Improved Imputation of Missing Pavement Performance Data Using Auxiliary Variables

J. Farhan[1] and T. F. Fwa, M.ASCE[2]

Abstract: Missing data in pavement condition and performance records of pavement management systems (PMS) are commonly encountered in practice. Imputation of missing data is often required in the analysis of pavement performance and decision making for maintenance and management of pavement networks. The traditional methods of handling missing data by pavement engineering professionals include deletion of affected records, and imputation of missing data by means of interpolation substitution, mean substitution, or regression substitution. Today, the advancement of computer technology has permitted the use of computationally complex stochastic methods of multiple imputation to improve the accuracy of imputed data. This study proposes an improved multiple imputation approach using a joint multivariate model on the understanding that, besides pavement performance data, a typical PMS database also collects data of related pavement properties and nonpavement variables such as traffic and weather conditions. The proposed approach develops an imputation strategy that uses selected pavement properties and nonpavement data as auxiliary variables in the multiple imputation analysis for missing pavement performance data. The theoretical basis and imputation procedure of the proposed approach are first presented, followed by a case study using the Long-Term Pavement Performance (LTPP) database to illustrate the choice of auxiliary variables and the steps involved in imputing missing rutting and roughness data respectively. The merits of the proposed approach are demonstrated by comparing the imputed results with actual measured data and by comparing the results with those obtained using the multiple imputation method without the inclusion of auxiliary variables. **DOI: 10.1061/(ASCE)TE.1943-5436.0000725.**

Author keywords: Pavement management system; Pavement performance; Missing data impuation; Multiple imputation; Aauxiliary variables; Pavement roughness; Pavement rut depth.

Introduction

Missing data are a common occurrence in the pavement performance database of a pavement management system (PMS) (Flintsch and McGhee 2009). Recollection of missing data is often costly and in the case of time-dependent pavement performance data, it is not always possible if missing data are not detected promptly. A common but undesirable practice is to discard records containing missing data. Some statistical packages discard records with missing values during analysis by default, thereby depriving analysts of potentially useful usable records (Enders 2010). To overcome this problem, highway agencies are increasingly employing data quality management programs to discover missing data, inconsistencies, and other anomalies in the data, and perform data cleansing activities (Larson and Forma 2007). In a study conducted by Flintsch and McGhee (2009), 61% of the highway agencies surveyed employed software routines to check for missing data.

The way that pavement condition/performance records with missing data are handled in a PMS decision-making process can have a significant effect on the conclusions drawn concerning pavement maintenance treatments and rehabilitation (Amado and Bernhardt 2002; Keleman et al. 2003). Instead of deleting or discarding records with missing data, another approach is to fill in or "impute" missing data. This approach allows the full dataset to be kept, which is a desirable feature in pavement management studies, as it retains a full coverage of all pavement segments over the complete time period of analysis.

In the analysis of missing data in pavement management database, the available missing-data imputation methods can be broadly classified into two main groups, namely the deterministic single imputation approach and the stochastic imputation approach (Enders 2010). The deterministic single imputation approach includes the traditional methods such as imputation by interpolation, mean, or regression. The stochastic imputation approach is represented by the multiple imputation technique, which is a computationally more complex method involving an iterative process. The latter approach is considered to be superior in providing statistically better estimates of missing data (Allison 2001; Enders 2010).

For the purpose of imputing missing data in a PMS database, this study proposes an improved procedure using the multiple imputation approach on the understanding that, besides pavement performance data, a typical PMS database also collects data of related pavement properties and traffic conditions. The proposed approach identifies relevant pavement properties and traffic conditions as auxiliary variables in the multiple imputation analysis of missing pavement performance data. The theoretical basis and imputation procedure of the proposed approach are presented in this paper. The application of the proposed approach is illustrated with a case study using the Long-Term Pavement Performance LTPP (2013) database. This case study also compares the quality of the imputed data by the proposed approach with those by other existing imputation methods.

[1]Research Fellow, Dept. of Civil and Environmental Engineering, National Univ. of Singapore, 10 Kent Ridge Crescent, Singapore 119260.

[2]Professor, Dept. of Civil and Environmental Engineering, National Univ. of Singapore, 10 Kent Ridge Crescent, Singapore 119260 (corresponding author). E-mail: ceefwatf@nus.edu.sg

Note. This manuscript was submitted on October 10, 2013; approved on June 2, 2014; published online on July 16, 2014. Discussion period open until December 16, 2014; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering***, © ASCE, ISSN 0733-947X/04014065(8)/$25.00.

Limitations of Traditional Single Imputation Methods

Traditionally, several methods of the deterministic single imputation approach have been employed for the purpose of estimating missing data. The main methods in this approach are the mean substitution method, the interpolation method, and the regression substitution method. A brief description of these methods and their limitations is provided in this section.

Mean Substitution Method

In this method, the missing values are substituted by a single constant value for a particular variable. Most commonly, this constant value is computed as the arithmetic mean of the available values of the variable. Though convenient to use, it distorts the actual distribution of collected pavement performance data. Statistically, it produces biased parameter estimates and an imputed dataset with a lower variance than the variance in the original dataset (Allison 2001; Enders 2010).

Interpolation Substitution Method

This method replaces missing values using a linear interpolation approach. The last valid value before the missing value and the first valid value after the missing value are used for the interpolation. If the first or last case in the series has a missing value, the missing value is not replaced. This method has been employed in practice for imputing pavement condition data (Yang et al. 2003; Bennett 2004). It assumes a linear trend in the vicinity of the missing data. It can be viewed as a form of conditional mean imputation that estimates a missing value only from the available data values immediately before and after it. Thus, it also produces biased parameter estimates and attenuates estimates of variability.

Regression Substitution Method

The method of regression predicts missing data based on the variable's relationship with other variables in the data set. In other words, the dependent variable with missing values is regressed on the independent variables to establish a regression equation used to impute missing values. Instead of using only two adjacent data points in estimating a missing value as in the case of interpolation substitution method, the regression substitution method makes use of the entire available data to predict each missing value. This method produces estimates of missing data that fit perfectly along the regression line without any residual variance, hence implying that a substantial correlation exists between variable with missing data and other attributes in the data set. Overall, it suffers from the same limitations as the mean and interpolation substitution methods, and, in particular, fails to account for the variability inherent in the original dataset (Little and Rubin 2002).

Stochastic Multiple Imputation Method

Concept and Advantages of Multiple Imputation Method

Imputation refers to the process of estimating a set of plausible values to substitute missing data. Single imputation methods, described in the preceding section, fail to reflect the true distributional relationship between observed and missing values, and they treat imputed data deterministically as though the data were actual measured pavement performance data. Hence, instead of substituting each missing value in a dataset with one randomly imputed value, it is reasonable to substitute each with several imputed values that reflect uncertainty about the missing data. This process of substituting missing data with multiple values is termed multiple imputation.

The key advantage of generating multiple imputations is that instead of using a point estimate as the imputed value, it simulates a distribution of missing data. Thus the overall variability in the population is maintained while preserving relationships between variables in the data set. Researchers have claimed that even a small number of simulated imputations could significantly improve the quality of missing-data estimation (Rubin 1987; Little and Rubin 2002).

Analysis Procedure of Multiple Imputation

The framework of analysis of the multiple imputation process adopted in this study is depicted in Fig. 1. A multivariate distribution is assumed for all imputation variables, and thus the multivariate normal distribution, known as the MVN (multivariate normal) method (Schafer 1997), is adopted in this study. Under the MVN method, the imputation model for the missing data given the observed data is a fully specified joint model (e.g., multivariate normal). A joint distribution of the data is specified with estimated parameters, and imputed values are drawn from this distribution. As shown in Fig. 1, a four-step procedure is followed to implement the MVN method in imputing missing pavement distress and performance data, as described below:

Step I: Since the MVN method assumes all data are normal, the observed data for all variables are transformed to approximately normal before imputation using a logit or square root transformation function and then transformed back to their original scale after imputation. Nevertheless, past research indicates that this type of imputation is quite robust even when the normality assumption is violated (Schafer 1997). The logit or logistic transformation is defined as

$$\log it(x) = \log\left(\frac{x}{1-x}\right) \tag{1}$$

Step II: This step initializes with estimates of parameters θ (mean vector and covariance matrix) from observed data Y_{obs}, and imputation of missing data Y_{mis} by using estimated parameters θ in a multivariate normal imputation model. Subsequently, newly computed values of missing data Y_{mis} and observed data Y_{obs}, i.e., complete data Y_{com}, are used to establish an improved estimate for the parameters θ. This entire process iterates between E (estimation of missing data) and M (computation of parameters) steps until convergence, which is defined as the maximum relative parameter change in the value of any parameter from one cycle to the next. The convergence criterion of the expectation maximization (EM) algorithm (to be explained in the next section) is set as 0.0001.

Step III: This step forms the basis of multiple imputation process, and generates a specified number M of data sets, each of which contains different estimates of the missing values. This step is performed using a data augmentation (DA) algorithm (to be explained in "Data Augmentation Algorithm"), which creates m multiple imputations using a technique called Markov chain Monte Carlo (MCMC). The DA algorithm initializes by estimating missing data Y_{mis} using final parameter estimates θ from Step II, and then draws new parameter estimates θ from a Bayesian posterior distribution based on the observed and imputed data. The process alternates between simulating missing data and parameters, thus creating a Markov chain $[\theta^{(t)}, Y_{\text{mis}}^{(t)}]$ that eventually converges.

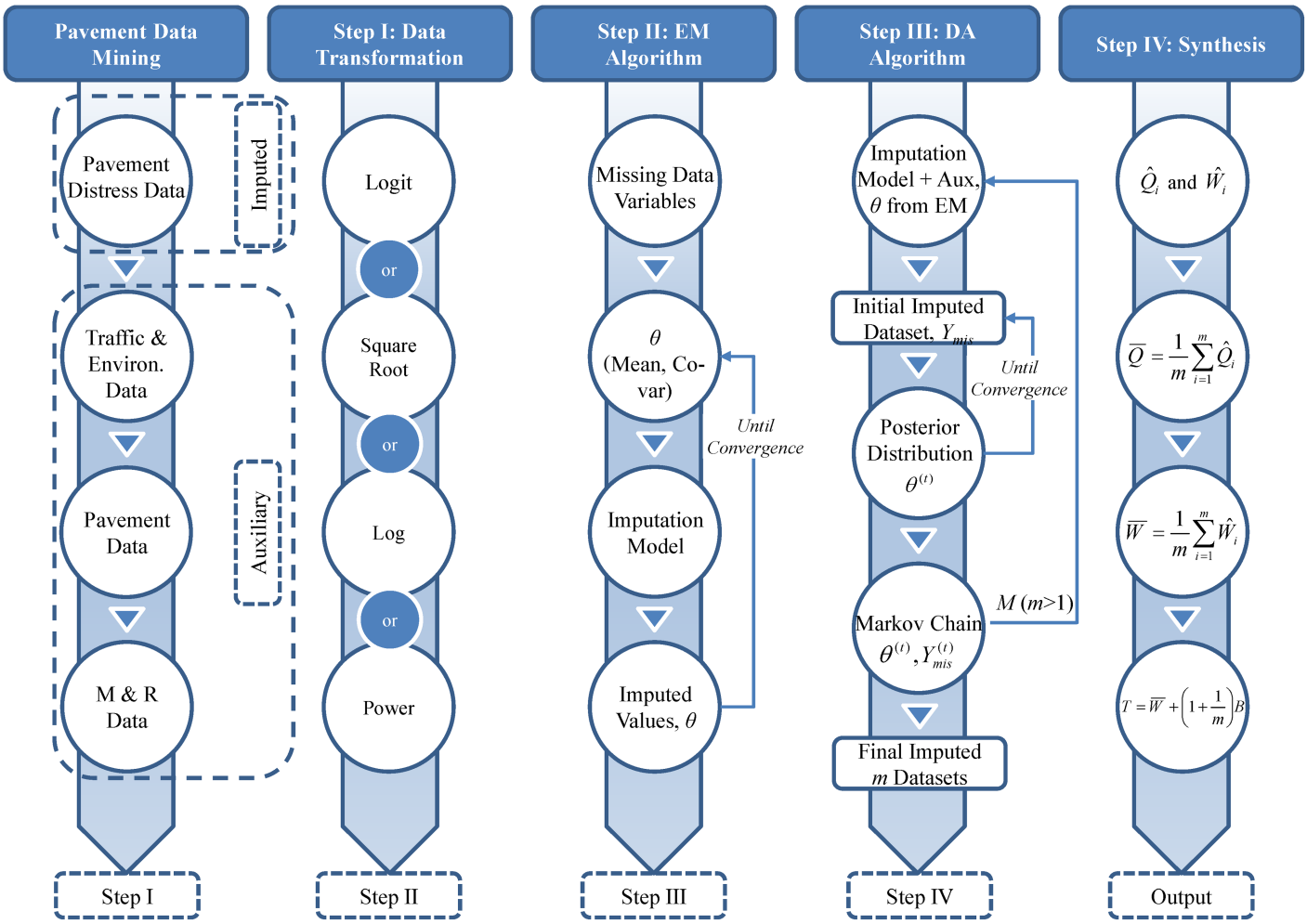

Fig. 1. Concept of the proposed approach

Table 1. Actual Datasets from LTPP Database Used in Imputation Analysis

State	ID	Proportion of actual missing distress data in LTTP database	
		Rut (%)	Roughness (%)
Arizona	04-1006	39.13	47.82
Connecticut	09-1803	34.78	21.73
Florida	12-1030	30.43	52.17
Maine	23-1026	52.17	43.47
Maine	23-1028	47.82	30.43
North Carolina	37-1006	60.87	56.52
North Carolina	37-1352	47.82	43.47
North Carolina	37-1814	52.17	39.13
Texas	48-9005	43.48	52.17

Table 2. Additional Proportion of Missing Data Created

State	ID	Proportion of created missing data	
		Rut (%)	Roughness (%)
Arizona	04-1006	21.43	16.67
Connecticut	09-1803	40	33.33
Florida	12-1030	25	36.36
Maine	23-1026	18.18	61.54
Maine	23-1028	33.33	50
North Carolina	37-1006	11.11	40
North Carolina	37-1352	16.67	23.08
North Carolina	37-1814	27.27	28.57
Texas	48-9005	23.08	18.18

The convergence is evaluated by visually inspecting the plots of the distribution (θ, Y_{mis}) from successive iterations to determine if a stationary distribution (Schafer 1997) is attained. Each iteration or cycle consists of an imputation or I-step followed by a posterior or P-step. Once DA has converged after k cycles, M runs each of length k are performed to generate M imputations.

Step IV: This final step combines M sets of point estimates and standard errors to obtain a single point estimate, and a standard error. Suppose $\hat{Q}_i$ and $\hat{W}_i$ are the point and variance estimates from the ith of m number of imputed datasets, then following the set of rules provided by Rubin (1987), the aggregated point estimate $\bar{Q}$ is the average of the m complete-data estimates given by

$$\bar{Q} = \frac{1}{m}\sum_{i=1}^{m}\hat{Q}_i \tag{2}$$

The between-imputation variance B and the total variance T associated with $\bar{Q}$ are given by

J. Transp. Eng. 2015.141.

Table 3. Total Proportions of Missing Data in Example of Imputation Analysis

State	ID	Proportion of missing data for analysis	
		Rut (%)	Roughness (%)
Arizona	04-1006	52.17	56.52
Connecticut	09-1803	60.86	47.82
Florida	12-1030	47.82	69.56
Maine	23-1026	60.86	78.26
Maine	23-1028	65.21	65.21
North Carolina	37-1006	65.21	73.91
North Carolina	37-1352	56.52	56.52
North Carolina	37-1814	65.21	56.52
Texas	48-9005	56.52	60.86

$$B = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{Q}_j - \bar{Q})^2 \tag{3}$$

$$T = \bar{W} + \left(1 + \frac{1}{m}\right)B \tag{4}$$

where $\bar{W}$ = within-imputation variance equal to the mean value of $\hat{W}_i$; and $i = 1, 2, \ldots, m$.

Expectation Maximization Algorithm

The expectation maximization algorithm (EM) is an iterative technique that finds maximum likelihood estimates in parametric models for incomplete data (Little and Rubin 2002; Roth 1994). This method involves two steps, namely expectation (E) and maximization (M). The first step computes the expected value for the missing data, and the second step is aimed at maximizing the likelihood function of the expected variables, obtained in the previous step, to find the parameter estimates. The EM strategy is based on the notion that the missing data have information that is useful in estimating parameters, such as mean vector and covariance matrix, and the estimated parameter has information that is useful in finding the most likely value of the missing data.

Consider an incomplete data matrix with observed data denoted as Y_{obs}, missing data as Y_{mis}, and a vector of parameters as θ. The complete dataset with imputed data is $Y_{com} = (Y_{obs}, Y_{mis})$. With the complete data log-likelihood function, $L(\theta) = f(Y_{com}|\theta)$, and the observed data log-likelihood function, $L(\theta) = f(Y_{obs}|\theta)$, the expected complete data log-likelihood function can be defined as

$$Q(\theta|\theta') = E\{\ln[f(Y_{com}|\theta)]Y_{obs}, \theta'\} \tag{5}$$

The EM algorithm starts at some value of θ and alternates between the following two steps (Ripley 1996):

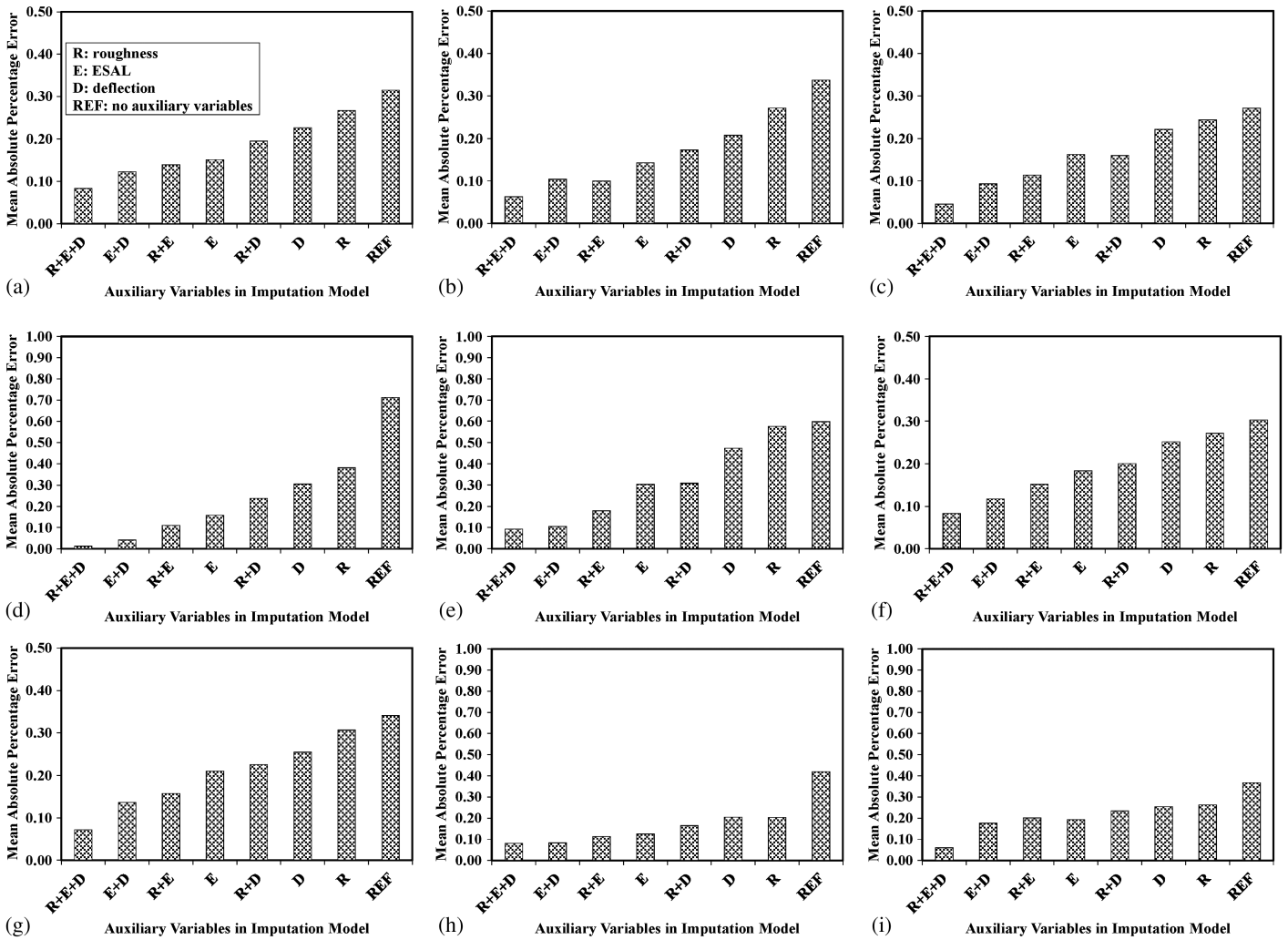

Fig. 2. Effects of auxiliary variables on MAPE for Case A rut-depth data

J. Transp. Eng. 2015.141.

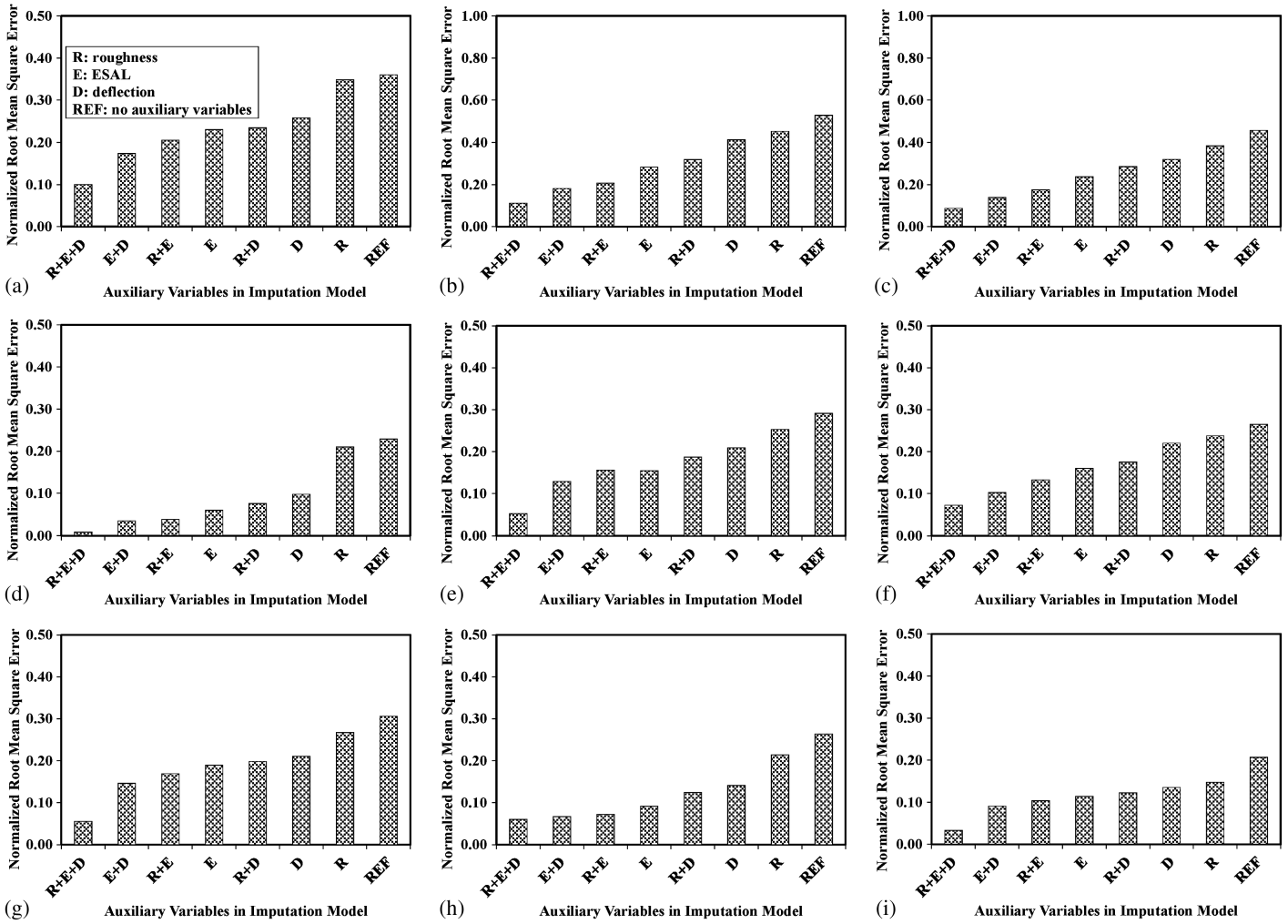

Fig. 3. Effects of auxiliary variables on NRMSE for Case A rut-depth data

1. Expectation step (E-step), i.e., computing $Q[\theta|\theta^{(t)}]$ as a function of θ;
2. Maximization step (M-step), i.e., finding $\theta^{(t+1)}$ that maximizes $Q[\theta|\theta^{(t)}]$

The log-likelihood function $L(\theta)$ increases with each iteration of the EM algorithm until converging to a local or global maximum (Dempster et al. 1977). The rate of convergence is directly related to the amount of unobserved or missing information in a data matrix, i.e., the greater the amount of missing data, the slower the convergence (Fraley 1999).

Data Augmentation Algorithm

Data augmentation (DA) is an iterative process that alternately fills in the missing data and makes inferences about the unknown parameters in a stochastic or random fashion (Tanner and Wong 1987). DA first performs a random imputation of missing data under the assumed values of the parameters, and then draws new parameters from a Bayesian posterior distribution based on the observed and imputed data (Schafer 1997). It uses starting values for the parameters θ obtained from the EM algorithm described in the preceding section. Starting at some value of θ, each cycle of the DA algorithm alternates between the following two steps:

1. Imputation step (I-step): Draws $Y_{\text{mis}}^{(t+1)} \sim P[Y_{\text{mis}}|Y_{\text{obs}}, \theta^{(t)}]$, and
2. Posterior step (P-step): Draws $\theta^{(t+1)} \sim P[\theta|Y_{\text{obs}}, Y_{\text{mis}}^{(t+1)}]$

The procedure of alternately simulating missing data and parameters creates a Markov chain $[\theta^{(1)}, Y_{\text{mis}}^{(1)}], [\theta^{(2)}, Y_{\text{mis}}^{(2)}], \ldots, [\theta^{(t)}, Y_{\text{mis}}^{(t)}]$ that eventually stabilizes or converges in distribution to $P(\theta|Y_{\text{obs}}, Y_{\text{mis}})$ (Schafer 1997). The imputed values in one data set should be independent of the imputed values from other data sets, and therefore a certain number of data augmentation cycles (I and P steps) are allowed to lapse between each saved data set.

Use of Auxiliary Variables in Multiple Imputation

Auxiliary variables are incorporated in the imputation process, based on their potential correlations with incomplete analysis variables, to improve the quality of imputations. This usually leads to a reduction in the bias of the estimates (Collins et al. 2001; Graham 2003). In pavement management systems, field surveys are conducted at regular time intervals to collect pavement distress and performance data, and typically, data of more than one performance indicator are collected. The common types of pavement distress and performance data recorded or derived from field survey measurements include the following:

1. Roughness data (e.g., international roughness index, IRI),
2. Rutting data (e.g., rut depths along wheel-paths),
3. Skid resistance data (e.g., skid number),
4. Pavement distress condition (e.g., pavement condition index), and

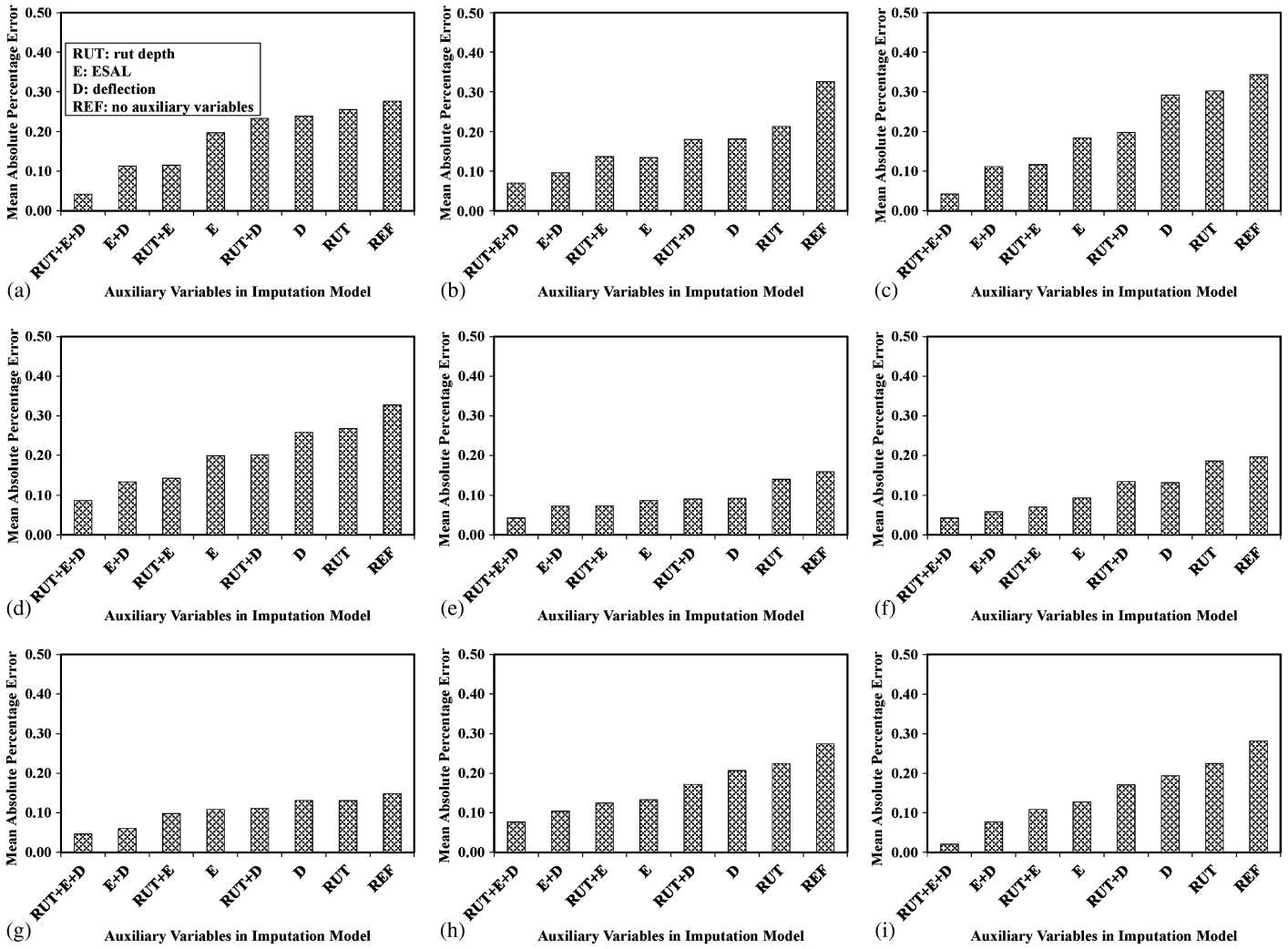

Fig. 4. Effects of auxiliary variables on MAPE for Case B roughness data

5. Pavement structural condition (e.g., maximum deflection by falling weight deflectometer test).

In addition to the aforementioned pavement performance data, most pavement field surveys also record the following associated nonpavement data:

1. Date and time of measurements,
2. Temperature,
3. Traffic volume and loading, and
4. Date and type of maintenance or rehabilitation.

In imputing missing values of the dataset of a particular pavement performance variable, other pavement performance variables as well as the nonpavement variables could be used as auxiliary variables.

Illustrative Example

Pavement Performance Data and Auxiliary Variables

The LTPP database (LTPP 2013) is an extensive data set that has been collected since 1987 and provides a reliable source of pavement performance and distress data. As a part of this study, nine pavement segments were selected randomly from the LTTP database as shown in Table 1. The imputation example illustrates missing data imputation for the following two cases:

1. Case A: Imputation of missing rut-depth data in rut database—the auxiliary variables used are roughness data in IRI, deflection measurements collected from a falling weight deflectometer (FWD), cumulative traffic loading data in equivalent single-axle loads (ESAL), and test date data in year; and
2. Case B: Imputation of missing roughness data in IRI database—the auxiliary variables used are rutting data measured in rut depths, deflection measurements collected from an FWD, cumulative traffic loading data in ESAL, and test date data in year.

Since the primary purpose of this analysis is to illustrate the use of auxiliary variables to improve quality of imputation, the wide range of possible options for auxiliary variables were not considered thoroughly.

To illustrate the benefits of the proposed procedure in imputing missing data, some data points in the original LTPP dataset were removed to become the additional missing data for the purpose of comparing between imputed and actual data. The example datasets are shown in Table 2. The total percentages of missing data to be imputed in the imputation analysis are delineated in Table 3.

Evaluation of Quality of Imputation using Auxiliary Variables

To evaluate the benefits of inclusion of auxiliary variables in imputing missing pavement performance data, a series of analyses

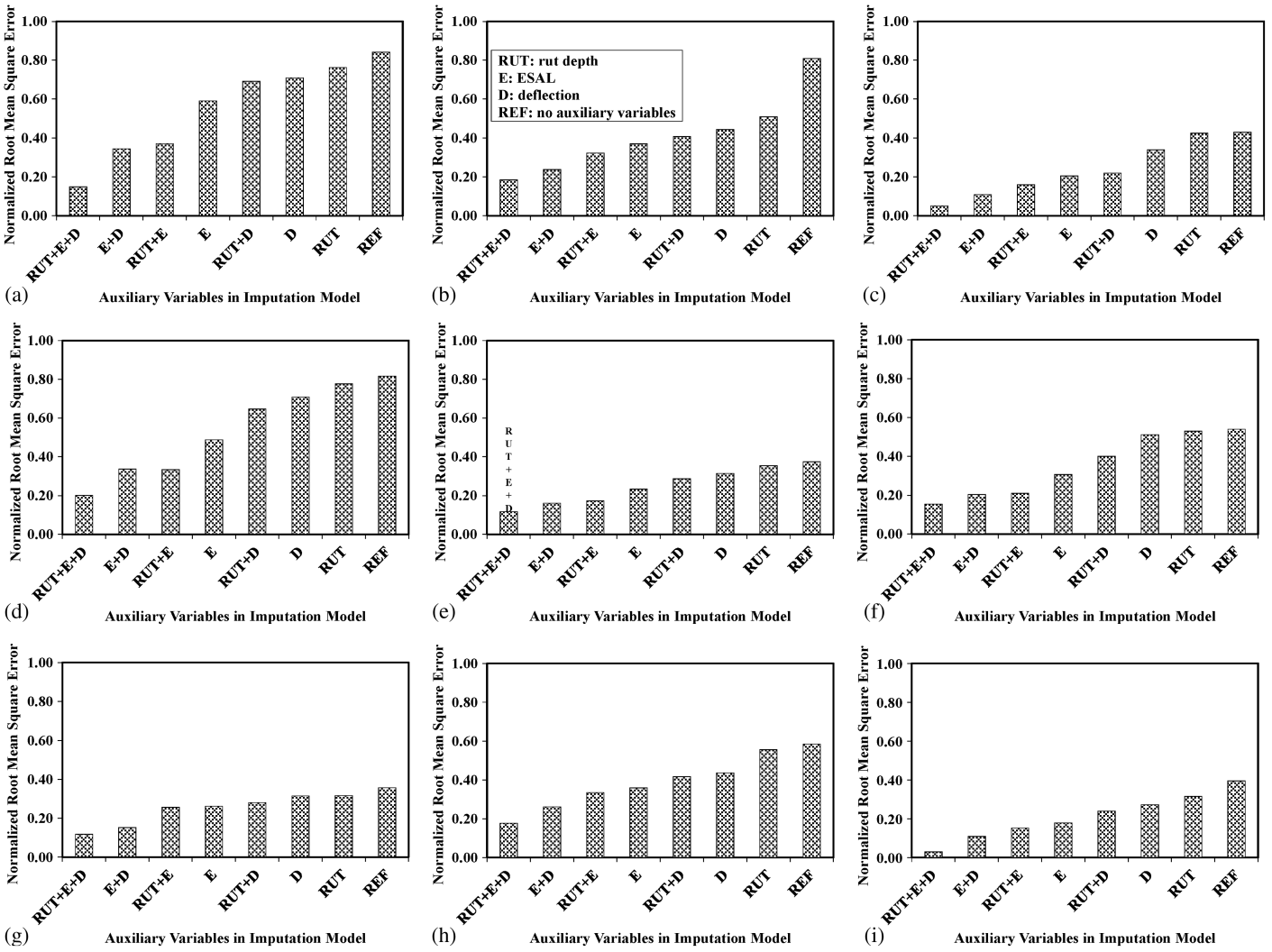

Fig. 5. Effects of auxiliary variables on NRMSE for Case B roughness data

were performed that added sets of auxiliary variables in a sequential manner for both Case A and Case B. The assessment of the imputation capability of the proposed approach is measured using the mean absolute percentage error (MAPE), and the normalized root mean squared error (NRMSE), respectively, as defined in Eqs. (6) and (7)

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - x_i'}{x_i'}\right| \tag{6}$$

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2}}{x_{\max} - x_{\min}} \tag{7}$$

where n = number of samples; x_i = actual value; x_i' = value estimated by the imputation model; $x_{\max}$ = maximum value in a dataset; and $x_{\min}$ = minimum value in a dataset. The actual values x_i are the values in the original LTPP dataset, which were initially removed (see Table 2) to be estimated in the example problem.

Analysis of Results

The results for Case A (i.e., imputation of missing rut-depth data) are plotted in Figs. 2 and 3, and those for Case B (i.e., imputation of missing roughness data) are plotted in Figs. 4 and 5. The improvements in the estimates by the inclusion of different auxiliary variables can be judged by the magnitudes of MAPE and NRMSE values. The smaller the values of MAPE and NRMSE, the smaller are the variations of the imputed values from the actual values. Based on the results, the performance of the various schemes of inclusion of auxiliary variables, in terms of their ability in improving the quality of estimates for missing data, are ranked in Table 4. The following observations may be made:

Table 4. Relative Performance of Various Schemes with Different Auxiliary Variables

Relative performance ranking	Auxiliary variables for Case A–Imputation of missing rut-depth data	Auxiliary variables for Case B–Imputation of missing roughness data
(Smallest error) 1	Roughness + ESAL + deflection	Rut + ESAL + deflection
2	ESAL + deflection	ESAL + deflection
3	Roughness + ESAL	Rut + ESAL
4	ESAL	ESAL
5	Roughness + deflection	Rut + deflection
6	Deflection	Deflection
7	Roughness	Rut
(Largest error) 8	No auxiliary variable	No auxiliary variable

1. Compared to the results for the case with no auxiliary variable, all other imputation models with the consideration of auxiliary variables produced improved estimates of missing data of either roughness or rut depth. This finding confirms the benefit of including auxiliary variables in the imputation analysis of missing pavement performance data;
2. In general, the quality of imputed values improved with the number of auxiliary variables considered. For both Case A and Case B, the imputation model including all the auxiliary variables available resulted in highest predictive performance; and
3. The influence of individual auxiliary variables on the quality of imputed values varies. For the present example problem, traffic loading ESAL yields the best beneficial results, followed by deflection, with roughness (for Case A imputing missing rut-depth data) and rut depth (for Case B imputing missing roughness data) the least effective. This observation underscores the benefit of identifying relevant auxiliary variables in imputation analysis.

Conclusions

This study has proposed a multiple imputation approach using auxiliary variables to address missing pavement condition and performance data in PMS. The proposed approach emphasized the use of auxiliary variables in the imputation model to improve its predictive performance. The benefits of using auxiliary variables in an imputation model were demonstrated through the use of an illustrative example. The results from the analysis indicated that the improvements to the quality of imputed values for the missing data of either roughness or rut depth were achieved with the inclusion of other pavement performance data and nonpavement data as auxiliary variables. Given the usual availability of auxiliary data, such as pavement properties data and other nonpavement data such as traffic and weather factors, in a typical PMS database, the study suggests that the proposed approach is a useful tool for imputing missing data in pavement management.

References

Allison, P. D. (2001). *Missing data: Quantitative applications in the social sciences*, Sage, Thousand Oaks, CA.

Amado, V., and Bernhardt, K. (2002). "Expanding the use of pavement condition data through knowledge discovery in databases." *Proc., of the Int. Conf. on Applications of Advanced Technologies in Transportation Engineering*, ASCE, Reston, VA, 394–401.

Bennett, C. R. (2004). "Sectioning of road data for pavement." *Presented at 6th Int. Conf. on Managing Pavements*, Queensland Dept. of Main Roads, Brisbane, Australia.

Collins, L. M., Schafer, J. L., and Kam, C. (2001). "A comparison of inclusive and restrictive strategies in modern missing data procedures." *Psychol. Methods*, 6(4), 330–351.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *J. R. Stat. Soc. Ser. B*, 39(1), 1–38.

Enders, C. K. (2010). *Applied missing data analysis*, Guilford Press, New York.

Flintsch, G. W., and McGhee, K. K. (2009). "Quality management of pavement condition data collection." *National Cooperative Highway Research Program Synthesis Rep. No. 401*, Transportation Research Board, Washington, DC.

Fraley, C. (1999). "On computing the largest fraction of missing information for the EM algorithm and the worst linear function for data augmentation." *Comput. Stat. Data Anal.*, 31(1), 13–26.

Graham, J. W. (2003). "Adding missing-data relevant variables to FIML-based structural equation models." *Struct. Eq. Model. Multi. J.*, 10(1), 80–100.

Keleman, M., Henry, S., and Farrokhyar, A. (2003). *Pavement management manual*, Colorado Department of Transportation, Denver, CO.

Larson, C. D., and Forma, E. H. (2007). "Application of analytic hierarchy process to select project scope for video logging and pavement condition data collection." *Transportation Research Record 1990*, Transportation Research Board, Washington, DC, 40–47.

Little, R. J. A., and Rubin, D. B. (2002). *Statistical analysis with missing data*, 2nd Ed., Wiley, Hoboken, NJ.

Long-Term Pavement Performance (LTPP) Database. (2013). "LTPP DataPave online." ⟨http://www.lttp-products.com/DataPave⟩ (Accessed Jun., 2013).

Ripley, B. D. (1996). *Pattern recognition and neural networks*, Cambridge University Press, Cambridge.

Roth, P. L. (1994). "Missing data: A conceptual review for applied psychologists." *Personnel Psychol.*, 47(3), 537–560.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*, Wiley, New York.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*, Chapman & Hall, London.

Tanner, M. A., and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *J. Am. Stat. Assoc.*, 82(398), 528–540.

Yang, J., Lu, J. J., and Gunaratne, M. (2003). "Application of neural network models for forecasting of pavement crack index and pavement condition rating." *Transportation Research Record 1699*, Transportation Research Board, Washington, DC, 3–12.