Regression Model Evaluation for Highway Bridge Component Deterioration Using National Bridge Inventory Data

by Pan Lu, Shiling Pei, and Denver Tolliver

Accurate prediction of bridge component condition over time is critical for determining a reliable maintenance, repair, and rehabilitation (MRR) strategy for highway bridges. Based on bridge inspection data, regression models are the most-widely adopted tools used by researchers and state agencies to predict future bridge condition (FHWA 2007). Various regression models can produce quite different results because of the differences in modeling assumptions. The evaluation of model quality can be challenging and sometimes subjective. In this study, an external validation procedure was developed to quantitatively compare the forecasting power of different regression models for highway bridge component deterioration. Several regression models for highway bridge component rating over time were compared using the proposed procedure and a traditional apparent model evaluation method based on the goodness-of-fit to data. The results obtained by applying the two methods are compared and discussed in this paper.

INTRODUCTION

Bridge deterioration is a serious problem across the United States. According to the United States Department of Transportation (USDOT 2014), more than 604,000 bridges are located on public roads in the United States, with approximately 50% of them built before 1966 (during initial interstate highway construction). Bridges in this age group will reach their 50-year milestone in the next three years. Although 50 years was intended originally as the design life of many bridges, their service life can be extended through diligent maintenance and rehabilitation (Tolliver and Lu 2011). The efficient use of public funds for fixing and maintaining bridges to keep them in adequate condition requires an effective bridge asset management framework. To address this issue, transportation management agencies worldwide have begun to adopt bridge management systems (BMS). A BMS is used to determine the optimum future bridge maintenance, repair, and rehabilitation (MRR) strategy at the lowest possible life-cycle cost based on the forecasted bridge conditions (Frangopol et al. 2000).

In the United States, highway bridge ratings typically consist of three major components: deck, superstructure, and substructure. The components deteriorate as a result of operating conditions and external environmental loads. Because of the importance of these components for normal operation and safety, prediction models for component deterioration are routinely developed to assess the conditions of bridges for a given future time span.

Among all predictive models, regression models are the most widely adopted types for engineering applications (FHWA 2007). Regression models forecast future bridges' performances from a set of explanatory variables via equations developed based on past data.

Specifically for bridge condition prediction, as stated by J. Lee et al. (2008), the major challenge faced by current bridge deterioration modeling techniques is the lack of reliable prediction modeling of future bridge condition ratings. For a given set of past bridge condition data, researchers can potentially apply all the predictive models that are selected based on in-sample goodness-of-fit statistics, and it is likely that predictions from all these models will be different. This presents a difficulty with regard to BMS because these different prediction results can lead to different management strategies. From an end-user perspective, no matter how complicated or simple the models are, prediction accuracy is the most important characteristics of the model. Thus the selection of predictive models should not be based on the modeler's experience or the model goodness-of-fit

with respect to in-sample data, but on the model's ability to accurately predict future behavior. This study will develop and apply an objective model evaluation procedure to reveal the true forecast power of the bridge component regression models so that engineers and decision makers can use this approach to gain confidence in bridge deterioration prediction. No existing literature has addressed this issue. The proposed procedure will be applied to the comparison of two regression models constructed using bridge component rating data in Illinois.

LITERATURE REVIEW

Regression model forecast verification is sometimes called validation, or evaluation. The purpose of this process is to help assess the specific strengths and deficiencies of regression models when they are used to forecast values of the dependent variable using forecasted values of the explanatory variables. This process has the potential to justify uses of the model for forecasting and support better decision making (Wilks 2006). Depending on the process used in validation, there are three types of evaluation procedures: apparent, internal, and external (Steyerberg and Harrell 2014). Apparent validation validates a model's goodness-of-fit on the entire dataset used to construct the model, which may not reveal the true predictive ability of the model because the exact same dataset is used for model development and validation test. The estimation of the prediction ability has been shown to be overly optimistic with this validation method (Witten and Frank 1999).

Internal validation evaluates sample data from the same underlying population as the sample data used to develop the model. Usually, it is difficult to obtain new sample data from the same population for validation. In the literature, the most commonly found method for internal validation is data-splitting (McCarthy 1976; Clementi et al. 2001; Shao 1993; Snee 1977; Lu and Tolliver 2012).

Simple data splitting is a subsampling procedure also known as resampling. It re-sizes the sample data into two sub-datasets and uses one random subset for validation and the other for model development. There are several data-splitting techniques found in literature (Snee 1977): 1) Cross validation is a repeated data-splitting technique, it repeats the simple data-splitting analysis many times and the predictions are averaged. 2) The bootstrapping procedure is different from splitting in that bootstrapping samples are taken with replacements from the original sample while data-splitting samples are selected without replacement. 3) The jack-knife technique is very similar to repeated data-splitting except it only takes one of the records from the original sample out at one time and repeated as many times as the total number of records in the original sample.

These procedures are powerful techniques when external validation is not possible. However, external validation is the most accurate and unbiased test for the model and the entire data collection process as stated by Harrell et al. (1996). External validation's main emphasis is that predictions from the previously developed model are tested on a new dataset that is different from the development population.

In the following sections, it will be shown that the data-based procedure proposed in this study can also be considered as a type of external validation with time delay and will reveal the model's true forecasting power in the past. The procedure is unique in that it focuses on long-term prediction power evaluation, which has not been investigated extensively in engineering applications. Several key prediction accuracy measures will be used in the examples and are introduced in this section.

DATABASE

The National Bridge Inventory (NBI) ASCII database is a unified database compiled by the Federal Highway Administration (FHWA) for all bridges and tunnels in the United States that have public roads passing above or below (FHWA 2007).The database provides information on 116 items and 432 characteristics of each bridge, including, but not limited to, bridge type and specification, bridge geometric information, bridge functional description, operational condition, bridge inspection data, and bridge construction and reconstruction records. The detailed information for each item and

characters can be found in the FHWA NBI reference report (FHWA 1995). The data in NBI is collected by state highway agencies and reported to FHWA annually.

Within NBI inspection data, there are three primary component ratings of special importance to bridge asset management: deck condition rating (DCR), superstructure condition rating (SPCR), and substructure condition rating (SBCR). The NBI rating system includes eight interim levels between excellent (9) and failure (0). For detailed information regarding to NBI inspection data, viewers are referred to NCHRP (2009).

METHODOLOGY

Prediction Accuracy Measures

Cook and Kairiukstis (1990) state that reduction of error (RE) "should assume a central role in the verification procedure" (p. 181). RE is an example of a forecast skill statistic (Wilks 2006). Wilks (2006) defined forecast skill as the relative accuracy of a set of forecasts with respect to some set of standard controls, which are usually the average values of the predictions. The equation used to calculate RE can be expressed in the following Equation (1).

$$(1) \quad RE = 1 - \frac{SSE_v}{SSE_{ref}}$$

Where SSE_v = sum of squares of validation errors between observed and predicted values over the validation period and SSE_{ref} = sum of squares of validation errors between observed values and mean of the predictions often known as control values or reference values over the validation period.

The difference between observed and predicted values is defined as validation error noted as $e_{(i)}$. It can be mathematically expressed as Equation (2).

$$(2) \quad e_{(i)} = y_i - \widehat{y_{(i)}}$$

Where y_i and $\widehat{y_{(i)}}$ are the observed and predicted values of the predictions for validation data point i.

The sum of the squares of errors for validation, SSE_v, can be expressed as Equation (3) and the sum of squares of errors for reference, SSE_{ref}, can be expressed as Equation (4).

$$(3) \quad SSE_v = \Sigma_{i=1}^{n_v} e_{(i)}^2$$

$$(4) \quad SSE_{ref} = \Sigma_{i=1}^{n_v} (y_i - \bar{y})^2$$

Where n_v is the total number of data points in the validation dataset and $\bar{y}$ is the mean of the predictions, which usually serves as a reference or control value. Theoretically, the value of RE can range from negative infinity to one, where one indicates perfect prediction for the validation data set. It will only occur when all the residuals for validation data are zero. On the other hand, if SSE_v is much greater than SSE_{ref}, RE can be negative and large. As a rule of thumb, a positive RE indicates that the regression model on average has some forecast skill. Contrastingly, if $RE \leq 0$, the model is deemed to have no skill to predict (Wilks 2006; Cooks and Kairiukstis 1990). The similarity in form of the equations for RE and regression R^2 expressed as Equation (5) suggests that RE can also be used as validation evidence for R^2. The closer the values of RE and R^2 are to each other, the more the model is accepted as a predictive tool.

$$(5) \quad R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}$$

In this research, adjusted R-squared is reported to take account of the phenomenon of the R-squared automatically and spuriously increasing when extra explanatory variables are added to the model. Adjusted R-squared can be written as Equation (6).

(6) $R^2 = 1 - \dfrac{SSE/df_e}{SST/df_t} = 1 - \dfrac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-p-1)}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2/(n-1)}$

Other commonly used apparent model validation criteria, including Akaike's Information Criteria (AIC), Bayesian Information Criteria (BIC), and Mean Absolute Error (MAE), are also selected in this research and compared with the RE external model evaluation method. The model with the smallest AIC, BIC, and MSE is deemed the "best" model based on apparent validation since it minimizes the difference from the given model to the "true" model. In other words, that smaller AIC, BIC, or MSE indicates better model/predictor. They can be mathematically expressed as Equation (7), (8), and (9).

(7) $\text{AIC} = n * \ln\left(\dfrac{\text{SSE}}{n}\right) + 2k$

(8) $\text{BIC} = n * \ln\left(\dfrac{\text{SSE}}{n}\right) + \dfrac{2(k+2)n\sigma^2}{SSE} - \dfrac{2n^2\sigma^4}{SSE^2}$

(9) $\text{MSE} = (\Sigma_{i=1}^{nv} e_i^2)/n$

Where *n* is in-sample size, SSE is the sum of the squares of errors, e_i is validation error, k is the number of independent variables, MSE is mean squared error, and σ^2 is error variance.

Forecasting Power Evaluation Procedure

For a predictive model, it is critical to have a quantitative measure of prediction accuracy. More important, given the many similar regression modeling choices, the user needs to be able to tell with confidence which model will yield the best prediction over a given period of time, which can be the analysis period. The deterioration of a bridge is a relatively slow process. With bridge rating indicators collected annually, it is difficult to detect any trend unless one uses data over a long time span. If all existing data are used to construct the model, there will be no independent data left for validation purposes. In this study, it is proposed to artificially "reserve" part of the data as the independent validation "pool" (external data) to use in evaluating long-term regression model quality. Splitting the existing dataset for this process can be application-specific. For example, if it is of interest to a transportation agency to predict bridge condition in five years, the validation datasets should cover at least five years of data to reflect the prediction span, which will at least provide the forecasting power within the analysis periods.

The procedure is rather straightforward and can be summarized in four steps: (1) A user select prediction analysis period that is typical for the implementation of the model of interest. (2) Data should be segregated into two parts, one containing the latest data at least within the analysis period identified in step 1, which will be kept unused during the development of the model, with the rest of the data used to develop regression parameters for the selected model of choice (assuming regression models are used). The data set aside will provide forecasting accuracy validation information for the time horizon that is at least the same as the analysis period. (3) The predictive model and its variations (may include other model types also) will be used to predict the data in the desired time horizon and compared to the true data, which is the set-aside data. Based on this comparison, a group of physically meaningful indicators can be derived (based on the application of the model prediction) for the quality and accuracy of the model. (4) Based on these indicators, the forecasting skill trend over time is analyzed and the forecasting model selection is judged based on the analysis results. In short, this approach essentially "rolls back" the modeler in time, assuming the prediction was done N years ago (N is the needed prediction time span) without knowing the current data. However, the forecasting model is selected by using the current data. The resulting best model is truly the best model N years ago. As a result, presumption of the procedure is that the model that performs well N years ago will continue to perform well N years in the future.

This presumption brings about the limitation of using this approach: The interactive mechanism between independent variables and dependent variables needs to be close to stationary over the prediction time span. Or in other words, the prediction accuracy of the model will not change dramatically over time. For example, if a phenomenon was affected by certain factors (independent variables) that were not present in the past (e.g., adoption of a new bridge design detail, new deicing chemical), this approach will not apply. However, under such conditions, almost all statistical regression approaches will be invalid as that piece of information does not exist in any data pool. In such cases, only the physically or mechanistically derived models can be used for prediction. The proposed approach should have very accurate results for any process that has a stationary underline-driven mechanism, i.e., the influence of independent variables on the dependent variables does not change dramatically with time. Even for cases where the internal driven mechanism does change over time, as long as the change is slow relative to the prediction time span, the approach could still be used in a piece-wise linear fashion where only the portion of the data closest to the prediction time span is used to construct the model. It is envisioned that this approach can be applied to many engineering problems to assist decision making.

BRIDGE RATING PREDICTION MODEL EVALUATIONS AND FINDINGS

Bolukbasi et al. (2004) recommended bridge component deterioration models with third-degree polynomial functions of bridge age. The research developed third-degree polynomial regression models with 2,601 Illinois bridge NBI inspection records from 1976 to 1998 and recommended that the data should be filtered by eliminating bridges for which reconstruction works are not recorded.

To illustrate the application of the proposed procedure, two candidate regression models for predicting steel bridge deck component ratings performance in Illinois were evaluated and compared by both in-sample goodness-of-fit statistics as apparent validation method and the data-based prediction power evaluation procedure as external validation method. One of the prediction models duplicates the work of Bolukbasi et al. (2004) with and without a filtering process. The filtering method is based on trends in rating values. If the data show an increase in the value of the condition rating, then unrecorded repair and/or reconstruction activity is assumed and the data observation is deleted. For detailed filtering method, readers are referred to Bolukbasi et al. (2014). Steel bridge deck deterioration models are constructed based on Illinois NBI data from 1994 to 1999. The same data set is used to construct the other multiple linear regression model through an explicit enumeration of all possible independent variables to the best knowledge of the authors. Both regression models will be based on the dataset with and without a filtering process, so for comparison, there will be four models: polynomial model (Bolukbasi et al.'s third degree polynomial regression model) with a data filtering process, polynomial model without a data filtering process, full model (multiple linear regression model) with a data filter process, full model without a data filtering process. Illinois NBI bridge population data from 2000 to 2014 is set aside for the purpose of external data validation.

This example looked at the accuracy of various models in a 15-year prediction horizon, with some interesting observations obtained and discussed in this paper.

Overview of Models and Comparison

Regression models for bridge component deterioration are used by many transportation agencies to assist in bridge inventory management. In research on this subject, a few key explanatory variables were found to influence the prediction accuracy. These include the age of the bridge, traffic load, jurisdiction of the bridge, and deicing practices, just to name a few. A multiple linear regression model is constructed through an explicit enumeration of all possible explanatory variables available to the authors, and the "best" fitted model is selected based on model selection and is referred to as the "full" model afterward. The "full" model has "age" as continuous independent variable, traffic (Average Daily Traffic [ADT]) and bridge design as categorical variables. Traffic, ADT,

can be forecasted with various travel demand models and adjusted by seasonal, directional count data. In the United States, prototypical vehicles are defined for analyzing bridge load designs. The alphanumeric codes "H" or "HS" denote a single-unit truck and a tractor pulling a semitrailer, respectively. The numeric suffix represents the gross weight in tons for H truck or weight on the first two axle sets of the HS truck. For example, H_10 denotes a truck with a gross weight of 10 tons. A duplication of the work of Bolukbasi et al. (2004) to construct a third degree polynomial model is also selected for the purpose of comparison and refers to "polynomial" model. Both two models are constructed based on the data set with and without filtering process. In this study, both apparent and external validation will be performed for both "full" and "polynomial" models, and the indicators described in earlier sections will be used to evaluate model quality.

First, two sets of regression models were constructed as shown in Table 1 and some selected apparent model validation indicators, adjusted R-squared, AIC, MSE, and BIC, are also shown in Table 1. One can tell that model constructed with data filtering has smaller AIC and BIC values and larger adjusted R-squared values by comparing column 2 and 3 for polynomial model and column 4 and 5 for full model. This finding conforms with Bolukbasi et al. (2014), so the data filtering process to remove unrecorded rehabilitation or reconstruction effects is needed before model construction.

Full models are more likely to be selected over polynomial models based on larger adjusted R-squared value, smaller MSE values, and smaller AIC and BIC values with respect to apparent validation criteria by comparing column 2 and 4 for adjusted data set and column 3 and 5 for unadjusted data set. Note that the full model contains polynomial model, and, based on apparent validation criteria, the full model is believed to be the best model and is preferred.

Note that even the full model only has 35% of variances explained by the independent variables, that is because many variances that could cause the bridge deterioration are not available and excluded in the model, such as climate information and maintenance funding level status. The apparent evaluation method is designed to identify significant explanatory factors and can select the best-fitted model for the sample data. However, it is still questionable if the model selected by the apparent validation method is suitable for short- and long-term forecasting. In other words, the model with the highest R-squared may not yield the best forecasting results.

Model Validation and Comparison Results

Once the regression model apparent validation analysis was done using all Illinois NBI 1994-1999 data and the NBI datasets from 2000-2014 were kept as the external validation data pool; model external validation can be performed and analyzed by using the RE indicator. The validation datasets can be considered external datasets for three reasons: 1) NBI reports data annually in separate datasets that will lead to differences in various aspects of the data such as data collection techniques, data collection personnel, and possibly the definition of variables. 2) the bridges included in validation datasets and model construction data sets vary because bridges may be closed or open to traffic for various reasons in different years. 3) NBI inspection records in each single year contain the whole bridge population information but not sample information.

To illustrate the external validation method result, both deck models constructed with a data filtering procedure to remove unrecorded rehabilitation or reconstruction effects are compared with prediction power indicator, RE. As shown in Table 2, RE values for both models are calculated for each validation data set.

Table 1: Parameter Estimates and Statics with Data of 1994-1999

Deck Model	Polynomial model with data filtering	Polynomial model without data filtering	Full model with data filtering	Full model without data filtering
N	20643	63439	20643	63439
Adjusted R-squared	0.3258	0.2686	0.3561	0.2967
AIC	15573	38767	14637	36284
MSE	2.13	1.84232	2.03079	1.7717
BIC	15575	38769	14639	36286
Intercept	8.2709	8.19543	7.85967	8.0987
Age	-0.08254	-0.095	-0.09026	-0.10561
Age²	0.00097214	0.00123	0.00128	0.00148
Age³	-0.0000053	-0.00000613	-0.00000723	-0.00000754
ADT<=1000	N/A	N/A	-	-
ADT<=5000	N/A	N/A	-0.36832	-0.27658
ADT<=10000	N/A	N/A	-0.43792	-0.44108
ADT>10000	N/A	N/A	-0.64814	-0.44701
H_10	N/A	N/A	0.68265	0.47337
H_15	N/A	N/A	0.68574	0.45804
H_20	N/A	N/A	0.62199	0.45024
HS_15	N/A	N/A	0.90066	0.83339
HS_20	N/A	N/A	0.84992	0.51066
HS_20P	N/A	N/A	0.74399	0.35579
HS_25	N/A	N/A	-	-

Note: "N/A" indicate that the corresponding variables are not considered in the model; "-"indicate that the corresponding variable is a reference variable for the categorical variable; all independent variables are significant at 99% of confidence.

Table 2: Reduction of Error Forecasting Power with Data of 2000-2014

Obs	Year	Observations	Full Model RE	Polynomial Model RE	RE Difference
1	2000	5491	0.26985	0.24409	0.026
2	2001	7219	0.25568	0.24016	0.016
3	2002	7122	0.21916	0.20442	0.015
4	2003	7034	0.20004	0.19137	0.008
5	2004	6943	0.21934	0.21097	0.008
6	2005	6869	0.20258	0.19101	0.012
7	2006	6763	0.19298	0.18161	0.011
8	2007	6707	0.17798	0.16430	0.014
9	2008	6664	0.17893	0.16412	0.015
10	2009	6631	0.19466	0.18045	0.014
11	2010	6630	0.17018	0.16121	0.009
12	2011	6680	0.18777	0.18349	0.004
13	2012	6660	0.18815	0.18490	0.003
14	2013	6664	0.18011	0.18374	-0.004
15	2014	6601	0.14751	0.15336	-0.006

One can tell from Table 2 that there are 15 external validation datasets and the data size ranges from 5,491 to 7,219. As stated earlier, an RE of 1 indicates perfect prediction for the validation dataset. As a rule of thumb, a positive RE indicates that the regression model, on average, has some forecast ability with higher values indicating better forecasting, and the closer the values of RE are to each other, the more the model is accepted as a predictive tool (Wilks 2006; Cooks and Kairiukstis 1990). Both models' RE values start at values close to, but less than, the models' adjusted R-squared value: 0.26985 vs 0.3561 (0.086 difference) and 0.24409 vs 0.3258 (0.082 difference). The finding shows that the evaluation adjusted R-squared factor provides close estimation of the true prediction accuracy for the example models at the beginning of the validation years, but the adjusted R-squared evaluation might be optimistic for forecasting power, especially for long-term analysis.

RE values all are positive, indicating a certain level of forecasting power, which is promising with the true external observations. The full model has a relatively higher forecasting power when compared to the polynomial model for the first 13 years, and then the polynomial model shows a higher forecasting power for the 14th and 15th years. In general, the forecasting power difference between the two models, column 6, shows that forecasting power of the two models decreases over the 15 validation periods. This finding indicates that the full model may be the preferred model to forecast near-term condition; but for long-term forecasting purpose, e.g., forecasting condition over 10 years, the polynomial model may be preferred.

Reasons for this can be that apparent validation selects the full model by best describing the interactive mechanism between independent variables and dependent variables for the same sample dataset used to build the model. However, when applying the model forecasting ability to a future dataset, the interactive mechanism between independent variables and dependent variables may be unchanged over the short prediction time span but may change for a longer prediction period. The full model involves more independent variables than the polynomial model (which only involves one variable, age), and the likelihood that certain factors included in the full model will change in long time span is higher than in the polynomial model. When that happens, the full model loses its forecasting power compared to the polynomial model or, in other words, the full model introduces more errors than the polynomial model.

Also note that by examining RE values for both models by year, column 4 and column 5 from Table 2, one can tell that forecasting power represented by RE can both increase and decrease for a short time, but in general and overall, the forecasting power decreases over the long term.

To further illustrate the forecasting ability of the two models, the authors repeat the proposed procedures for both superstructure and substructure models with a data filtering procedure considered. The apparent validation results are shown in Table 3, and external validation results are shown in Figure 1. From Table 3, the full model is preferred according to the apparent validation method for both substructure and superstructure models with higher adjusted R-squared, smaller AIC, smaller MSE, and smaller BIC values.

Table 3: Sub- and Super-Structure Model Statics with Data of 1994-1999

Deck Model	Substructure Polynomial model	Substructure full model	Superstructure Polynomial model	Superstructure full model
N	20643	20643	20643	20643
Adjusted R-squared	0.4287	0.4949	0.4225	0.4660
AIC	15853	13322	13668	12068
MSE	2.155	1.91	1.94	1.79
BIC	15855	13324	13670	12070

Shown from Figure 1, RE values for all models show up-and-down patterns in the short-run; however, they all show decreasing trends over the 15 validation periods. The full model is preferred at the beginning and the polynomial model shows better forecasting ability at validation years 13 and 14, respectively, for deck and superstructure models. For the substructure models, the full model is preferred over the 15 validation periods. Note that to determine the "best" model for its forecasting ability, especially for long-term planning, external data evaluation with roll-back data is recommended. External data evaluation will result in a different model preference compared with apparent model validation method.

Figure 1: Reduction of Error for Deck, Super-, and Sub- Structure Full and Polynomial Models
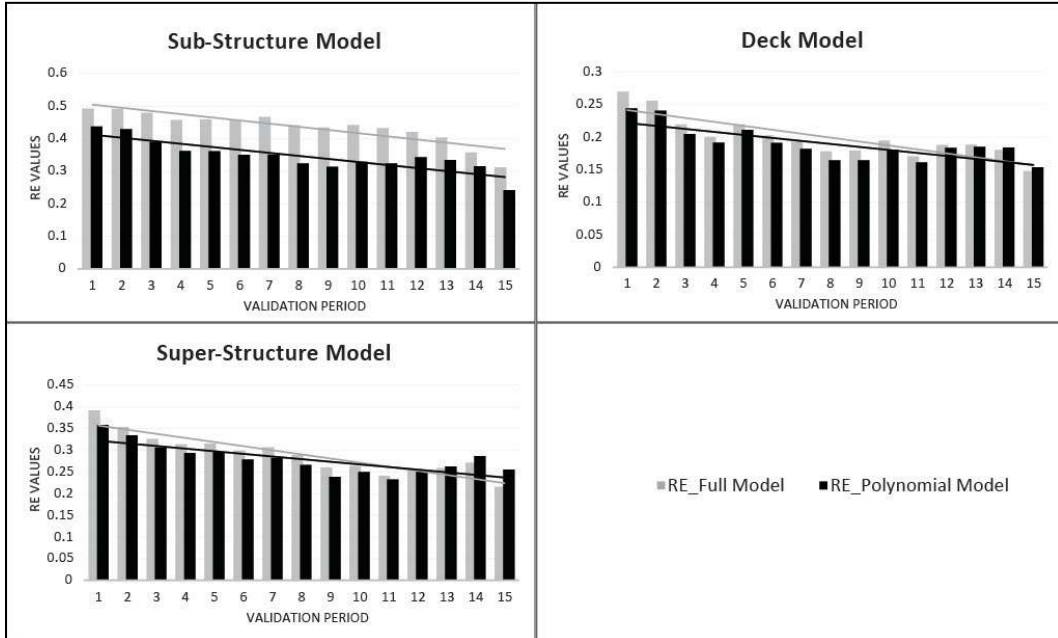
CONCLUSIONS AND RECOMMENDATIONS

The following are the main conclusion the study.

- The research by Bolukbasi et al. (2014) is verified and confirmed with new data from Illinois. The NBI data need to be filtered to remove the effect of unrecorded rehabilitation or reconstruction work to develop reasonable deterioration curves for bridges.
- The apparent evaluation method is valid for discovering the explanatory relationship between dependent and independent variables. However, it may not suitable for forecasting model selection.
- For the purpose of forecasting ability, external data evaluation with roll-back data is recommended. And the roll-back period should cover the forecasting period to reveal both short- and long-term forecasting power.

This paper proposes and demonstrates an objective, data-based approach for regression model forecasting ability evaluation. If the model is selected based on apparent evaluation only, then the forecasting outcome may not be accurate, especially for long-term planning, maintenance, rehabilitation, and replacement decisions. In this study, both a simple model as polynomial model and a full model selected by the apparent evaluation method have been generated for steel bridge component deterioration that can be used in any MRR decisions with confidence. In addition to producing deterioration curves, the methods described in the paper allow engineers to select the best forecasting model depending on their planning horizon. Finally, it is recommended to expand the use of the proposed procedure to help DOT decision making by developing a performance-based prediction model evaluation criteria.

References

Bolukbasi, Melik, Jamshid Mohammadi, and David Arditi. "Estimating the Future Condition of Highway Bridge Components Using National Bridge Inventory Data." *ASCE Practice Periodical on Structural Design and Construction* 9, (2004): 16-25.

Clementi, M., S. Clementi, M. Fornaciari, F. Orlandi and B. Romano. "The GOLPE Procedure for Predicting Olive Crop Production from Climatic Parameters." *Journal of Chemometrics* 15, (2001): 397-404.

Cook, E. and L. Kairiukstis. *Methods of Dendrochronology: Applications in the Environmental Science.* Kluwer Academic Publisher and International Institute for Applied Systems Analysis. Dordrecht, The Netherlands, 1990.

FHWA. *Information Public Disclosure of National Bridge Inventory (NBI) Data, 2007*. [available from: http://www.fhwa.dot.gov/bridge/nbi/20070517.cfm, date accessed: 07-12-2014,]

FHWA. *Recording and Coding Guide for the Structure Inventory and Appraisal of the Nation's Bridges*. Federal Highway Administration, U.S. Department of Transportation, 1995.

Frangopol, D.M., E.S. Gharaibeh, J.S. Kong, and M. Miyake. "Optimal Network Level Bridge Maintenance Planning Based on Minimum Expected Cost." *Proceedings of the Transportation Research Record*, Florida, (2000): 26–33.

Harrell, F.E., K.L. Lee, and D.B. Mark. "Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors." *Statistics in Medicine* 15, (1996): 361-387.

Lee, Jaeho, K. Sanmugarasa, M.Blumenstein, and Y.C. Loo. "Improving the Reliability of a Bridge Management System (BMS) Using an ANN-Based Backward Prediction Model (BPM)." *Automation in Construction* 17, (2008): 758-772

Lu, Pan, and Denver Tolliver. "Pavement Treatment Short-Term Effectiveness in IRI Change Using Long-Term Pavement Program Data." *ASCE Journal of Transportation Engineering* 138 (11), (2012): 1297-1302.

McCarthy P.J. "The Use of Balanced Half Sample Replication Cross-Validation Studies." *Journal of the American Statistical Association* 71, (1976): 596-604.

National Research Council (NCHRP). *NCHRP Synthesis 397: Bridge Management Systems for Transportation Agency Decision Making*. The National Academies Press, Washington, D.C., 2009.

Shao, J. "Linear Model Selection by Cross-Validation." *Journal of the American Statistics Association* 88, (1993): 486–494.

Snee, R.D. "Validation of Regression Models: Methods and Examples." *Technometrics* 19, (1977): 415–428.

Steyerberg, E.W. and Frank E. Harrell. "Validation of Predictive Regression Models 2014." 2014, date accessed: 07-12-2014, www.pitt.edu/~super7/7011-8001/7071.ppt

Tolliver, D., and P. Lu. "Analysis of Bridge Deterioration Rates: A Case Study of the Northern Plains Region." *Journal of the Transportation Research Forum* 50 (2), (2011): 87-100.

U.S. Department of Transportation (USDOT). *2013 Status of the Nation's Highways, Bridges, and Transit: Performance Report to Congress*, Washington, D.C., 2014.

Wilks, S. D. *Statistical Methods in the Atmospheric Sciences* (2nd ed.). Academic Press, Elsevier, London, 2006.

Witten, I.H., and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers, San Francisco, 1999.

Pan Lu *is assistant professor of transportation at the Upper Great Plains Transportation Institute at North Dakota State University. She received her Ph.D. in transportation and logistics from North Dakota State University in 2011. Her current research interests include statistical modeling in transportation, asset management, multimode transportation efficiency analysis, transportation system models, operation research in transportation, and transportation safety analysis.*

Shiling Pei *is assistant professor of civil and environmental engineering at Colorado School of Mines. He received his Ph.D.in civil engineering from Colorado State University in 2007. His current research interests include mechanistic models and non-linear structure dynamics, hazard mitigation and resilient infrastructure system, traditional and innovative timber systems, large scale dynamic testing and monitoring, performance-based engineering, structural reliability, and earthquake induced loss analysis.*

Denver Tolliver *is director of the Transportation & Logistics program and director of the Upper Great Plains Transportation Institute at North Dakota State University. He is also the director of the five-university Mountain-Plains Consortium, which is one of USDOT's designated university transportation centers. He holds doctoral and master's degrees from the Virginia Polytechnic Institute in environmental design & planning and urban & regional planning, respectively. Since 1989, Tolliver has been awarded $40 million in competitive research grants from federal and state agencies and private industries to conduct highway, railway, waterway, and multimodal analyses.*