1989), ILLI-BACK (Ioannides 1990), best fit algorithm (Hall *et al.* 1997, Smith *et al.* 1998), ELMOD (Ullidtz 1987), WESDEF (Cauwelaert *et al.* 1989), DIPLOBACK (Khazanovich, Roesler 1997), and MODCOMP (Irwin, Szenbenyi 1991; Irwin 1994) are examples of FWD interpretation programs and algorithms for rigid, flexible, and composite pavements.

Backcalculation programs based on multilayer elastic layer theory are generally used for AC pavements. For rigid pavements, plate theory for a slab resting on a Winkler foundation or elastic solid foundation is modeled. There is no widely accepted methodology for AC overlaid PCC-type of composite pavements on a Winkler foundation. The backcalculation programs WESDEF, BISDEF, and ELSDEF are based on multilayer elastic analysis programs WESLEA, BISAR and ELSYM, respectively. These programs require the thickness, Poisson's ratio, and a seed modulus as inputs. The forward elastic layer program iterates the given seed modulus until the observed deflections match with calculated deflections. Thus, the modulus of pavement layer is highly affected by the seed modulus. Consequently, experienced engineers are required to use these backcalculation programs (Lytton 1989).

2.2. Non-Traditional Approaches

The use of a new class of computational intelligence paradigm, known as soft computing techniques, in the field of geomechanical and pavement engineering has steadily increased over the past decade owing to their ability to admit approximate reasoning, imprecision, uncertainty and partial truth (Gopalakrishnan *et al.* 2010). Since real-life infrastructure engineering decisions are made in ambiguous environments that require human expertise, the application of soft computing techniques has been an attractive option in pavement and geomechanical modeling.

The term 'soft computing' applies to variants of and combinations under the four broad categories of evolu-

tionary computing, artificial neural networks (ANNs), fuzzy logic, and Bayesian statistics. Although each one has its separate strengths, the complementary nature of these techniques when used in combination (hybrid) makes them a powerful alternative for solving complex problems where conventional mathematical methods fail.

Among various soft computing techniques, the interests in ANNs have been increased for use in pavement systems applications over the past 15 years (Use of Artificial Neural... 1999). There have been several successful studies of using ANNs to predict the pavement layer moduli using the falling weight deflectometer (FWD) deflection data (Gucunski, Krstic 1996; Khazanovich, Roesler 1997; Kim, Y., Kim, Y. R. 1998; Meier, Rix 1994). The NCHRP1-37A research project team in charge of developing the Mechanistic-Empirical Pavement Design Guide (MEPDG) incorporated the ANN models (Ceylan 2002) in preparing the MEPDG concrete pavement analysis package. Recently, data mining tools are attracting attention among researchers in various fields for discovering knowledge and underlying relationships in simulated or actual data (Miradi 2009).

3. Data Mining Tools: Brief Review

3.1. Linear Regression

Linear regression probably the oldest and most widely used predictive model, which commonly represents a regression that is linear in the unknown parameters used in the fit. The most common form of linear regression is least squares fitting (Weher 1977).

3.2. Pace Regression

It evaluates the effect of each feature and uses a clustering analysis to improve the statistical basis for estimating their contribution to overall regression. It can be shown that pace regression is optimal when the number of coefficients tends to infinity. We use a version of Pace Regression described in (Wang 2000; Wang, Witten 2002).

3.3. Additive Regression

It is a meta learner that enhances the performance of a regression based classifier. Each iteration fits a model to the residuals left by the classifier on the previous iteration (Friedman 2002). The predictions of each of the learners are added together to get the overall prediction. It is generally used with Decision Stump as the base learner.

3.4. Instance-Based

This is a lazy classification technique which implements nearest-neighbour classifier. It uses normalized Euclidean distance to find the training instance closest to the given test instance, and predicts the same class as this training instance (Aha *et al.* 1991).

3.5. Conjunctive Rule

This is a rule-based learner that can predict both numeric and nominal class labels. The goal of rule induction

is to induce rules from data capturing all generalizable knowledge within it, while being as small as possible (Cohen 1995).

3.6. Decision Table

Decision table typically constructs rules involving different combinations of attributes, which are selected using an attribute selection search method. Simple decision table majority classifier (Kohavi 1995) has been shown to sometimes outperform state-of-the-art classifiers.

3.7. Decision Stump

A decision stump (Witten *et al.* 2011) is a weak tree-based machine learning model consisting of a single-level decision tree with a categorical or numeric class label. Decision stumps are usually used in ensemble machine learning techniques.

3.8. Artificial Neural Networks (ANNs)

ANNs are networks of interconnected artificial neurons, and are commonly used for non-linear statistical data modeling to model complex relationships between inputs and outputs. Several good descriptions of neural networks are available (Bishop 1996; Fausett 1993).

3.9. Support Vector Machines

SVMs are based on the Structural Risk Minimization (SRM) principle from statistical learning theory. A detailed description of SVMs and SRM is available in (Vapnik 1995). In their basic form, SVMs attempt to perform classification by constructing hyperplanes in a multidimensional space that separates the cases of different class labels. It supports both classification and regression tasks and can handle multiple continuous and nominal variables.

3.10. Reduced Error Pruning Trees

REPTree (Witten *et al.* 2011) is a implementation of a fast decision tree learner. REPTree builds a decision/regression tree using information gain/variance and prunes it using reduced-error pruning (with backfitting). It deals with missing values by splitting the corresponding instances into pieces.

3.11. M5 Model Trees

M5 Model Trees (Wang, Witten 1997) are a reconstruction of Quinlan's M5 algorithm (Quinlan 1992) for inducing trees of regression models, which combines a conventional decision tree with the option of linear regression functions at the nodes. It also uses the techniques used in CART (Breiman *et al.* 1984) to effectively deal with enumerated attributes and missing values.

3.12. Random SubSpace

The Random Subspace classifier (Ho 1998) constructs a decision tree based classifier that also consists of multiple trees. It tries to achieve a balance between over fitting and achieving maximum accuracy. The algorithm main-

tains highest accuracy on training data and improves on generalization accuracy as it grows in complexity.

3.13. Bagging

Bagging (Breiman 1996) is a meta-algorithm to improve the stability of classification and regression algorithms by reducing variance. Bagging is usually applied to decision tree models to boost their performance. It involves generating a number of new training sets (called bootstrap modules) from the original set by sampling uniformly with replacement. The bootstrap modules are then used to generate models whose predictions are averaged to generate the final prediction.

4. Theoretic Database Development

The synthetic data used in conducting pavement inverse analysis with data mining in this study were generated from a two-dimensional axi-symmetric pavement FE software developed at the University of Illinois at Urbana-Champaign (Raad, Figueroa 1980). It incorporates stress-sensitive geo-material models and has been reported to provide a more realistic representation of the flexible pavement structure and its response to loading. Numerous research studies have analyzed and validated this FE model's AC pavement structural response prediction for highway and airfield pavements (Thompson, Elliott 1985; Garg *et al.* 1998).

FWD tests are generally performed by dropping a 9000-lb (40-kN) load on the top of a circular plate, in contact with the pavement surface, with a radius of 150 mm (6 inches). Deflections are measured at offsets of 0 (D_0), 300 (12) (D_{12}), 600 (24) (D_{24}), 900 (36) (D_{36}), 1200 (48) (D_{48}), 1500 (60) (D_{60}) mm (inches) from center of loading plate. The FWD loading was simulated using the flexible pavement FE program.

The AC surface layer was treated as linear elastic material with Young's Modulus, E_{ac} , and Poisson ratio, μ . Stress-dependent elastic models along with Mohr-Coulomb failure criteria were applied for the unbound aggregate base and fine-grained soil subgrade layers. The (stress-hardening) $K_b-\theta$ model (Hicks, Monismith 1971) was used for the base layer ($E_R = K_b \cdot \theta^n$; E_R is resilient modulus (psi), θ is bulk stress (psi) and K and n are statistical parameters). Based on extensive testing of unbound aggregate materials, (Rada, Witczak 1981) proposed the following relationship between K and n : $\log_{10}(K_b) = 4.657 - 1.807 \cdot n$. The (stress-softening) bilinear model (Thompson, Robnett 1979) was used for the subgrade layer.

Asphalt concrete modulus E_{ac} , granular base $K-\theta$ model parameter K , and the subgrade soil break point deviator stress E_{ri} in the bilinear model were used as the layer stiffness inputs for all the different conventional flexible pavement FE simulations. The 40-kN (9-kip) wheel load was applied as a uniform pressure of 550 kPa (80 psi) over a circular area of radius 6 in. The thickness and moduli ranges used in the database generation are provided elsewhere (Ceylan *et al.* 2007).

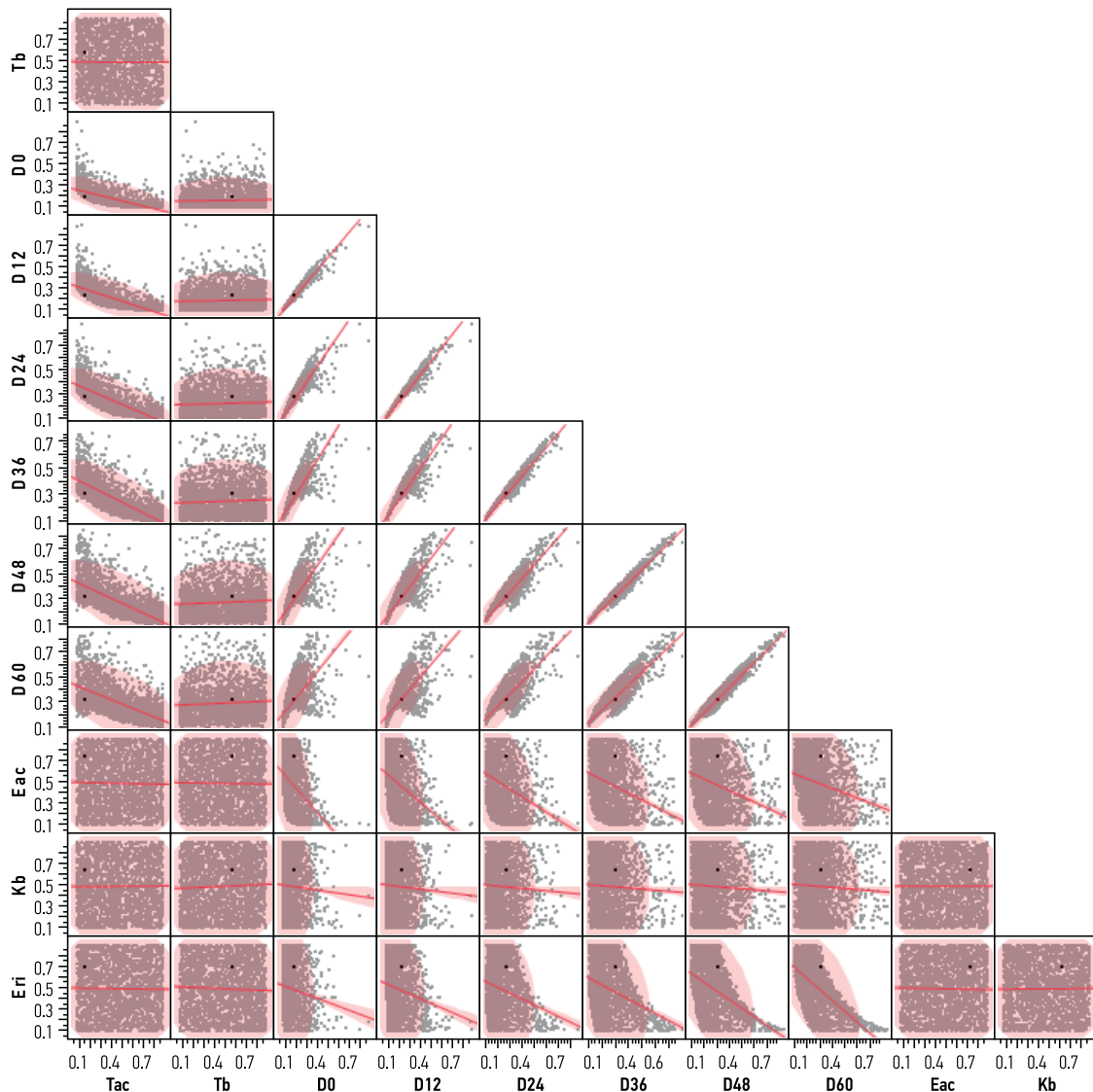


Fig. 3. Scatterplot matrix of input and output variables in the theoretic database

A total of 30000 FE runs were conducted by randomly choosing the pavement layer thicknesses and input variables within selected ranges to generate a knowledge database for inverse analysis using data mining tools. All the datasets were normalized within the range of 0.1 to 0.9 to facilitate learning. A scatterplot for each pair of variables (pavement layer thickness, surface deflections and layer moduli values) from the synthetic database used in data mining is displayed in a matrix arrangement and compiled in Fig. 3.

5. Experiments and Discussion of Results: Theoretic Data

A suite of data mining tools discussed in a previous section was employed in the experimental runs using theoretic data. The goal was to identify the best-performance predictive models which could be applied on the actual field FWD data for real-time inverse analysis of pavements. The following variables define the inputs and

outputs in the knowledge discovery and data mining process:

- inputs: Surface deflections ($D_0, D_{12}, D_{24}, D_{36}, D_{48},$ and D_{60}); AC layer thickness (T_{ac}); and base layer thickness (T_b);
- outputs: Modulus of the AC surface layer (E_{ac}); Modulus of the base layer (K_b); and Modulus of the Subgrade layer (E_{ri}).

Thus, data mining based backcalculation models were developed with eight input parameters and one output parameter per model. However, the unbound aggregate base layer modulus could not be predicted using just the eight inputs (deflections and thicknesses). Therefore, in the development of K_b backcalculation model, the predicted E_{ac} and E_{ri} were used as additional inputs along with the six FWD deflections as well as the thicknesses of the AC surface and base layer. The results for both scenarios are discussed later in the paper.

The data were divided randomly into two different subsets of the training data subset and the testing

data subset in such a way that they are representative of same statistical population. Both datasets were normalized within the range of 0.1 to 0.9 for input and output values to facilitate the training process. The training data subset was used for model learning and the testing data subset was used to examine the statistical accuracy of the developed models. Further, 5-fold cross-validation was employed to increase the robustness of prediction accuracy and avoid any over-training. The R (R: A Language and Environment... 2009) and WEKA (Hall et al. 2009) software toolkits were used in this study for data mining.

Quantitative assessments of the degree to how close the models could predict the actual outputs are used to provide an evaluation of the models' predictive performances. A multi-criteria assessment with various goodness-of-fit statistics was performed using all the data vectors to test the accuracy of the trained models. The criteria that are employed for evaluation of models' predictive performances are the coefficient of correlation (R), Mean Absolute Error (MAE), and Root-Mean-Squared Error ($RMSE$) between the actual and predicted values. The definitions of these evaluation criteria are as follows:

$$R = \frac{\sum_{i=1}^n (y_i^t - \bar{y}^t)(y_i^p - \bar{y}^p)}{\sqrt{\sum_{i=1}^n (y_i^t - \bar{y}^t)^2} \sqrt{\sum_{i=1}^n (y_i^p - \bar{y}^p)^2}}; \quad (1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i^t - y_i^p}{y_i^t} \right|; \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^t - y_i^p)^2}{n}}, \quad (3)$$

where: y_i^t and y_i^p are the target and predicted modulus values, respectively; \bar{y}^t and \bar{y}^p are the mean of the target and predicted modulus values corresponding to n patterns.

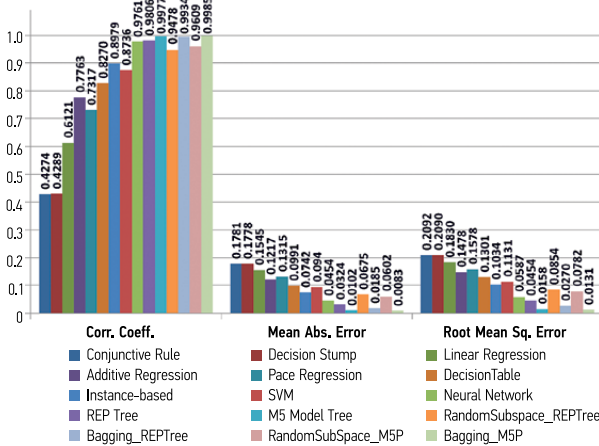


Fig. 4. Summary of asphalt layer moduli (E_{ac}) prediction performance with data mining techniques using theoretic deflection basins

R is a measure of correlation between the predicted and the measured values and therefore, determines accuracy of the fitting model (higher R equates to higher accuracy). The MAE and $RMSE$ indicate the relative improvement in prediction accuracy. Relative smaller magnitudes indicate better prediction accuracy.

The values of performance statistics for the developed data mining based inverse prediction models are summarized in Figs 4–7, for E_{ac} , E_{ri} , and K_b . It is observed that excellent performance is achieved using REPTree and M5 Model trees as underlying regression algorithms with Bagging meta-learner for all three pavement layer moduli. Among the three pavement layers, the prediction accuracy for K_b is the worst as expected even after including E_{ac} and E_{ri} as additional inputs. This is further confirmed by the prediction error histograms plotted in Fig. 8 for E_{ac} , K_b (using E_{ac} and E_{ri} as additional inputs), and E_{ri} using Bagging_M5P (Bagging meta-learning technique with M5 model trees as the base learner), for instance. The Bagging_M5P predictor was chosen as the best-performance data mining predictive technique to be used in real-time pavement inverse analysis described in the next section.

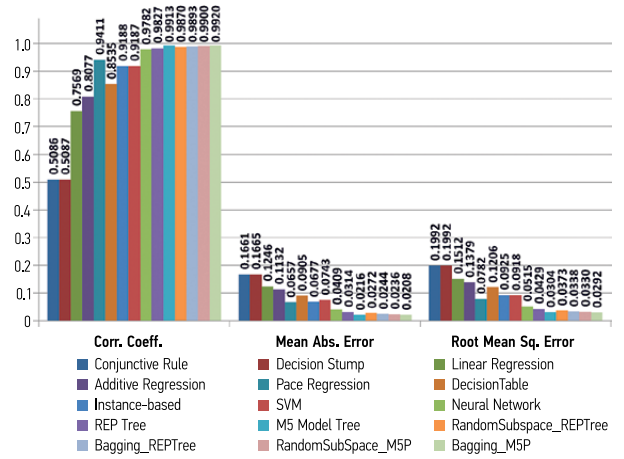


Fig. 5. Summary of subgrade layer moduli (E_{ri}) prediction performance with data mining techniques using theoretic deflection basins

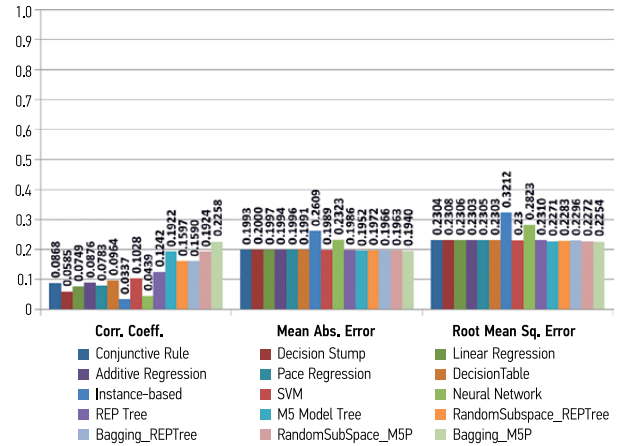


Fig. 6. Summary of base layer moduli (K_b) prediction performance with data mining techniques (without including E_{ac} and E_{ri} as additional inputs) using theoretic deflection basins

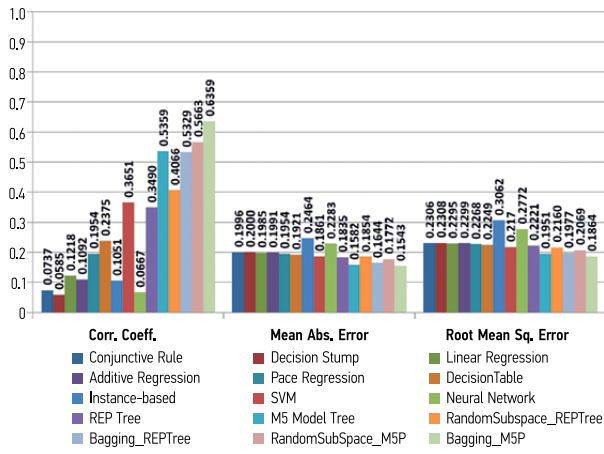


Fig. 7. Summary of base layer moduli (K_b) prediction performance with data mining techniques (including E_{ac} and E_{ri} as additional inputs) using theoretic deflection basins

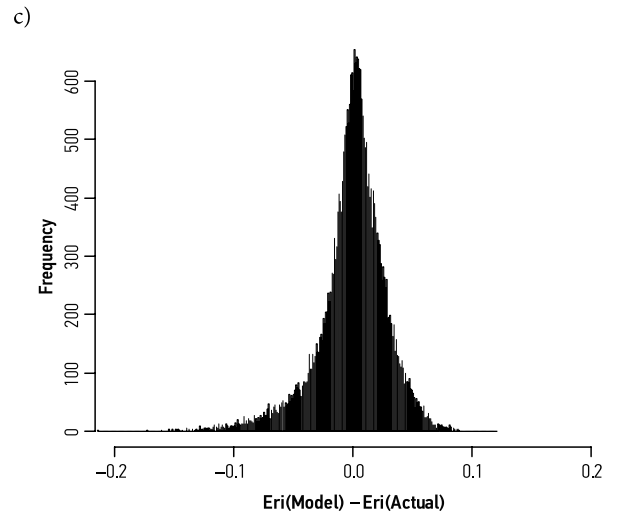
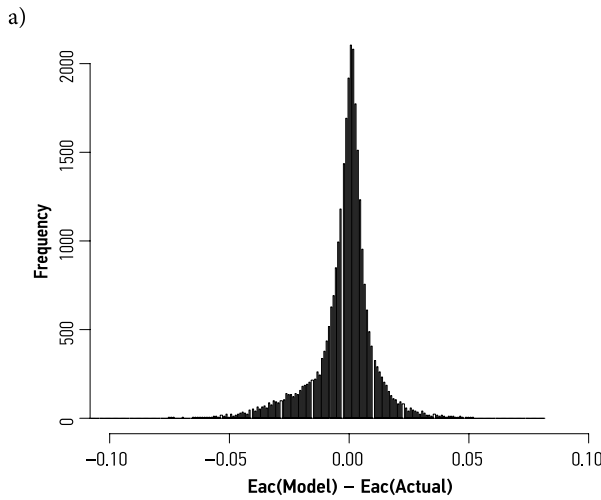
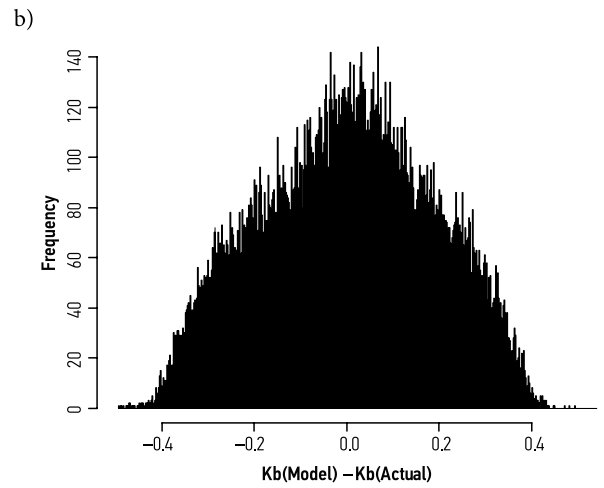


Fig. 8. Histograms of prediction errors with Bagging_M5P: a - E_{ac} ; b - K_b (using E_{ac} and E_{ri} as inputs); c - E_{ri}

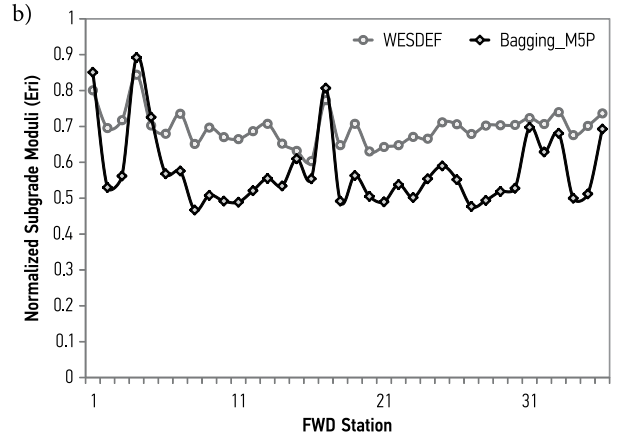
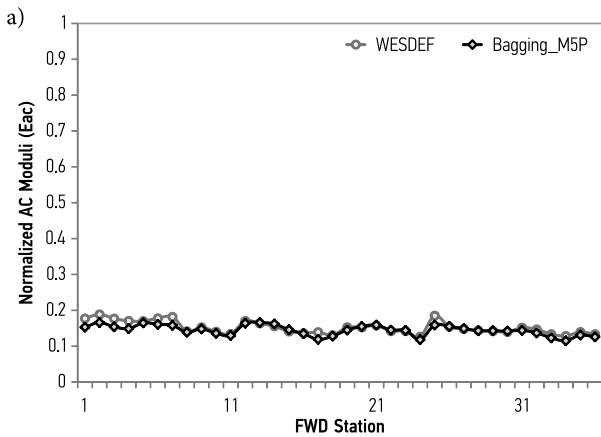


Fig. 9. Asphalt pavement moduli predictions with Bagging_M5P using actual FWD deflections basins acquired in the field: a - E_{ac} ; b - E_{ri}

6. Experiments and Discussion of Results: Field Data

Bagging_M5P models constructed on the theoretic data were applied on the actual FWD data acquired from an airport flexible pavement test section at the U.S. National Airport Pavement Test Facility (NAPTF). The selected test section is a typical conventional granular base flexible pavement resting over a medium-strength subgrade. It consists of 127-mm (5-in) thick AC surface course, 200-mm (8-in) thick crushed stone granular base, 307-mm (12-in) thick granular subbase on top of the subgrade. For this analysis, the granular base and subbase layer thicknesses were combined.

A clayey material known as Dupont Clay (DPC) was used for the subgrade (target California Bearing Ratio of 8). The naturally-occurring sandy-soil material at the full-scale test site underlies the subgrade layer. Detailed information related to NAPTF flexible test sections, material properties, analysis of NDT data can be found in (Gopalakrishnan 2004). The FWD data referenced in this paper is accessible for download at the Federal Aviation Administration (FAA) Airport Technology Website: <http://www.airporttech.tc.faa.gov>.

Nondestructive tests using the FWD equipment were conducted on the selected test section prior to traffic testing to verify the uniformity of pavement and subgrade construction and strength. Surface deflection basins from FWD tests conducted on June 14, 1999 (pavement temperature = 21.2 °C) at nominal force amplitudes of 40-kN (9-kip) were used in this study.

For the sake of comparison, WESDEF (Cauwelaert et al. 1989), a traditional pavement inverse analysis program, was also used for backcalculating the pavement layer moduli from field FWD data. The WESDEF backcalculation program uses the WESLEA multi-layer elastic analysis program. It utilizes an iterative procedure to obtain a set of moduli that, when used in linear-elastic calculations, will produce deflections similar to the measured values. The program has the ability to backcalculate moduli values using deflections with depth, such as those obtained using Multi-Depth Deflectometers (MDDs), as well as with surface deflections. The material type, entered for each layer in the pavement structure, is used to establish the default seed modulus, minimum and maximum moduli, the Poisson's ratio, and the interface slip values.

In WESDEF, the modulus for the stiff layer was set to 6.9 GPa (1000000 psi) with a Poisson's ratio of 0.50. The pavement layer moduli predicted by Bagging_M5P predictor based on field data are plotted together with those predicted by WESDEF in Fig. 9. In general, the Bagging_M5P moduli predictions are consistent and agreeable with those predicted by WESDEF. Note that WESDEF assumes the subgrade to be linear elastic and requires seed moduli values to start the optimization process while Bagging_M5P considers the non-linear stress-dependent subgrade properties and employs knowledge discovery and data mining principles to find the solutions.

Irrespective of the high prediction accuracy of any developed backcalculation model, there are some major factors that can lead to erroneous results in pavement backcalculation (Irwin 2002; Von Quintus, Killingsworth 1998). For instance, major cracks in the pavement, or testing near a pavement edge can cause the deflection data to depart drastically from the assumed conditions. Pavements with cracks or various discontinuities and other such features are ill-suited for any backcalculation analysis or moduli determination. Also, layer thicknesses are not uniform in the field, nor are materials in the layers completely homogeneous. The spatial and seasonal variations of pavement layer properties in the field should also be considered.

Summary and Conclusions

The Falling Weight Deflectometer (FWD) is one of the most widely used test methods for assessing the structural integrity of existing pavements in a non-destructive manner. Used in combination with sampling and laboratory testing techniques, the FWD provides an effective and efficient means for evaluation of existing pavement structures, and for development of input parameters for design procedures. Typically, the FWD deflection measurements are used to backcalculate the in-situ elastic moduli of each pavement layer. Backcalculation is the inverse process of characterizing the stiffness properties of the paving layers through the deflection data collected by the FWD. The backcalculated moduli themselves provide an indication of layer condition. They are also used in an elastic layered or finite element program to predict the critical pavement responses (stresses, strains and deflections) under applied loads.

This paper introduced some of the existing data mining techniques to pavement community which were also successfully used to conduct real-time asphalt pavement inverse analysis (i.e., backcalculation). Non-parametric modeling techniques in data mining such as decision trees are known to be useful in cases when it is not easily possible to formulate any credible and useful assumption about the data distributions. Among the examined data mining techniques, Bagging_M5P predictors (Bagging meta learning technique with M5 model trees as the base learner) produced the best results using both theoretic pavement deflection basins as well as actual FWD deflection basins acquired in the field, which were consistent and in agreement with the WESDEF predictions.

References

- AASHTO. 1993. *Guide for Design of Pavement Structures*. Washington, D.C.: American Association of State Highway and Transportation Officials (AASHTO).
- Aha, D. W.; Kibler, D.; Albert, M. K. 1991. Instance-based learning algorithms, *Machine Learning* 6(1): 37–66. <http://dx.doi.org/10.1023/A:1022689900470>
- Alavi, S; LeCates, J. F.; Tavares, M. P. 2008. *Falling Weight Deflectometer Usage: a Synthesis of Highway Practice*. NCHRP Synthesis 381. Washington, DC: Transportation Research

- Board. Available from Internet: http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_syn_381.pdf
- Barenberg, E. J.; Petros, K. A. 1991. *Evaluation of Concrete Pavements Using NDT Results: Final Summary Report*. Project IHR-512. University of Illinois at Urbana-Champaign and Illinois Department of Transportation. Available from Internet: <http://ict.illinois.edu/publications/report%20files/TES-065.pdf>
- Bishop, C. M. 1996. *Neural Networks for Pattern Recognition*. Oxford University Press. 504 p.
- Breiman, L. 1996. Bagging predictors, *Machine Learning* 24(2): 123–140. <http://dx.doi.org/10.1007/BF00058655>
- Breiman, L.; Friedman, J.; Olshen, R. A.; Stone, C. J. 1984. *Classification and Regression Trees*. Chapman and Hall/CRC. 368 p.
- Cauwelaert, F. J. V.; Alexander, D. R.; White, T. D.; Barker, W. R. 1989. Multilayer elastic program for backcalculating layer moduli in pavement evaluation, *Nondestructive Testing of Pavements and Backcalculation of Moduli, ASTM Special Technical Publication* 1026: 171–188. <http://dx.doi.org/10.1520/STP19806S>
- Ceylan, H. 2002. *Analysis and Design of Concrete Pavement Systems Using Artificial Neural Networks*. University of Illinois at Urbana-Champaign. 512 p.
- Ceylan, H.; Gopalakrishnan, K.; Guclu, A. 2007. Advanced approaches to characterizing nonlinear pavement system responses, *Transportation Research Record* 2005: 86–94. <http://dx.doi.org/10.3141/2005-10>
- Cohen, W. W. 1995. Fast effective rule induction, in *Machine Learning: Proceedings of the Twelfth International Conference on Machine Learning*. July 9–12, 1995, Tahoe City, California. 115–123.
- Fausett, L. V. 1993. *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. Pearson. 461 p.
- Foxworthy, P. T.; Darter, M. I. 1989. ILLI-SLAB and FWD deflection basins for characterization of rigid pavements, *Nondestructive Testing of Pavements and Backcalculation of Moduli, ASTM Special Technical Publication* 1026: 368–386. <http://dx.doi.org/10.1520/STP19818S>
- Friedman, J. H. 2002. Stochastic gradient boosting, *Computational Statistics and Data Analysis* 38(4): 367–378. [http://dx.doi.org/10.1016/S0167-9473\(01\)00065-2](http://dx.doi.org/10.1016/S0167-9473(01)00065-2)
- Garg, N.; Tutumluer, E.; Thompson, M. R. 1998. Structural modelling concepts for the design of airport pavements for heavy aircraft, in *Proceedings of BCRA 1998 Conference: Fifth International Conference on the Bearing Capacity of Roads and Airfields*. 6–July 1998, Trondheim, Norway. 115–124.
- Gopalakrishnan, K. 2004. *Performance Analysis of Airport Flexible Pavements Subjected to New Generation Aircraft*: PhD Thesis, University of Illinois at Urbana-Champaign. 629 p.
- Gopalakrishnan, K.; Ceylan, H.; Attoh-Okine, N. O. 2010. *Intelligent and Soft Computing in Infrastructure Systems Engineering: Recent Advances*. Springer. 336 p.
- Gucunski, N.; Krstic, V. 1996. Backcalculation of pavement profiles from spectral-analysis-of-surface-waves test by neural networks using individual receiver spacing approach, *Transportation Research Record* 1526: 6–13. <http://dx.doi.org/10.3141/1526-02>
- Hall, K. T.; Darter, M. I.; Hoerner, T. E.; Khazanovich, L. 1997. *LTPP Data Analysis. Phase I: Validation of Guidelines for K-Value Selection and Concrete Pavement Performance Prediction*. Federal Highway Administration (FHWA). 150 p.
- Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. H. 2009. The WEKA data mining software: an update, *ACM SIGKDD Explorations Newsletter* 11(1): 10–18. <http://dx.doi.org/10.1145/1656274.1656278>
- Hicks, R. G.; Monismith, C. L. 1971. Factors influencing the resilient response of granular materials, *Highway Research Record* 345: 15–31.
- Ho, T. K. 1998. The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(8): 832–844. <http://dx.doi.org/10.1109/34.709601>
- Hoffman, M. S.; Thompson, M. R. 1981. *Mechanistic Interpretation of Nondestructive Pavement Testing Deflections*. University of Illinois, Urbana, Illinois. Available from Internet: <http://ict.illinois.edu/publications/report%20files/TES-032.pdf>
- Ioannides, A. M. 1990. Dimensional analysis in NDT rigid pavement evaluation, *Journal of Transportation Engineering* 116(1): 23–36. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(1990\)116:1\(23\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(1990)116:1(23))
- Ioannides, A. M.; Barenberg, E. J.; Lary, J. A. 1989. Interpretation of falling weight deflectometer results using principals of dimensional analysis, in *Proceedings of the 4th International Conference on Concrete Pavement Design and Rehabilitation*, April 18–20, 1989, Purdue University. 231–247.
- Irwin, L. H. 2002. Backcalculation: an overview and perspective, in *Proceedings of the Pavement Evaluation Conference*, October 21–25, 2002, Roanoke, Virginia, USA. 22 p. [CD].
- Irwin, L. H. 1994. *Instructional Guide for Back-Calculation and the Use of MODCOMP*. CLRP Publication No. 94-10. Local Roads Program, Ithaca, Cornell University, NY.
- Irwin, L. H.; Szenbenyi, T. 1991. *User's Guide to MODCOMP3 Version 3.2*. CLRP Report Number 91-4. Local Roads Program, Ithaca, Cornell University, NY.
- Khazanovich, L.; Roesler, J. 1997. DIPLOBACK: neural-network-based backcalculation program for composite pavements, *Transportation Research Record* 1570: 143–150. <http://dx.doi.org/10.3141/1570-17>
- Kim, Y.; Kim, Y. R. 1998. Prediction of layer moduli from falling weight deflectometer and surface wave measurements using artificial neural network, *Transportation Research Record* 1639: 53–61. <http://dx.doi.org/10.3141/1639-06>
- Kohavi, R. 1995. The power of decision tables, in *Machine Learning: ECML-95: 8th European Conference on Machine Learning*, April 25–27, 1995, Heraclion, Crete, Greece. 174–189.
- Lytton, R. L. 1989. Backcalculation of pavement layer properties, *Nondestructive Testing of Pavements and Backcalculation of Moduli, ASTM Special Technical Publication* 1026: 7–38. <http://dx.doi.org/10.1520/STP19797S>
- Meier, R. W.; Rix, G. J. 1994. Backcalculation of flexible pavement moduli using artificial neural networks, *Transportation Research Record* 1448: 75–82.
- Miradi, M. 2009. *Knowledge Discovery and Pavement Performance: Intelligent Data Mining*. PhD Thesis. Delft University of Technology, The Netherlands. 324 p. Available from Internet: http://repository.tudelft.nl/assets/uuid:5b6e67a7-1a0a-4268-990d-ec2438f7c903/miradi_20090408.pdf
- NCHRP. 2004. *Mechanistic-Empirical Pavement Design Guide*. National Cooperative Highway Research Program (NCHRP). Washington, D.C.
- Newcomb, D. E. 1987. Comparison of field and laboratory estimated resilient moduli of pavement materials (with discussion), in *Proceedings of the Association of Asphalt Paving Technologists* 56: 91–110.

- Quinlan, J. R. 1992. Learning with continuous classes, in *AI'92: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, 16–18 November 1992, Hobart, Tasmania. 343–348.
- Raad, L.; Figueroa, J. L. 1980. Load response of transportation support systems, *Transportation Engineering Journal* 106(1): 111–128.
- Rada, G.; Witczak, M. W. 1981. Comprehensive evaluation of laboratory resilient moduli results for granular material, *Transportation Research Record* 810: 23–33.
- R: *A Language and Environment for Statistical Computing*. 2009. R Development Core Team. R Foundation for Statistical Computing. Vienna, Austria. 409 p.
- Sharma, S.; Das, A. 2008. Backcalculation of pavement layer moduli from falling weight deflectometer data using an artificial neural network, *Canadian Journal of Civil Engineering* 35(1): 57–66. <http://dx.doi.org/10.1139/L07-083>
- Smith, K. D.; Wade, M. J.; Peshkin, D. G.; Khazanovich, L.; Yu, H. T.; Dater, M. I. 1998. *Performance of Concrete Pavements. Volume II: Evaluation of Inservice Concrete Pavements*. FHWA Publication No FHWA-RD-95-110. 330 p.
- Thompson, M. R.; Elliott, R. P. 1985. ILLI PAVE based response algorithms for design of conventional flexible pavements, *Transportation Research Record* 1043: 50–57.
- Thompson, M. R.; Robnett, Q. L. 1979. Resilient properties of subgrade soils, *Transportation Engineering Journal* 105(1): 71–89.
- Ullidtz, P. 1987. *Pavement Analysis*. North Holland. 318 p.
- Use of Artificial Neural Networks in Geomechanical and Pavement Systems*. 1999. Transportation Research Circular. Number E-C012. 18 p. Available from Internet: <http://onlinepubs.trb.org/onlinepubs/circulars/ec012.pdf>
- Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. Springer. 334 p.
- Von Quintus, H. L.; Killingsworth, B. M. 1998. Comparison of laboratory and in situ determined elastic layer moduli, in *Proceedings of the 77th Annual Meeting of the Transportation Research Board*. Washington, D.C. [CD].
- Wang, Y. 2000. *A New Approach to Fitting Linear Models in High Dimensional Spaces*. This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Computer Science at The University of Waikato. 218 p.
- Wang, Y.; Witten, I. H. 1997. Induction of model trees for predicting continuous classes. Poster paper presented at the *Machine Learning: ECML'97: 9th European Conference on Machine Learning*, April 23–25, 1997, Prague, Czech Republic.
- Wang, Y.; Witten, I. H. 2002. Modeling for optimal probability prediction, in *ICML'02: Proceedings of the Nineteenth International Conference on Machine Learning*, July 8–12, 2002, Sydney, Australia. 650–657.
- Weher, E. 1977. Edwards, Allen, L.: An introduction to linear regression and correlation. (A series of books in psychology.) W. H. Freeman and Comp., San Francisco 1976. 213 S., Tafelanh., s 7.00, *Biometrical Journal* 19(1): 83–84. <http://dx.doi.org/10.1002/bimj.4710190121>
- Witten, I. H.; Frank, E.; Hall, M. A. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann. 664 p.