Derivation of Transition Probability Matrices for Pavement Deterioration Modeling

José J. Ortiz-García[1]; Seósamh B. Costello[2]; and Martin S. Snaith[3]

Abstract: The derivation of transition matrices has traditionally been effected using one of two methods. The standard approach is to observe, from historical data, the way in which a road network deteriorates from one year to the next, and use this to estimate the transition matrix probabilities. Alternatively, a panel of experienced engineers can be used to estimate the probabilities using expert opinion. This paper proposes and describes the development of three further methods for the determination of transition probabilities. The first method assumes that the historical condition data for each of the sites in the network is readily available. The second utilizes the regression curve obtained from the original data, and the third assumes that the yearly distributions of condition are available to assit in the process. In each case, an objective function aims at minimizing the difference between each of the method functions obtained from the original data and the corresponding functions obtained from the transition probabilities. An analysis of the results concluded that although the transition matrix fitted curve for the third method was not always as close to the regression curve as in the other methods, it did yield a distribution not only closer, but comparable, to the original distributions in all tested cases. The third method has therefore been taken forward and encapsulated in an analytical tool to assist the engineer in the formulation of transition matrices.

DOI: 10.1061/(ASCE)0733-947X(2006)132:2(141)

CE Database subject headings: Markov chains; Optimization; Pavement management; Deterioration; Probabilistic methods.

Introduction

Methods for predicting pavement deterioration can be broadly categorized into deterministic- and probabilistic-based models depending on the method employed to simulate aging of the pavement. Deterministic models are those for which condition is predicted as a precise value on the basis of mathematical functions of observed or measured deterioration (Robinson et al. 1998). Probabilistic-based models, on the other hand, predict the condition as the probability of occurrence of a range of possible outcomes.

At the project level, probabilistic models can be employed to predict the probability of a length of road being in a certain condition at a particular time. At the network level, however, probabilistic models are employed to predict the condition of a road network, as distinct from individual road lengths, and display the results as proportions of the network in a range of condition states. These models are commonly referred to as Markov prediction models.

[1]Group Engineer, Atkins Highways and Transportation, The Axis, 10 Holliday St., Birmingham, B1 1TF, U.K.

[2]Senior Lecturer, Dept. of Civil and Environmental Engineering, Univ. of Auckland, Private Bag 92019, Auckland, New Zealand.

[3]Professor of Highway Engineering, Civil Engineering, School of Engineering, Univ. of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

Note. Discussion open until July 1, 2006. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on February 10, 2004; approved on June 10, 2005. This paper is part of the ***Journal of Transportation Engineering***, Vol. 132, No. 2, February 1, 2006. ©ASCE, ISSN 0733-947X/2006/2-141–161/$25.00.

This research focused on the prediction of deterioration using probabilistic techniques at the network level. In particular, it addresses one of the major reasons for its underuse by engineers in the past. Namely, the fact that transition probabilities (i.e., the probability that a road at a certain level of deterioration will "transit" to another level of deterioration) are difficult to determine from empirical data or observation.

In response to this problem, three new methods have been suggested for determining transition probabilities. This paper describes the development of each of these alternatives, selects the optimum, and provides recommendations for its use.

Markov Theory Applied to Pavement Deterioration

It is not possible to describe the methodology adopted in determining transition probabilities without first providing a review of Markov theory and how it may be applied to pavement deterioration prediction.

The Markov prediction model is a stochastic process and is governed by three restrictions.

1. The first restriction is that the process is discrete in time;
2. The second is that the process should have a countable or finite state space; and
3. Finally, the process should satisfy the "Markov property" (Isaacson and Madsen 1976).

The Markov property is said to be satisfied if the future state of the process depends on its present state, but not on its past states. The stochastic process, in its usage for the prediction of pavement deterioration, is said to satisfy the Markov property if the future condition of the network is dependent on the present condition of the network and not on its past condition.

It is possible to show that a Markov Chain process may be used in the determination of pavement deterioration as follows:

- Pavement deterioration is continuous in time; however, to render it discrete in time the condition of the road network is analyzed at specific points in time. These usually take the form of duty cycles of 1 year.
- The state space, that is the number of possible outcomes, is infinite. However, in reality, the state space is defined as a finite number of fixed bands of condition for the particular defect under consideration.
- In pavement deterioration, it is assumed that the Markov property holds (Kerali and Snaith 1992).

Finally, a discrete-time Markov chain is said to be stationary, or homogeneous, in time if the probability of going from one state to another is independent of the time at which the step is being made (Isaacson and Madsen 1976).

Road condition may be modeled by either stationary or nonstationary Markov chains. In the case of stationary chains, it is considered that the road network will always deteriorate following the transition probabilities of one single transition matrix. If the pattern of deterioration of a particular road network is likely to change at a certain point in time, t, the deterioration process may be modeled by a nonstationary chain. This implies the use of a different transition matrix before and after t. In this case, the vector of the condition at t will become the starting vector for the second chain, which will operate with a different transition matrix. This type of arrangement may be performed as many times as required.

The initial state of any process may be described by a starting vector, $\mathbf{a_0}=(\boldsymbol{\alpha_1},\boldsymbol{\alpha_2},\ldots,\boldsymbol{\alpha_n})$. Using the analogy of pavement deterioration, the starting vector indicates the current condition of the network defined as the proportions in each band of condition. The starting vector should satisfy the following conditions:

- The sum of all α_i should be equal to one; and
- All entries should be nonnegative.

To model pavement deterioration with time, it is necessary to establish a transition probability matrix (TPM), denoted by **P**. The general form of **P** is given by

$$P=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

This matrix contains all of the information necessary to model the movement of the process among the condition states. The transition probabilities, p_{ij}, indicate the probability of the portion of the network in Condition i moving to Condition j in one duty cycle. A duty cycle in pavement deterioration refers to 1 year of traffic and environmental degradation. Similar to the starting vector, every TPM should satisfy the following conditions:

- The sum of the entries in each row should be equal to one; and
- All entries should be nonnegative.

In matrix notation, the probability distribution of the states of the process at a specific time, say $t=1$, is therefore given by

$$\mathbf{a_1} = \mathbf{a_0}\mathbf{P^1} \tag{1}$$

Similarly, the probability distribution of the states of the process at any Time t may be calculated by

$$\mathbf{a}_t = \mathbf{a_0}\mathbf{P}^t \tag{2}$$

The deterioration can therefore be modeled using the equation above, where $\mathbf{a_t}$=distribution of condition at Time t; $\boldsymbol{a}_0$=distribution of condition at Time 0, that is the starting vector; and $\mathbf{P^t}$=TPM raised to the power of t, the elapsed time in years.

Two more conditions apply to the process when it is used to simulate pavement deterioration. First, $p_{ij}=0$ for $i>j$, signifying the belief that roads cannot improve in condition without first receiving treatment. Second, $p_{nn}=1$, signifying a holding state whereby roads that have reached their worst condition cannot deteriorate further. Consequently, in pavement deterioration, the general form of the transition matrix **P** is denoted by

$$P=\begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1n} \\ 0 & p_{22} & p_{23} & \cdots & p_{2n} \\ 0 & 0 & p_{33} & \cdots & p_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

A further restriction allowing the condition to deteriorate by no more than one state in one duty cycle is commonly used in pavement deterioration modeling. The TPM is then denoted by

$$P=\begin{bmatrix} p_{11} & p_{12} & 0 & \ldots & 0 \\ 0 & p_{22} & p_{23} & \ldots & 0 \\ 0 & 0 & p_{33} & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

The Markov prediction model described above has been used in a number of pavement management systems to predict network deterioration. Prominent among these are *NOS* (Kulkarni 1984), *HIPS* (Thompson et al. 1987), *NETCOM* (Kerali and Snaith 1992), *MicroPAVER* (Butt et al. 1994), and *STRAT-2* (Costello et al. 2006).

Approaches to the Derivation of Transition Probability Matrices

Defining p_{ij} in the TPM has traditionally been effected by using one of two methods. The standard approach is to observe, from historical data, the way in which a road network deteriorates over time and use this to estimate p_{ij} using Eq. (3) below. N_{ij}=number of road sections in the network that moved from Condition i to Condition j during one duty cycle; and N_i=total number of road sections that started the year in Condition i:

$$p_{ij}=\frac{N_{ij}}{N_i} \tag{3}$$

The proportions are likely to vary from year to year thereby requiring an average to be determined for each p_{ij} to ensure accuracy in the model. In the event that insufficient amounts of reliable historical data are available, then a panel of experienced engineers can be used to estimate p_{ij} using expert opinion.

Three further methods for the determination of transition probabilities have been proposed, based on work by Ortiz-Garcia (2000). It is this research that provides the focus for the paper.

Experimental Procedure

The objective of the research was to determine the most appropriate method, from the candidate methods, for estimating the transition probabilities, p_{ij}. Three candidate methods of

J. Transp. Eng. 2006.132:141-161.

Table 1. Bands of Condition for Data Sets

Band of condition	Boundaries	
	Lower limit	Upper limit
1	90	100
2	80	90
3	70	80
4	60	70
5	50	60
6	40	50
7	30	40
8	20	30
9	10	20
10	0	10

estimation were suggested and subsequently tested on six different sets of artificial data specifically devised for the purpose. The three estimation methods were optimized using the generalized reduced gradient *(GRG2)* nonlinear optimization code incorporated as the Solver algorithm in the computer program *MICROSOFT EXCEL* (Fylstra et al. 1998).

The six artificial data sets (DS1–DS6) are all different in nature, but do share some common properties as follows. The network condition for all six data sets is represented on a scale ranging from zero to 100, where 100 indicates a perfect condition and conversely zero indicates complete disintegration. Each data set simulates the yearly collection of condition data on 30 sites of a particular network over a 20-year period. The number of data sites chosen was based on the fact that, with a sample size of about 30, a very good estimate of the standard deviate is obtained without placing undue demand on the road agency. The condition data from the 30 sites were typically distributed following a normal distribution. This can be supported by studies (Darter and Hudson 1973; AASHTO 1985) that have concluded that the forecast number of equivalent single axle loads on a pavement, in addition to other design variables, should be considered as normally distributed random variables. Based on this knowledge, other studies, including Li et al. (1997), have also assumed a normal distribution of condition data for predictive purposes.

Each data set represents a different trend in condition and is based on the general model forms identified by Paterson (1987). DS1 represents a standard S-shaped deterioration curve typical of the trend associated with either cracking or raveling progression. DS2 represents a standard deterioration curve where the rate of progression starts slowly but increases with age, as might be seen with roughness. DS3–DS5 represent a standard deterioration curve where the rate of progression starts fast but decreases with age, as for rutting, with each of the data sets representing a different rate of deterioration. Finally, DS6 represents a completely random rate of progression. These are further explained in the next section.

The 30 values of condition for each year are grouped into 10 bands of condition each of width 10, as defined in Table 1, thereby providing condition distributions, $\mathbf{a}_0'$ to $\mathbf{a}_{20}'$ representing the original data. Each data set was also subjected to a regression analysis based on the least-squares method. The resulting regression equation, $y(t)$, was taken as the deterministic model. The variance for each year was also calculated from the original data.

The distribution of observed condition in Year 0 constituted the starting vector, $\mathbf{a}_0$, for the process. The distributions, $\mathbf{a}_t$, for years $t=1$ to 20 were calculated using Eq. (2) in a nonstationary fashion: One transition matrix was derived for Years 1 to 10 and another one for Years 11 to 20. The vector $\mathbf{a}_{10}$ obtained with the first transition matrix was used as the starting vector for the second one. This variation would allow the study of a stationary chain when observing the behavior between Years 0 and 10, and at the same time, the study of a nonstationary chain when observing the whole analysis period.

The transition probabilities p_{ij} of $\mathbf{P}$ in Eq. (2) were obtained by optimizing an objective function, which was particular to each candidate method as explained below. A restriction imposed on the optimization process meant that the transition matrix derived allowed only transitions from one band of condition to the next during any duty cycle of 1 year. This assumption is widely used in Markov prediction modeling of pavement deterioration and had the added benefit of reducing calculation time.

Once the $\mathbf{a}_t$ distributions were obtained, it was possible to calculate their average, $\bar{y}(t)$, and variance, σ^2. The average was calculated using the following equation:

$$\bar{y}(t) = \mathbf{a}_t \cdot \mathbf{c} \tag{4}$$

where $\mathbf{c}$=vector of class marks (i.e., the midpoints of the condition bands). The variance was calculated using the following equation:

$$\sigma_t^2 = \sum_i (c_i^2 \cdot \alpha_{ti}) - \bar{y}(t)^2 \tag{5}$$

where c_i and α_{ti}=elements of the vectors $\mathbf{c}$ and $\mathbf{a}$, respectively.

To summarize the process, each data set was described by three parameters: The values of the regression equation $y(t)$ for each t, the variance of the original data, and the condition distributions of the original data, $\mathbf{a}_t'$ At the same time, the average condition $\bar{y}(t)$ and variance were calculated from the distributions $\mathbf{a}_t$, obtained by using the transition probabilities. In order to determine the relative success of each derivation method in simulating the deterioration patterns of the original data, the average condition $\bar{y}(t)$ for each t, the variance, and the distributions $\mathbf{a}_t$ obtained by using the transition probabilities were compared against the $y(t)$ for each t, the variance, and the distributions $\mathbf{a}_t'$ obtained from the original data, respectively. The following section describes the generation of the six data sets used in the analysis, followed by a description of the methods used for estimating the transition probabilities.

Generation of Data Sets

As noted previously, six different data sets were generated in order to test the methods of estimating transition probabilities. The deterioration trends of the generated data sets, together with the scatter associated with them, were selected to mirror various deterioration patterns observed in real data. This sought to ensure that the methods, if proven suitable, could be employed to obtain transition probabilities to predict the deterioration of actual networks.

Data Set 1

Data Set 1 (DS1) was created from an S-shaped deterioration curve, starting in the close to perfect condition of "95" at Year 0, deteriorating slowly at the beginning, increasing its deterioration rate between Years 5 and 15, and finally stabilizing at around Condition 20 in the final years. For each year, 30 data points were created by following a normal distribution with the average equal to the ordinate of the curve and standard deviation given by

J. Transp. Eng. 2006.132:141-161.

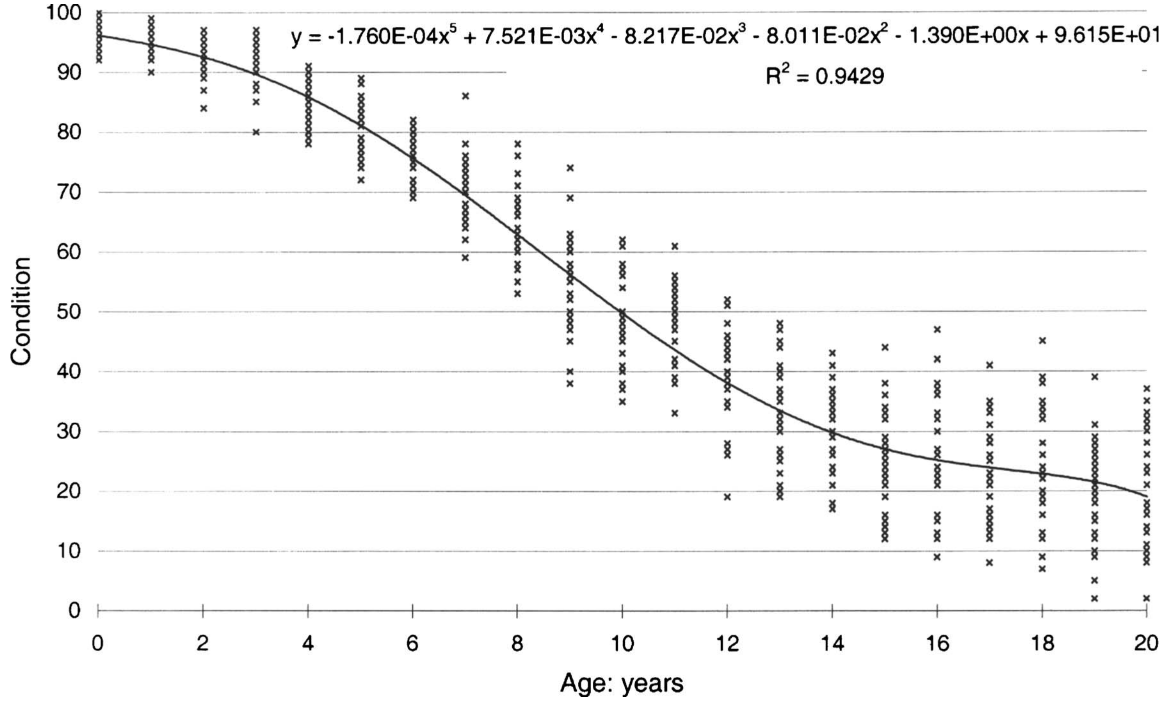

Fig. 1. Data Set 1. Scatter plot and regression curve.

$$\sigma_t = 2.5 + \frac{15}{40}t \tag{6}$$

As seen in Eq. (6), the standard deviation was set to increase linearly with time. This attempted to simulate the increase in condition variability as roads age. The boundary values of σ guaranteed that the created data fell within the condition scale boundaries. The pavement condition versus age plot for this data set is shown in Fig. 1.

Data Set 2

Data Set 2 (DS2) was created from a predefined transition matrix. The aim of creating a data set from a transition matrix was to check whether the methods would produce the same matrix from which the data were originally created. If this was the case, it would also produce the same deterioration curve and the same final distributions.

The data set was created as follows. A series of distributions of condition, a_t, were obtained from multiplying a starting vector $\mathbf{a_0=(1,0,\ldots,0)}$ by a matrix defined as follows: $p_{ii}=-10i+95$ except for $i=10$ when $p_{ii}=1$ representing the holding state; $p_{ij}=1-p_{ii}$ for $j=i+1$; all the other $p_{ij}=0$. In other words, 85% of the portion of the network in condition band one will remain in that band of condition and 15% will move to band two. Similarly, 75% of the portion of the network in band two will move to band three, etc. While these percentages may have been arbitrarily chosen for the purpose of this experiment, they were selected in this way to simulate a reasonable pattern of deterioration.

Once the distributions $\mathbf{a_t}$ were obtained, a number of data points were created in each condition band, according to the α_i of the particular band. The total number of points per year was, as for DS1, 30. The pavement condition versus age plot in Fig. 2 shows the scatter of this data set.

Data Sets 3, 4, and 5

Data Sets 3, 4, and 5 (DS3, DS4, and DS5, respectively) simulate three networks deteriorating at different rates. As for the previous data sets, these contain condition data for 30 sites of the same network. The average condition in Year 0 for each of them is between 80 and 100.

In DS3, the 30 sites deteriorate rapidly thereby simulating a high rate of deterioration. Sites in DS4 show a reduction in the rate of deterioration when compared to DS3 and were derived to simulate a medium rate of deterioration. Finally, DS5 represents sites with a slow rate of deterioration. The scatter of these data sets may be observed in Figs. 3–5.

Data Set 6

The final data set (DS6) was created in a way similar to Data Sets 3–5, by starting with all of the sites in a condition between 80 and 100 and then deteriorating them for each year. However, the pattern of deterioration introduced to this data set was completely random. Each site could exhibit any trend of deterioration as if none of them belonged to the same network. The aim was to generate a data set that simulated the unlikely phenomenon of a network in which roads deteriorate following completely different trends. The scatter of DS6 may be seen in the pavement condition versus age plot in Fig. 6.

Methods Used for Estimating the Transition Probabilities

Three candidate methods, A, B, and C, for estimating the transition probabilities were employed in the analysis. In Method A, it is assumed that the original data are available (i.e., historical condition data for each of the sites in the network) and may be used to estimate the transition probabilities. In Method B, the

J. Transp. Eng. 2006.132:141-161.

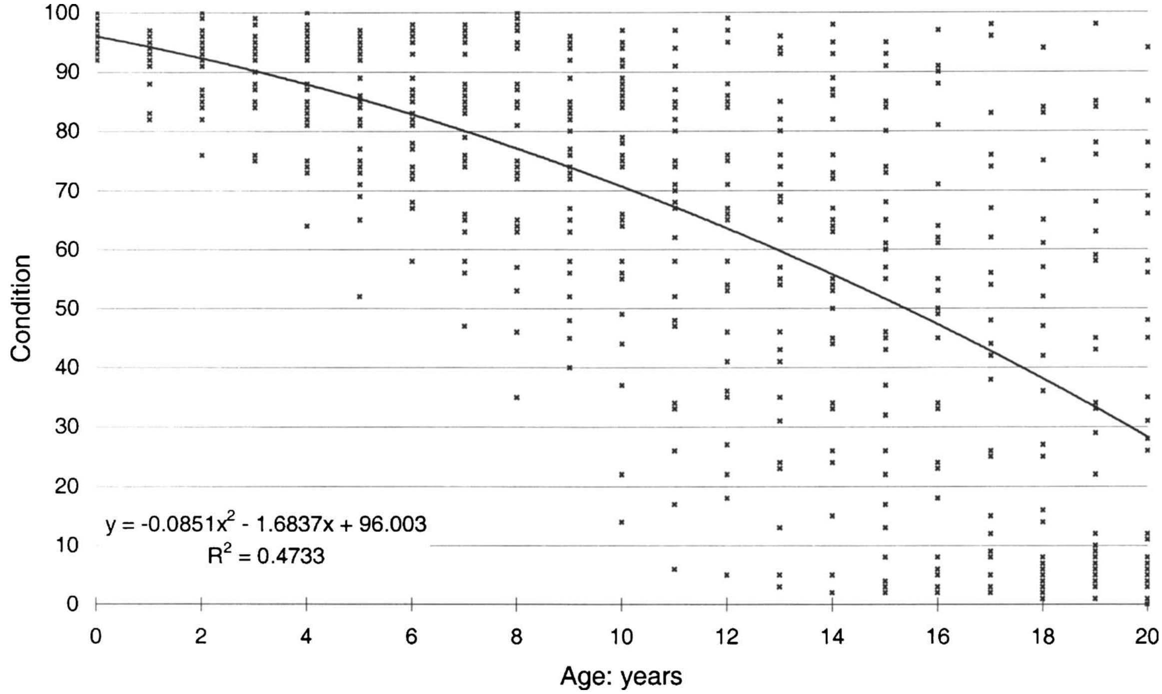

Fig. 2. Data Set 2. Scatter plot and regression curve.

regression equation obtained from the original data is used to estimate the transition probabilities. Finally, in Method C, the distributions of the original data (aggregated according to the bands of condition) are used. The mathematical derivation of each method is presented in the following sections.

Method A—Estimation of Transition Probabilities from Historic Data

Method A assumes that the raw data used in the regression analysis of the deterministic model are readily available. If the condition of site j at Time t is denoted by c_{jt}, the objective function Z for this method is given by

$$Z = \min \sum_{t} \sum_{j} [c_{jt} - \bar{y}(t)]^2 \tag{7}$$

The objective function aims, therefore, at minimizing the sum of the squared differences between each of the data points and the average condition calculated from the distributions of condition, $\mathbf{a_t}$.

Method B—Estimation of Transition Probabilities from Historic Data as Represented by a Regression Equation

Method B again uses the raw data, but after a regression equation has been obtained to describe the progression. The regression equations were obtained for the six data sets by least-squares fit. If $y(t)$ denotes the regression equation, the objective function, Z, employed to obtain the transition probabilities is as follows:

$$Z = \min \sum_{t} [y(t) - \bar{y}(t)]^2 \tag{8}$$

The objective function aims, therefore, at minimizing the difference between the average of $\mathbf{a_t}$ and the ordinates of the regression equation. This minimizes the distance between the regression curve and the transition matrix fitted curve.

Method C—Estimation of Transition Probabilities from Historic Data Grouped into Distributions

In Method C, the raw data are aggregated into bands of condition and presented in the form of distributions. Using the same nomenclature as above, if $a_t(i)$ denotes the ith element of the distributions obtained from Eq. (2) at Time t, and $a_t'(i)$ is the ith element of the original data distributions at Time t, the objective function Z takes the form:

$$Z = \min \sum_{t} \sum_{i} [\mathbf{a}_t(i) - \mathbf{a}_t'(i)]^2 \tag{9}$$

The objective function aims, therefore, at minimizing the difference between the distributions of condition obtained from the raw data and the distributions obtained from the transition probabilities.

It may be observed from the definition of the objective functions that the changing values in the optimization are the p_{ij} of the transition matrix. The nonlinear optimization algorithm starts the search for the optimum p_{ij} from initial p_{ij} values. It assesses the gradient of the objective function on the current region and changes the p_{ij} along the path of greatest gradient. The search continues until the objective function cannot be minimized further.

Experimental Results

Transition probabilities were derived for each of the data sets employing the three methods. The results were then compared by assessing the following three criteria:

J. Transp. Eng. 2006.132:141-161.

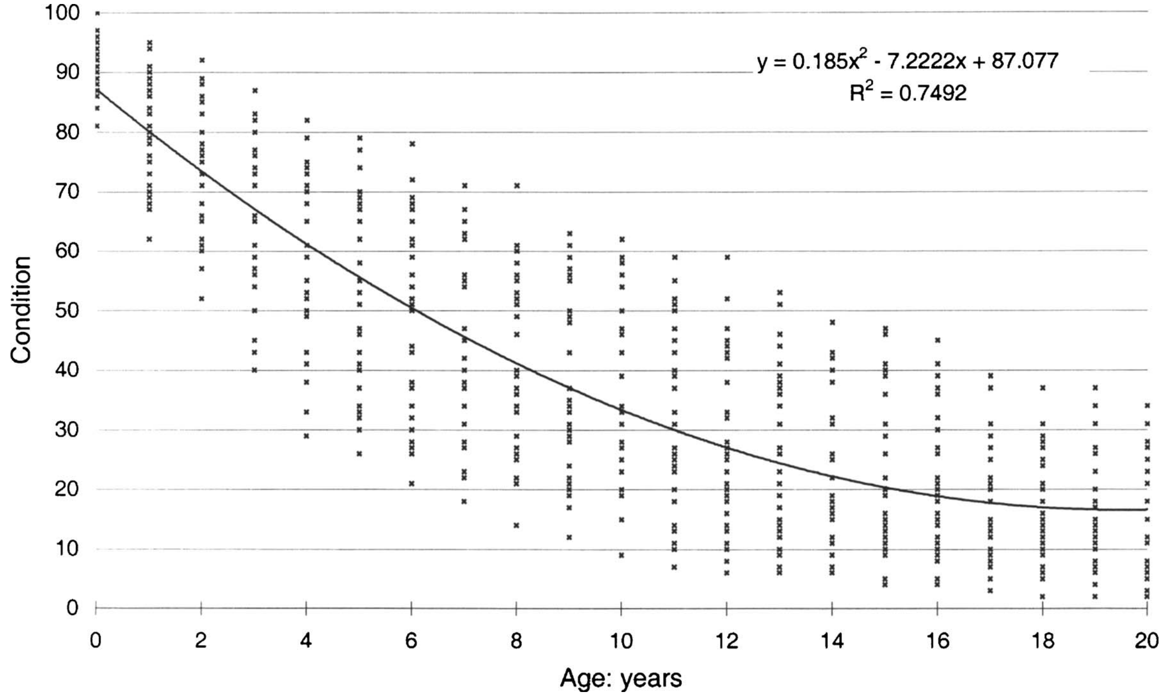

Fig. 3. Data Set 3. Scatter plot and regression curve.

1. The similarity between the transition matrix fitted curve $\bar{y}(t)$ and the regression curve, $y(t)$;
2. The similarity between the standard deviation of the original data and the standard deviation of the transition matrix fitted data; and
3. The similarity between the original condition distributions, $\mathbf{a}_\mathbf{t}'$, and the transition matrix fitted distributions, $\mathbf{a}_\mathbf{t}$.

A number of pavement condition versus age charts were plotted to assess the first two points. These charts may be found in Figs. 7–12. A snap shot of condition distributions for each data set in Year 15 was used to assess the third point. These are presented in Figs. 13–18. These results are discussed in the following sections.

Optimization Results for Data Set 1

Fig. 7 shows that the transition matrix fitted curve closely follows the regression curve for Methods A and B. It may also be noted

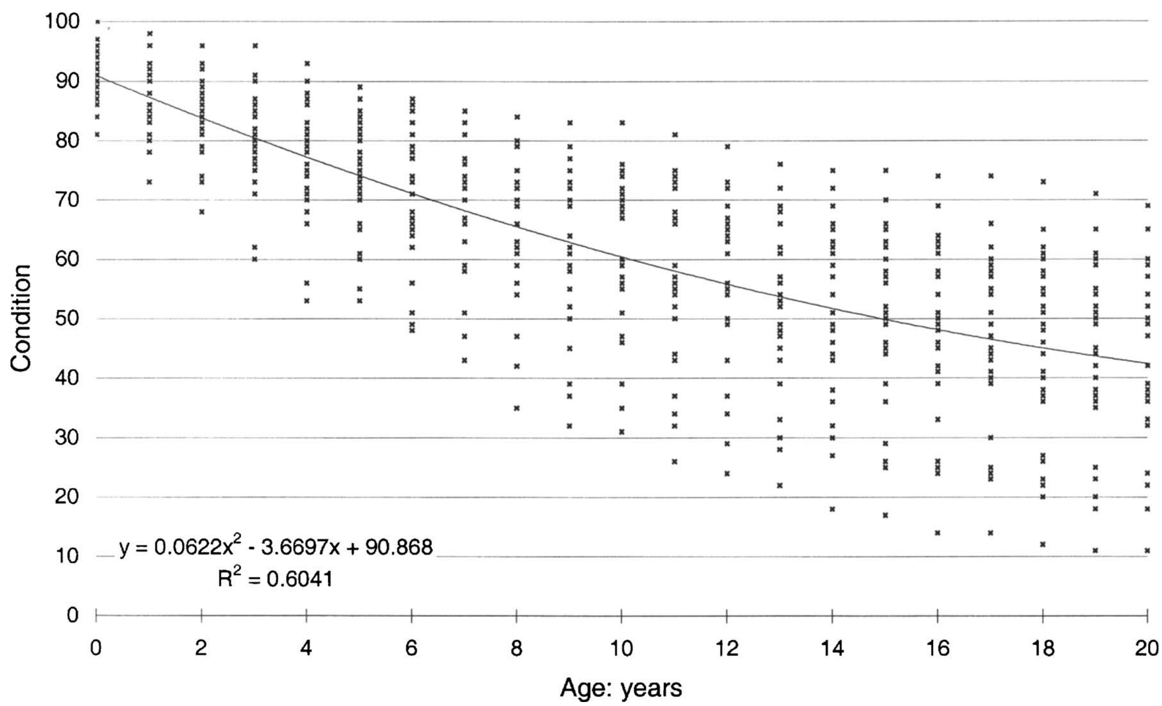

Fig. 4. Data Set 4. Scatter plot and regression curve.

J. Transp. Eng. 2006.132:141-161.

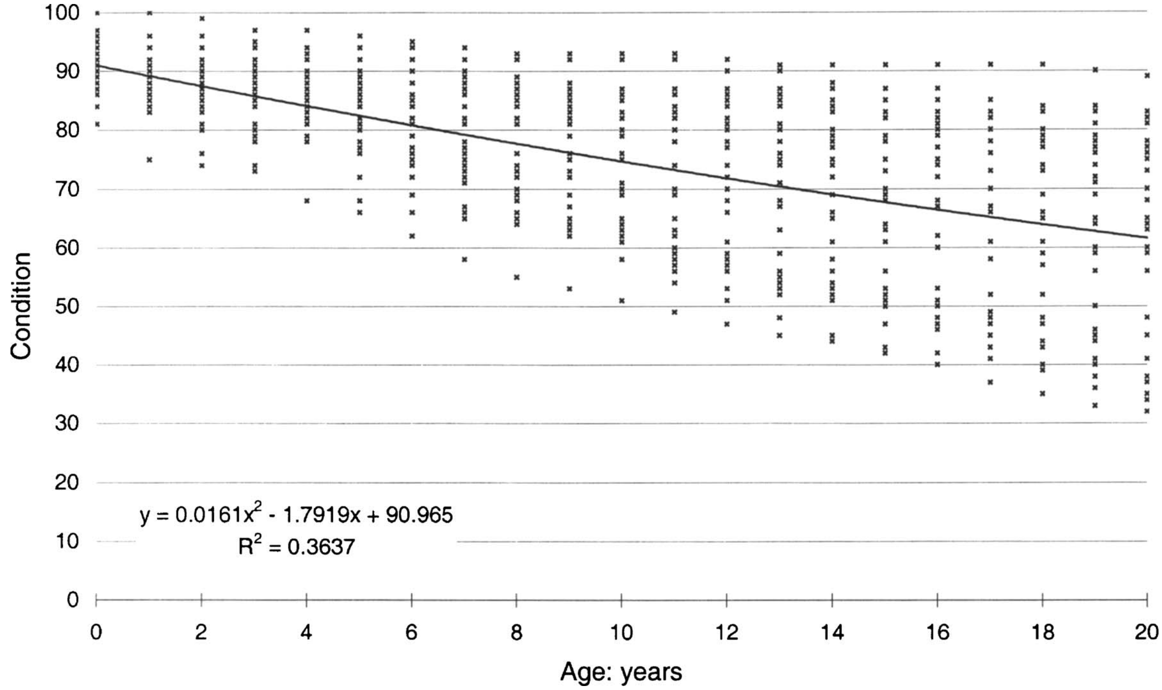

Fig. 5. Data Set 5. Scatter plot and regression curve.

that the standard deviation of the transition matrix fitted data is significantly greater than the standard deviation of the original data. However, the histograms in Fig. 13 clearly show that the optimization of the objective function for Methods A and B does not take into account the distributions of the original data; the distributions of the fitted data yield an average similar to the ordinates of the regression equation, but the distributions are completely different.

Conversely, the transition matrix fitted curve obtained with Method C, in comparison with Methods A and B, does not follow the regression curve as closely, but its standard deviation is closer to the standard deviation of the original data (see Fig. 7). Moreover, the histograms obtained with Method C present a similar shape to the distributions of the original data (see Fig. 13). As explained previously, the objective function of Method C aims at minimizing the differences between both distributions.

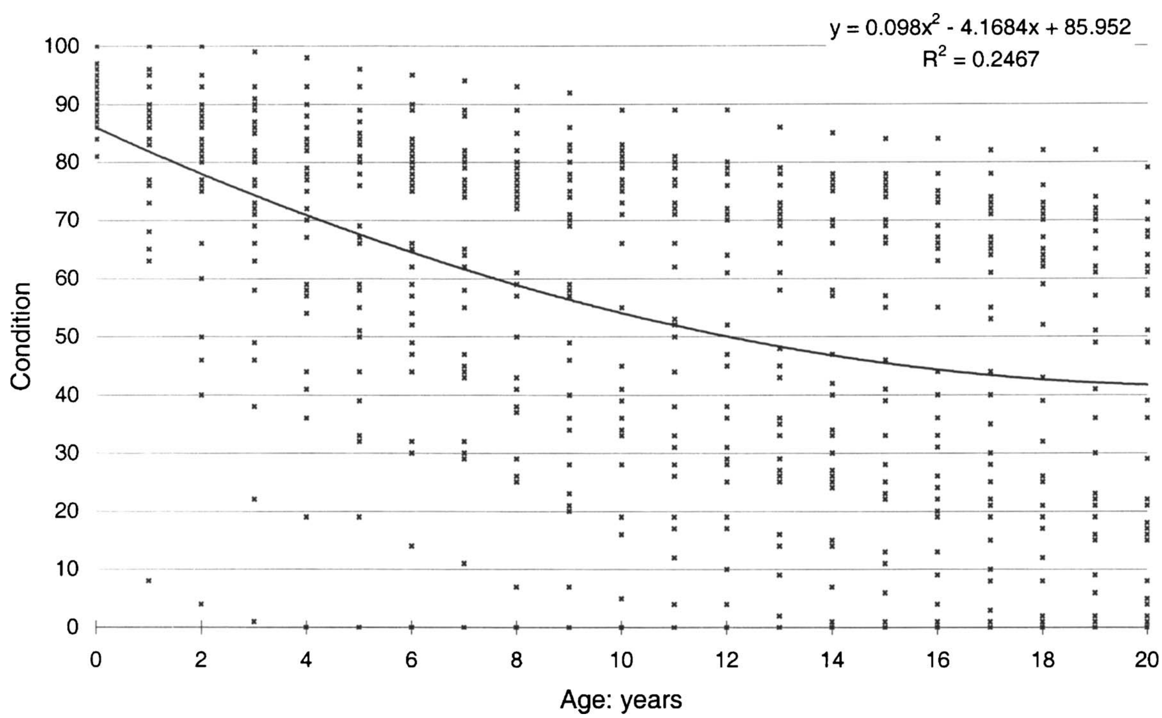

Fig. 6. Data Set 6. Scatter plot and regression curve.

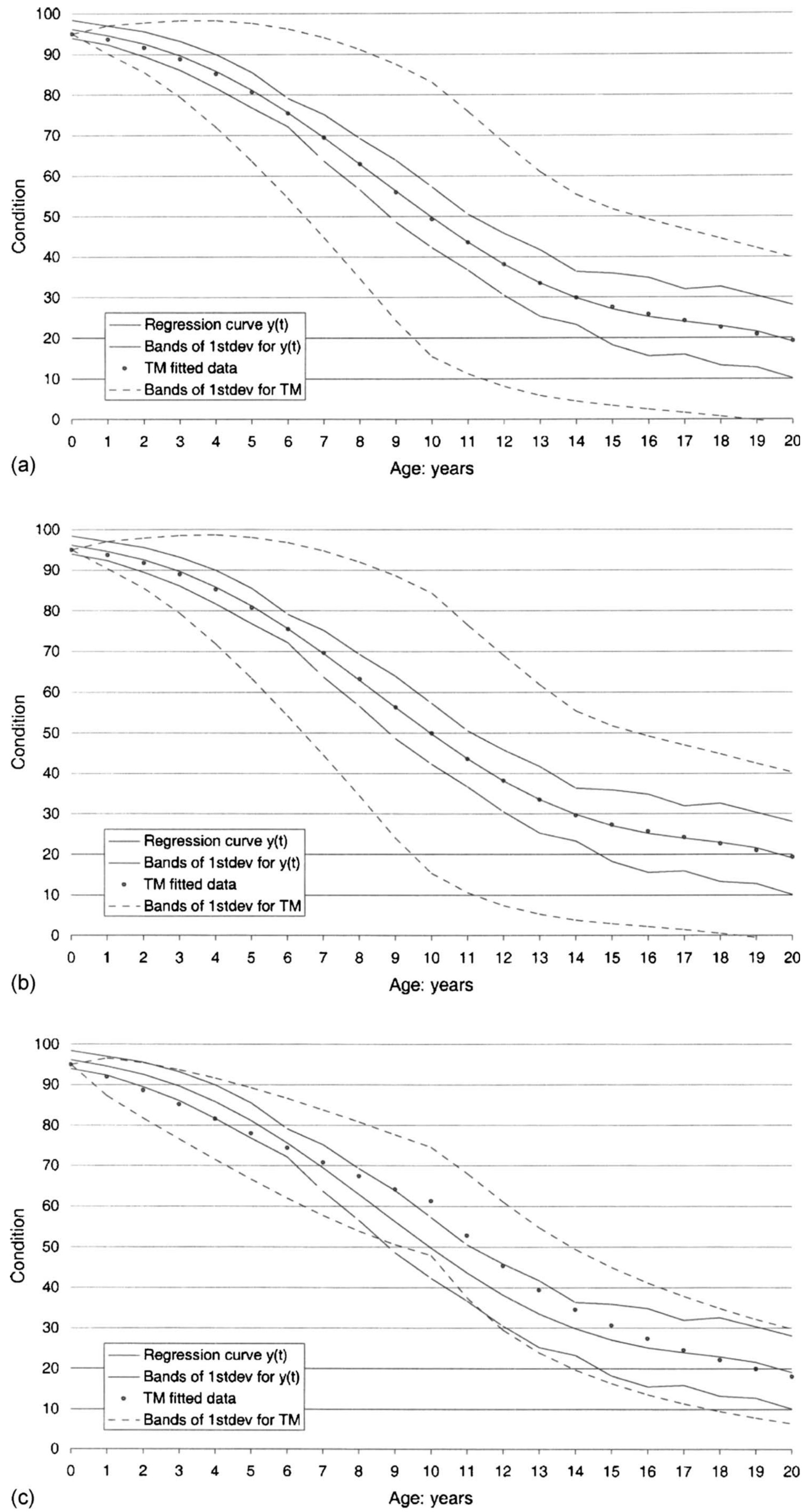

Fig. 7. Fitted data and bands of one standard deviation. Data Set 1. From top, Methods A, B, and C.

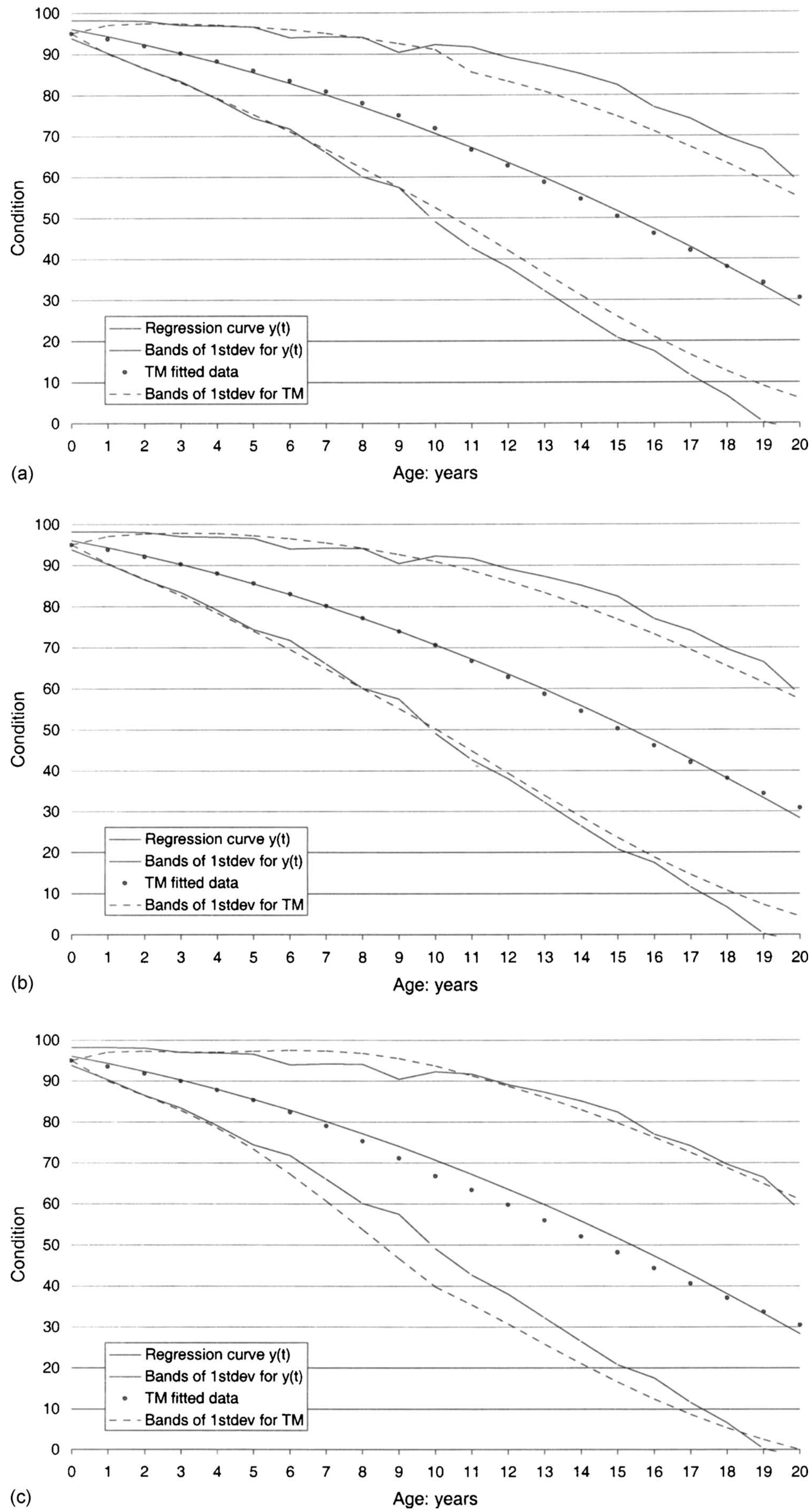

Fig. 8. Fitted data and bands of one standard deviation. Data Set 2. From top, Methods A, B, and C.

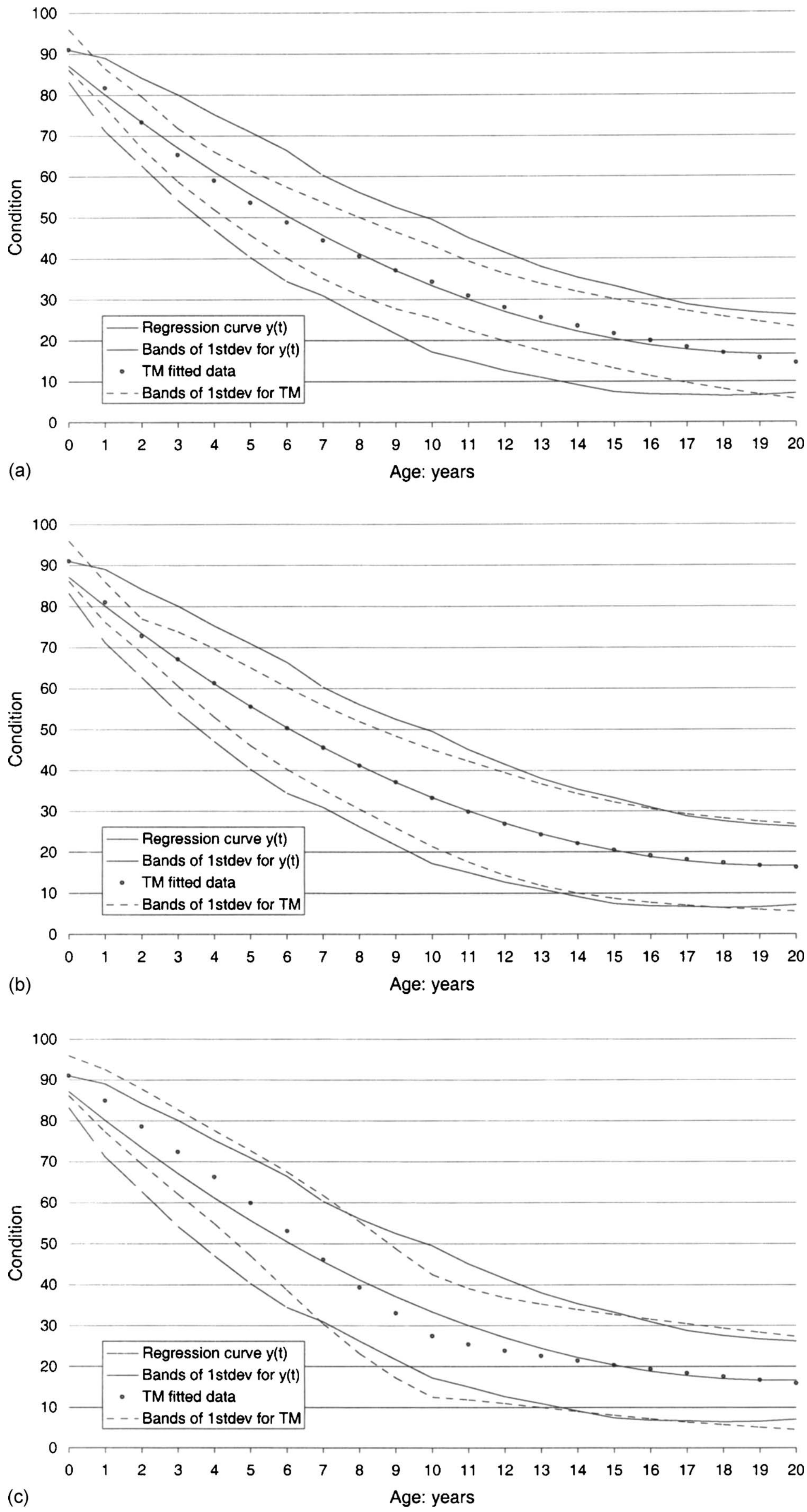

Fig. 9. Fitted data and bands of one standard deviation. Data Set 3. From top, Methods A, B, and C.

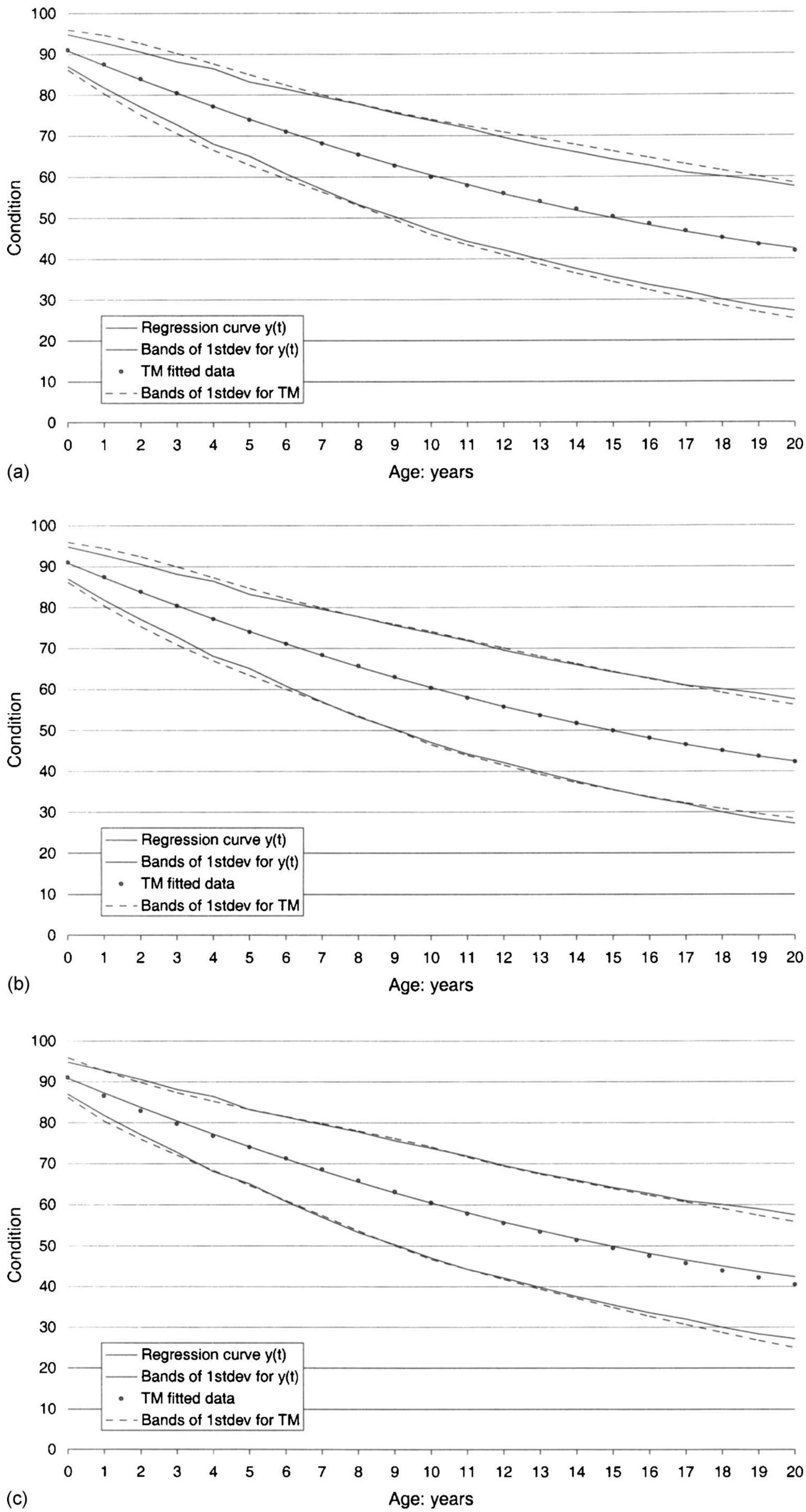

Fig. 10. Fitted data and bands of one standard deviation. Data Set 4. From top, Methods A, B, and C.

J. Transp. Eng. 2006.132:141-161.

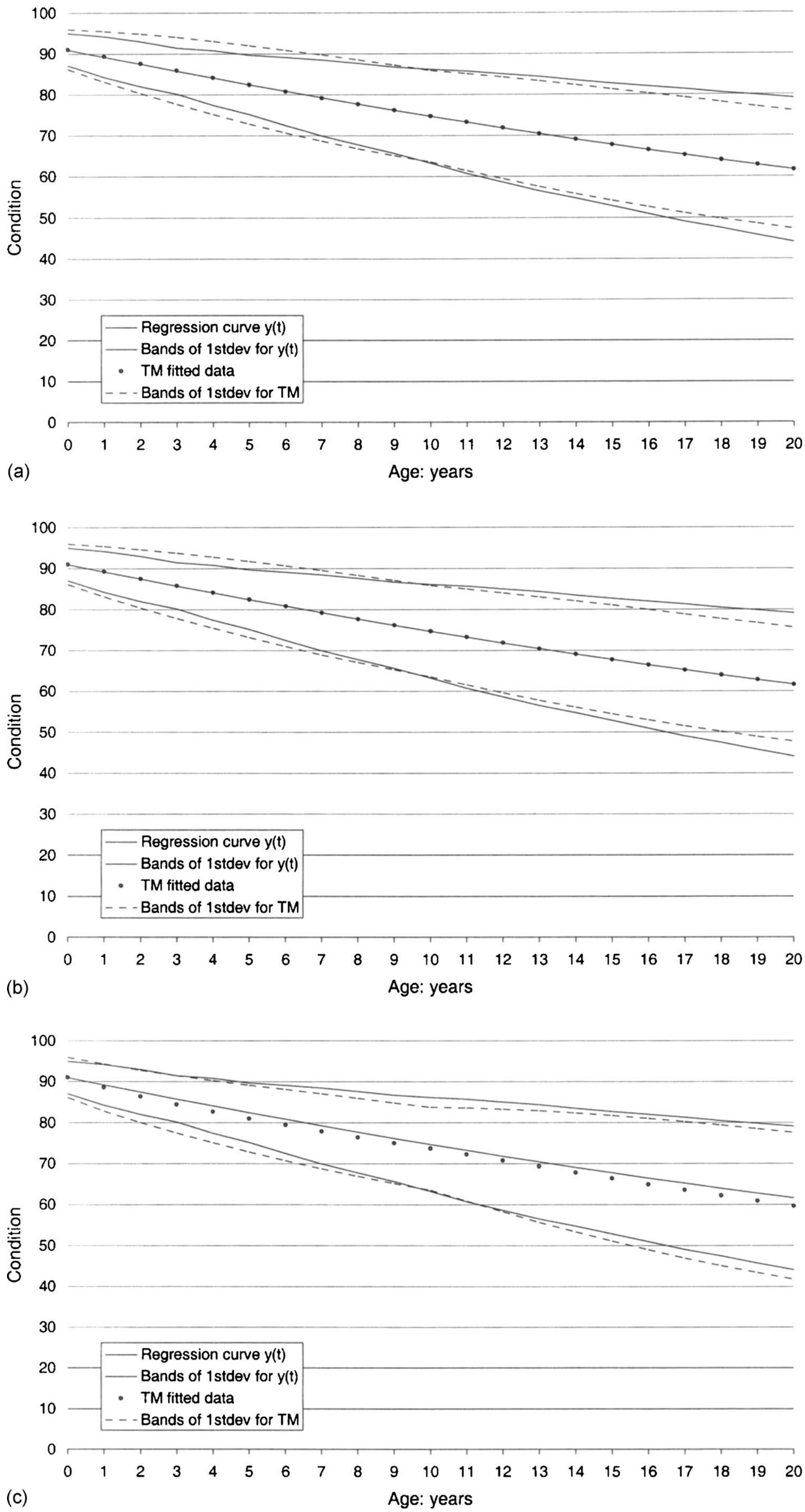

Fig. 11. Fitted data and bands of one standard deviation. Data Set 5. From top, Methods A, B, and C.

J. Transp. Eng. 2006.132:141-161.

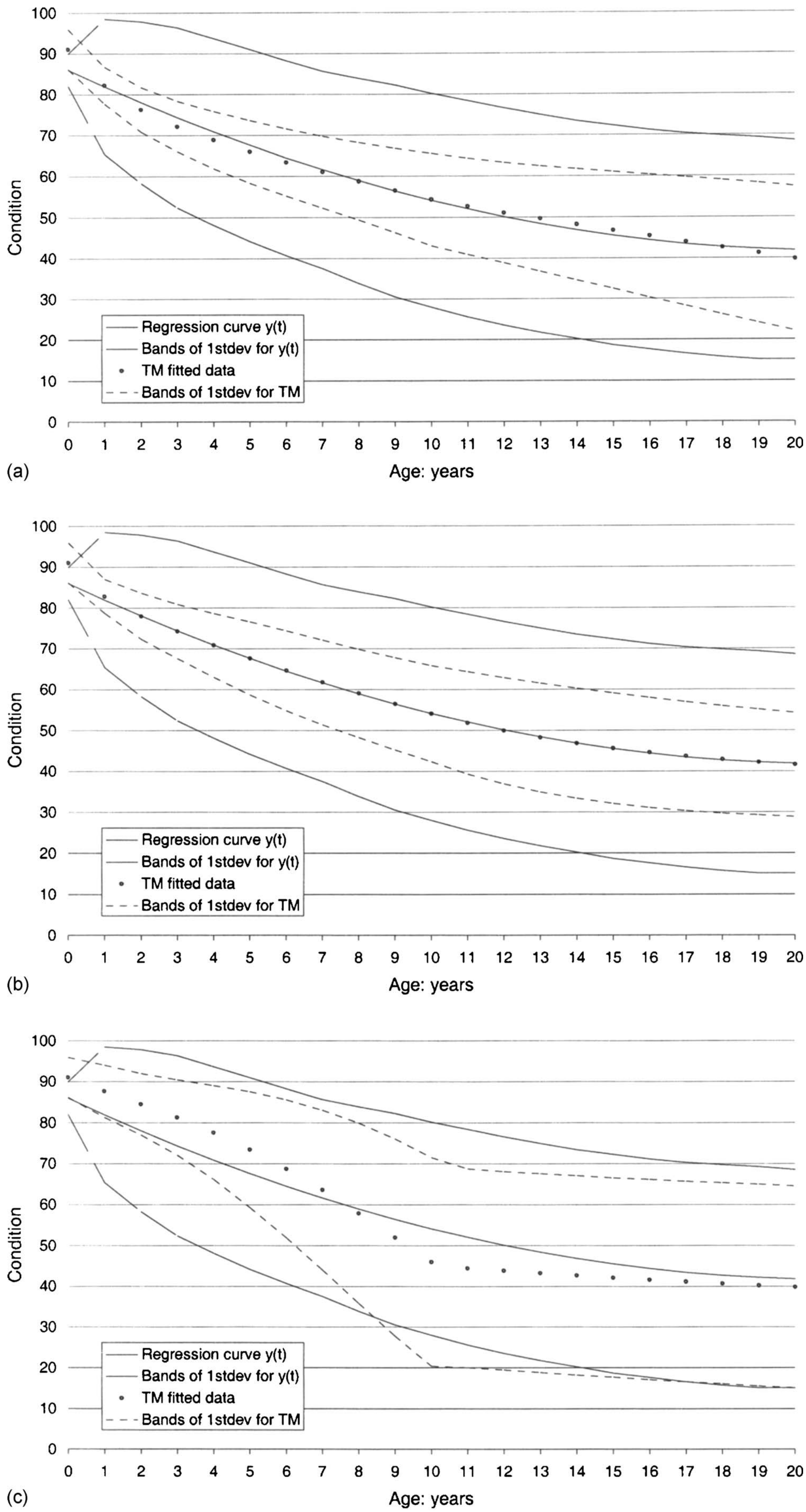

Fig. 12. Fitted data and bands of one standard deviation. Data Set 6. From top, Methods A, B, and C.

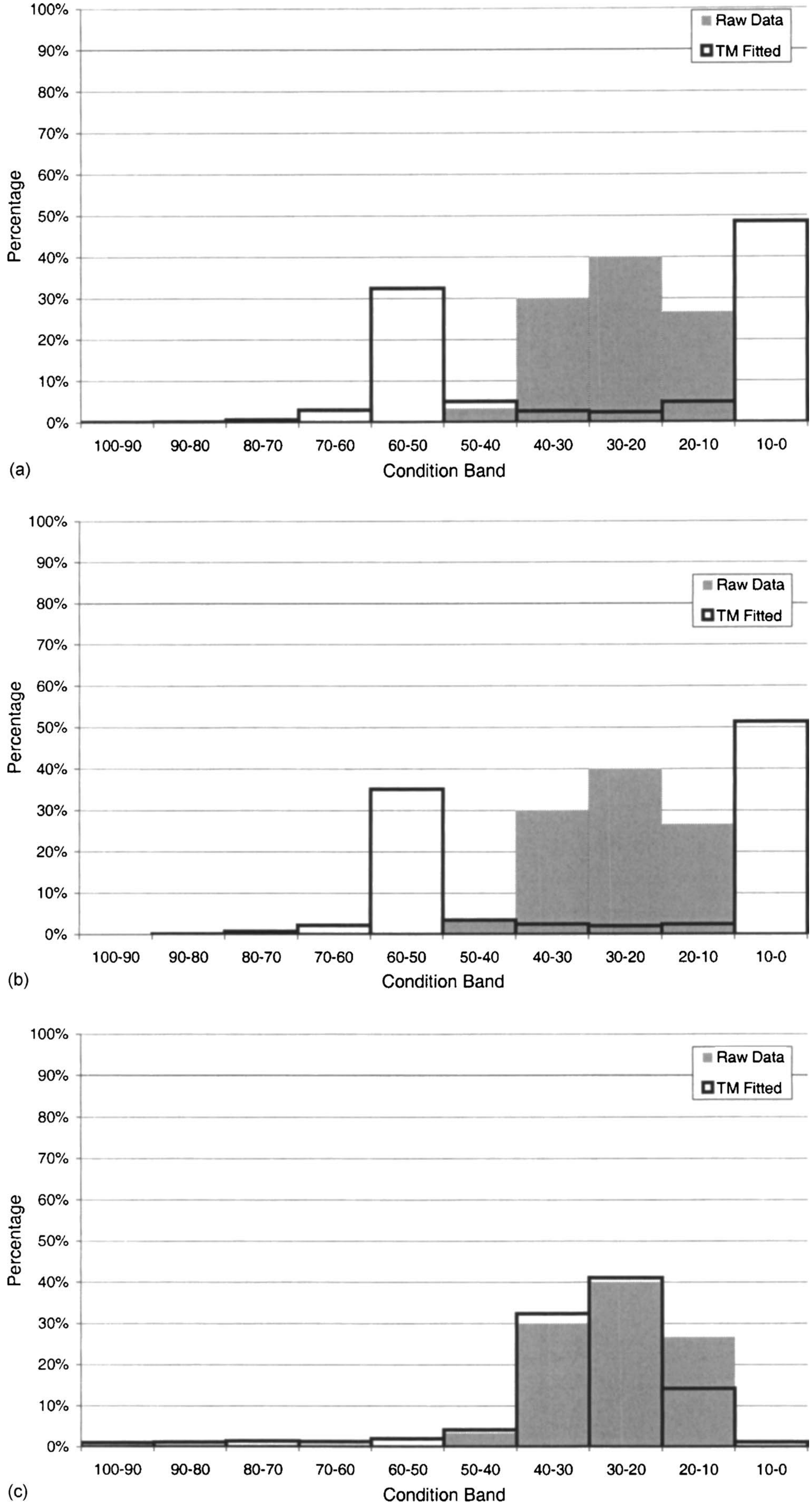

Fig. 13. Original versus fitted histograms. Data Set 1. Sample at Year 15. From top, Methods A, B, and C.

J. Transp. Eng. 2006.132:141-161.

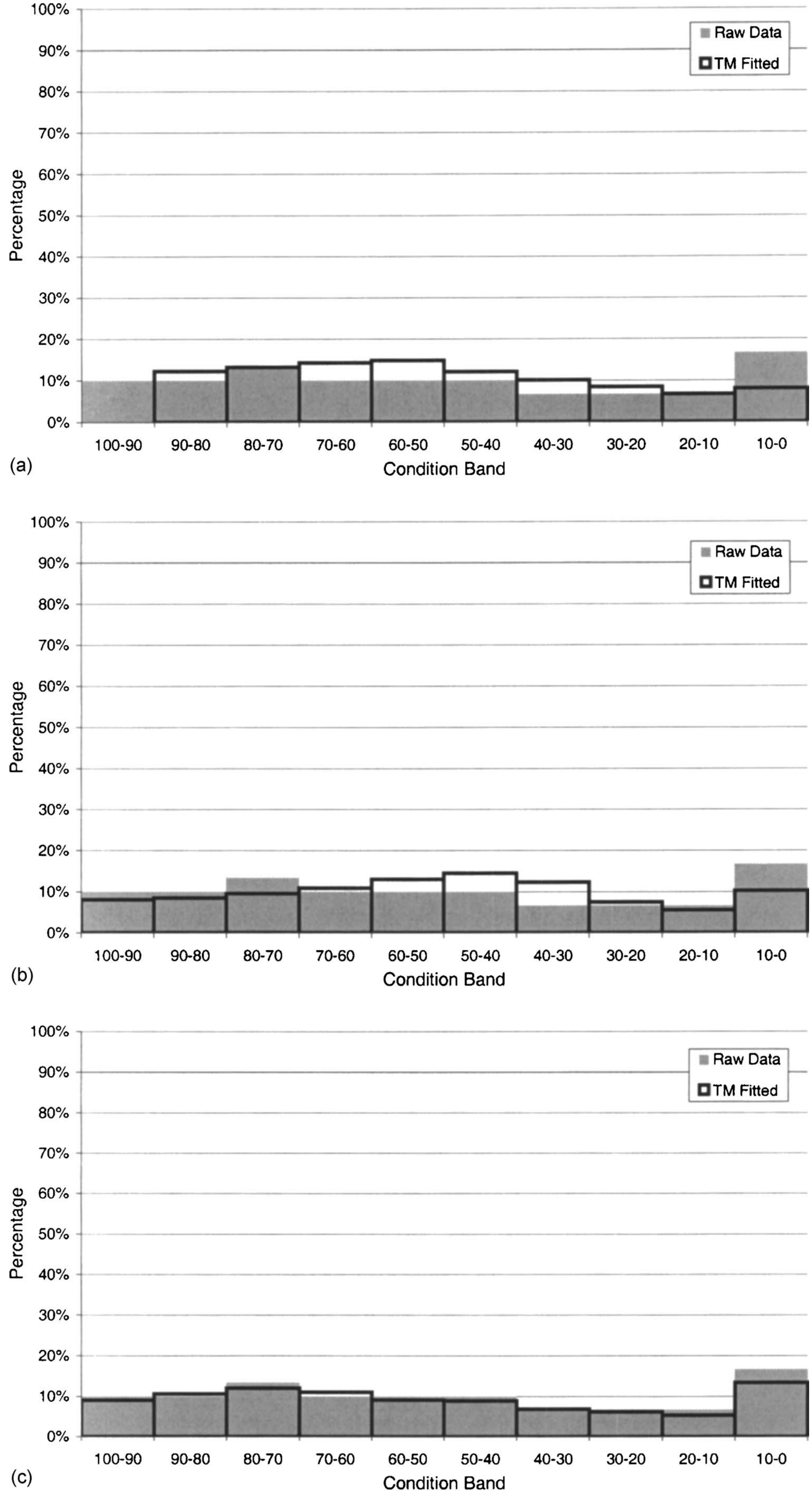

Fig. 14. Original versus fitted histograms. Data Set 2. Sample at Year 15. From top, Methods A, B, and C.

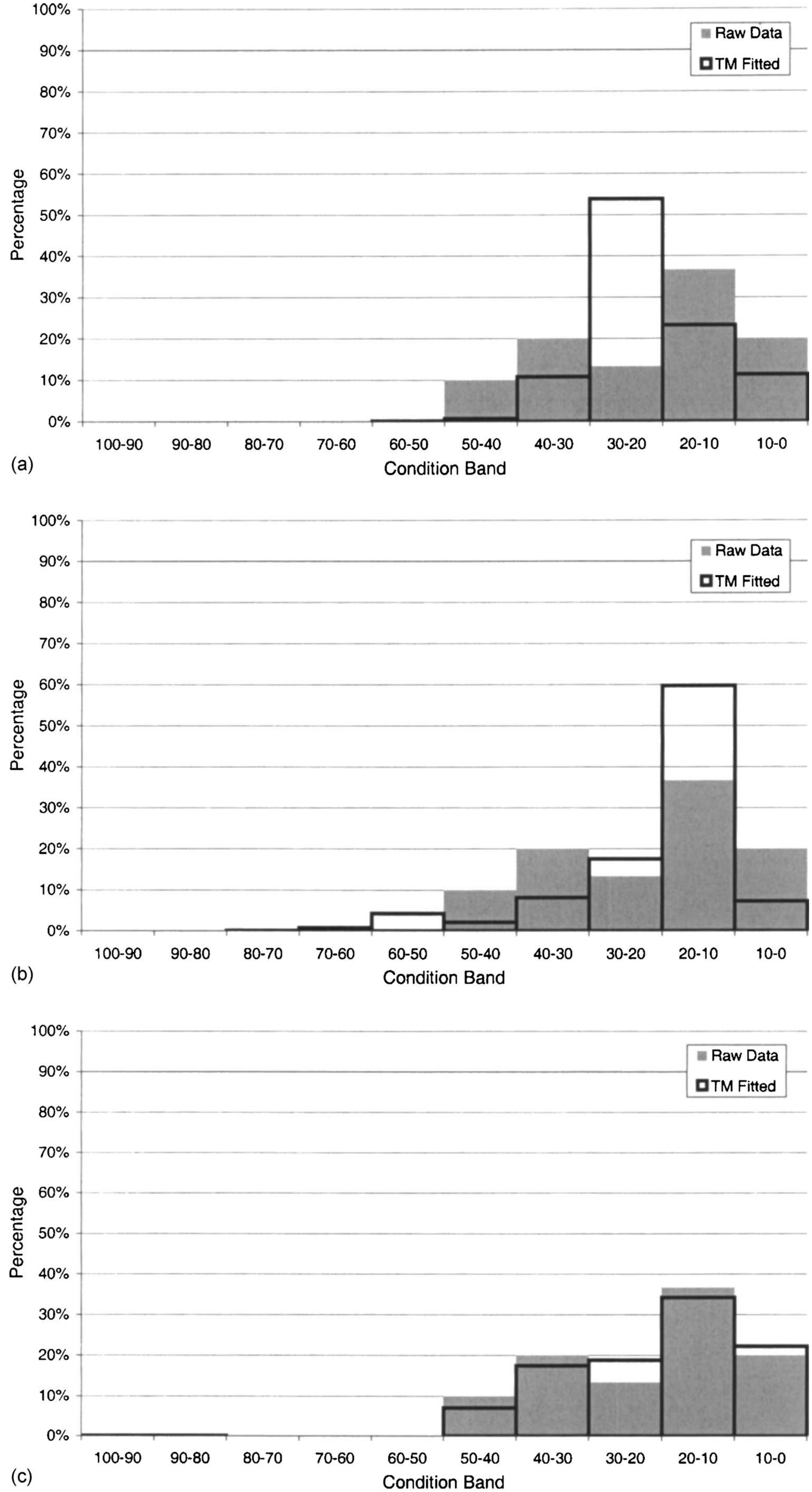

Fig. 15. Original versus fitted histograms. Data Set 3. Sample at Year 15. From top, Methods A, B, and C.

J. Transp. Eng. 2006.132:141-161.

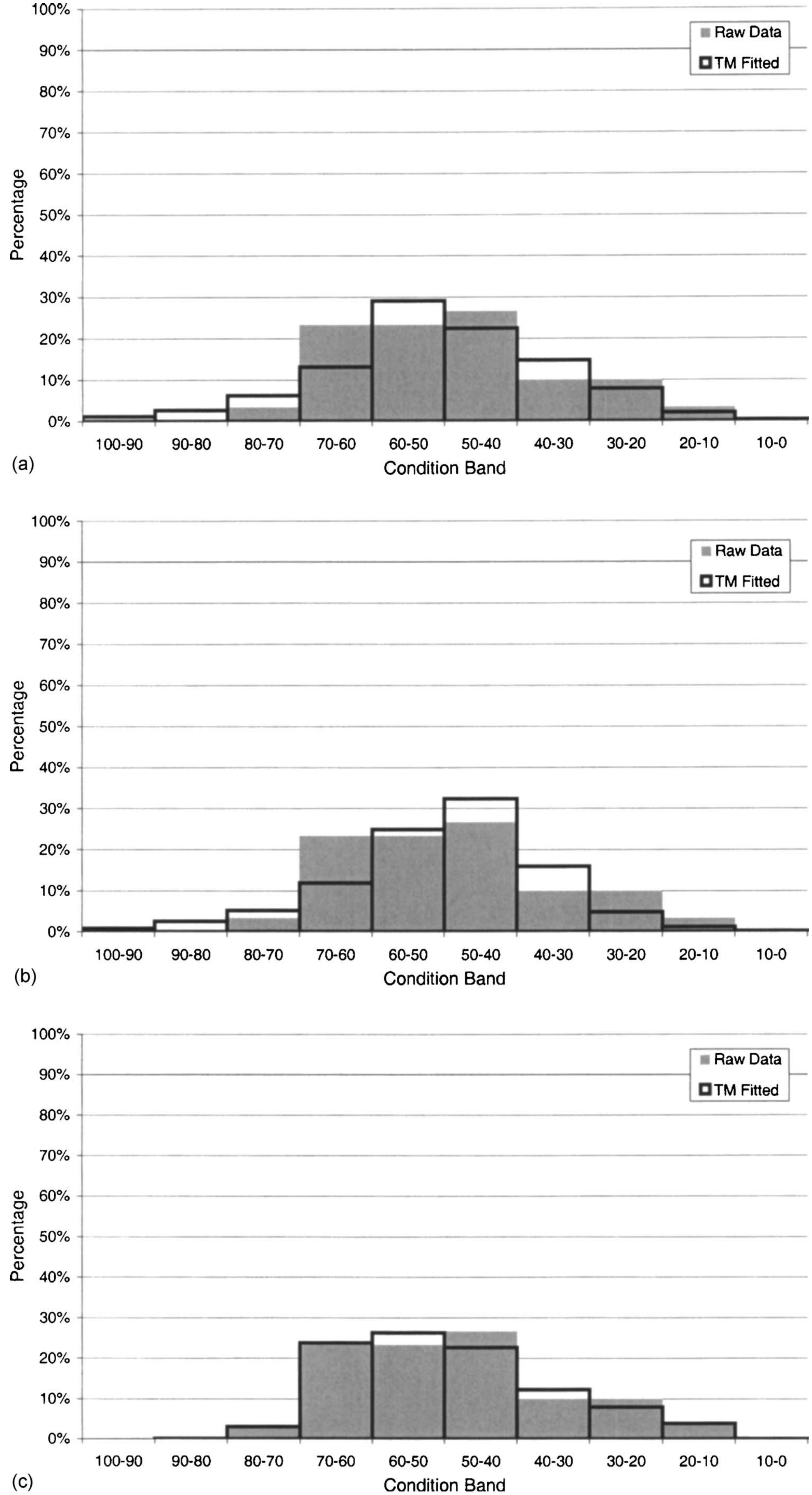

Fig. 16. Original versus fitted histograms. Data Set 4. Sample at Year 15. From top, Methods A, B, and C.

J. Transp. Eng. 2006.132:141-161.

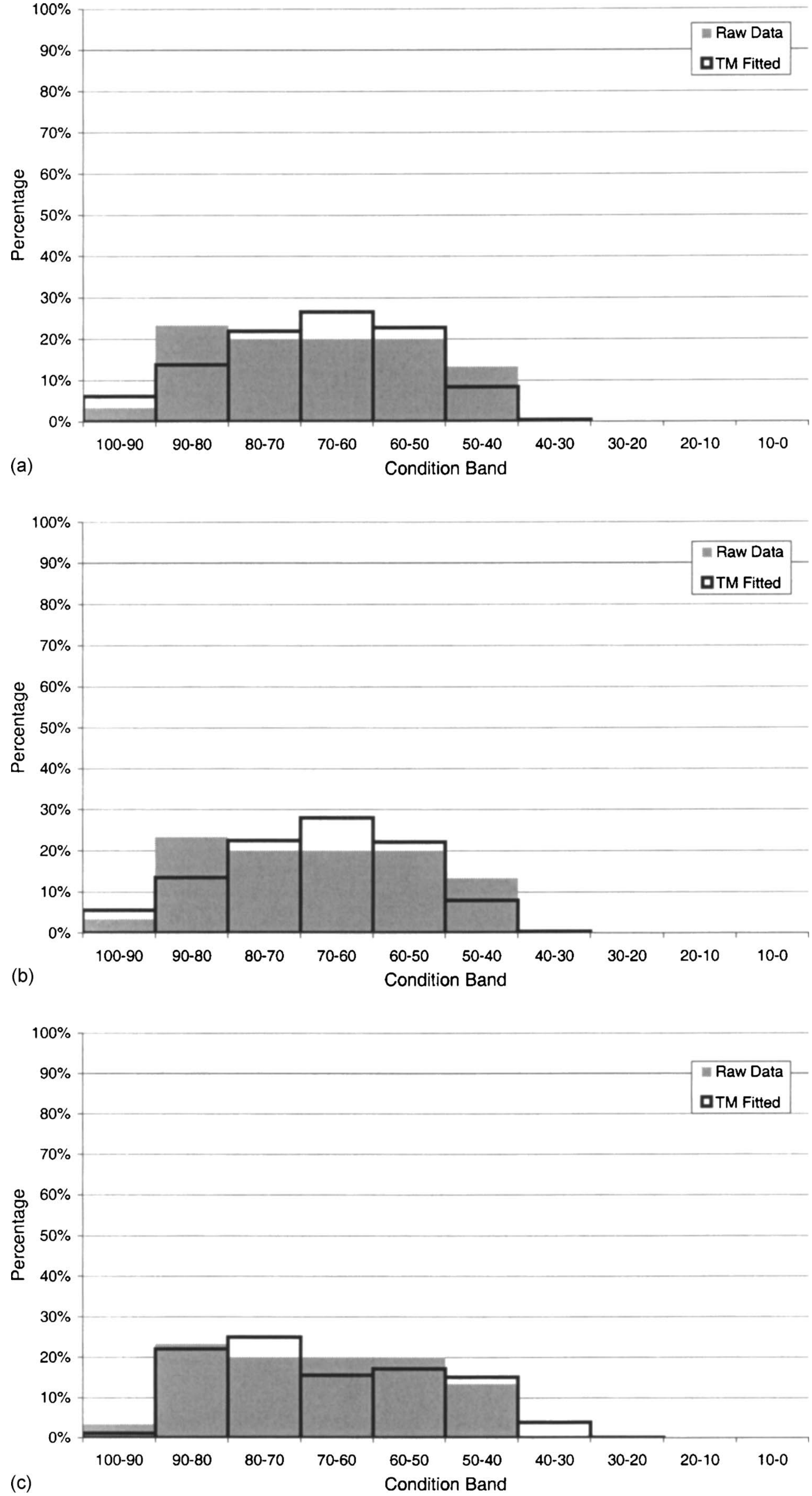

Fig. 17. Original versus fitted histograms. Data Set 5. Sample at Year 15. From top, Methods A, B, and C.

J. Transp. Eng. 2006.132:141-161.

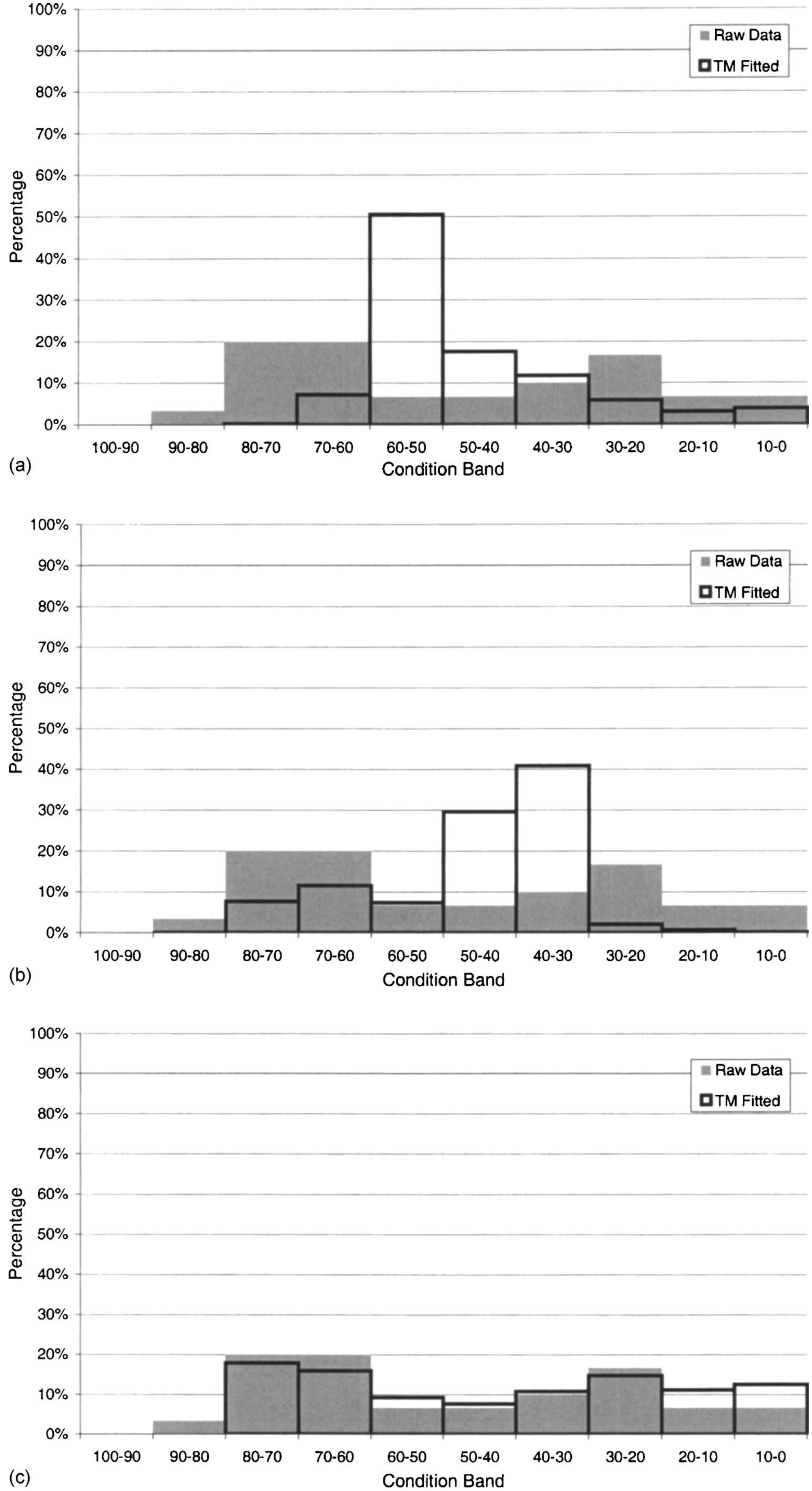

Fig. 18. Original versus fitted histograms. Data Set 6. Sample at Year 15. From top, Methods A, B, and C.

J. Transp. Eng. 2006.132:141-161.

Optimization Results for Data Set 2

It should be noted that DS2 was built from a predefined transition matrix and therefore it should be expected that the transition matrix fitted data are similar to the original data. The plots in Fig. 8 show that the transition matrix fitted curves obtained by Methods A and B are almost on the regression curve. The transition matrix fitted curve obtained by Method C is slightly displaced. The histograms in Fig. 14 show that the distributions obtained by Methods A and B closely reflect the original distributions and the distribution obtained by Method C is almost identical to the original distribution.

Optimization Results for Data Sets 3, 4, and 5

The results for these three data sets are analyzed together as the data set generation process was similar for each of them. The transition matrix fitted curve closely follows the regression curve for DS3 in Methods A and B. It was slightly displaced for Method C. This may be seen in Fig. 9. The distributions yielded by Methods A and B presented a different shape compared to the original distributions. The distribution yielded by Method C was, in contrast, similar to the original. This may be observed in Fig. 15.

DS4 and DS5 presented a similar behavior; the transition matrix fitted curve and standard deviations closely followed those from the original data for the three methods. The distributions obtained from the transition matrix and the original distributions were also comparable. This may be observed in Figs. 10 and 16 for DS4 and in Figs. 11 and 17 for DS5.

Optimization Results for Data Set 6

It should be noted that DS6 was built in such a way that the sites that composed the artificial network deteriorated in a completely random manner. This is reflected in the shapeless form of the distributions and their large standard deviations. Methods A and B, as in the previous examples, were quite efficient in yielding a deterioration curve that closely follows the original one. They were not capable, though, of producing similar standard deviations. This may be observed in Fig. 12. The distributions shown in Fig. 18, however, show that the distributions from the transition matrix process are different from the original ones, for Methods A and B.

Conversely, Method C yielded a transition matrix distribution that closely reflects the original one despite the displacement observed between the transition matrix fitted curve and the regression curve. The standard deviations were, however, reasonably similar.

Discussion of Results

From the results presented above, it may be observed that Methods A and B yielded, in most cases, a close fit to the original curve. However, in a number of cases, they did not yield the expected standard deviation, and in others the distributions obtained from the transition matrix process were completely different from the original distributions. In contrast, in all cases, Method C yielded a distribution closer to the original distributions than either Method A or B. Moreover, the distributions determined from Method C were also comparable, in many cases almost identical, to the original ones, although the transition matrix fitted curve was not always as close to the regression curve as in the aforementioned methods.

The displacements observed in the transition matrix fitted curves obtained by Method C might be explained by one or both of the following. First, although the transition matrix fitted curve is calculated as the average of the transition matrix distributions, it should be noted that the value of the regression curve for each point in time, t, is not necessarily equal to the average of the condition distributions. Second, even if the values of the regression curve were equal to the average of the condition distributions, as is the case for DS1, some accuracy is lost when the data are aggregated into bands of condition.

Conclusions

From the methods outlined above, it appears that Method C provides the most appropriate method for determining a TPM. However, to effect this method, the original distributions of condition are required to enable a TPM to be derived. In practice, this is probably not practical as most deterministic deterioration relationships provide a deterministic curve. This would lead one to believe that Method B is the most practical in terms of implementation. However, where a deterministic curve is available without an indication of its scatter (confidence or confidence interval), it can reliably be assumed that the curve has not been properly calibrated in the first instance. It should be appreciated that the predictive power of the transition matrix depends solely on the accuracy of the original deterministic model and the confidence placed in the deterministic model is reflected by its measure of scatter. Where available, a deterministic curve and its measure of scatter can be utilized to implement the preferred Method C. In such a scenario, it is assumed that the distribution of condition can be represented by a normal distribution defined by its average (the ordinate of the curve) and its standard deviation as a measure of scatter. Indeed, this has been taken forward since the work was completed and encapsulated in an analytical tool called the Transition Matrix Calculator. Details of this work along with an example of its use in practice are included in Costello et al. (2005).

Further Work

All comments thus far have been made independent of whether the road condition is being modeled by stationary or nonstationary chains. This study included both a stationary chain, from Years 0 to 10, and a nonstationary chain, when the analysis as a whole is taken into consideration. Both methods appear to have performed satisfactorily, however extensive studies are planned using the newly developed analytical tool, Transition Matrix Calculator, in order to develop a sound understanding of the relationships involved. These results will be published as they become available.

In addition, the method presented may be improved by employing more sophisticated search algorithms. The nonlinear optimization algorithm used lacks, as for any other of its type, the guarantee of reaching a global optimum. Other number-search methods, such as genetic algorithms, could be investigated in further research to determine whether more sophistication in the search algorithm improves the optimization results.

J. Transp. Eng. 2006.132:141-161.

References

AASHTO. (1985). *AASHTO guide for the design of pavements structures*, AASHTO, Washington D.C.

Butt, A. A., Shahin, M. Y., Carpenter, S. H., and Carnahan, J. V. (1994). "Application of Markov process to PMS at network level." *Proc., 3rd Int. Conf. on Managing Pavements*, Vol. 2, National Academy Press, Washington, D.C., 159–172.

Costello, S. B., Ortiz-Garcia, J. J., and Snaith, M. S. (2005). "Development of the transition matrix calculator." *Proc., 10th Int. Conf. on Civil, Structural, and Environmental Engineering Computing*, Civil-Comp, Stirling, U.K.

Costello, S. B., Snaith, M. S., Kerali, H. G. R., Tachtsi, V. T., and Ortiz-Garcia, J. J. (2006). "Stochastic model for strategic assessment of road maintenance." *Proc., Institution of Civil Engineers, Transport*, in press.

Darter, M. I., and Hudson, W. R. (1973). "Probabilistic design concepts applied to flexible pavement system design." *Rep. No. 123-18*, Center for Highway Research, Univ. of Texas at Austin, Austin, Tex.

Fylstra, D., Lasdon, L., Watson, J., and Warren, A. (1998). "Design and use of the Microsoft Excel Solver." *Interfaces*, 28(5), 29–55.

Isaacson, D. L., and Madsen, R. W. (1976). *Markov chains: Theory and applications*, Wiley, New York.

Kerali, H. R., and Snaith, M. S. (1992). "NETCOM: The TRL visual condition model for road Networks." *Contractor Rep. No. 321*, Transport Research Laboratory, Crowthorne, U.K.

Kulkarni, R. B. (1984). "Dynamic decision model for a pavement management system." *Transportation Research Record 997*, Transportation Research Board, Washington, D.C., 11–18.

Li, N., Haas, R., and Xie, W. (1997). "Development of a new asphalt pavement performance prediction model." *Can. J. Civ. Eng.*, 24, 547–559.

Ortiz-García, J. J. (2000). "Strategic planning of highway maintenance: Condition standards and their assessment." PhD thesis, Univ. of Birmingham, Birmingham, U.K.

Paterson, W. D. O. (1987). "Road deterioration and maintenance effects: Models for planning and management." *The Highway Design and Maintenance Standards Model (HDM-III)*, Vol. III, World Bank, Washington, D.C.

Robinson, R., Snaith, M. S., and Danielson, U. (1998). *Road maintenance management: Concepts and systems*, Macmillan, Basingstoke, U.K.

Thompson, P. D., Neumann, L. A., Miettinen, M., and Talvitie, A. (1987). "A micro-computer Markov dynamic programming system for pavement management in Finland." *Proc., 2nd North American Conf. on Managing Pavements*, Ontario Ministry of Transportation and Communications, Vol. 2, Toronto, 2.241–2.252.

J. Transp. Eng. 2006.132:141-161.