# Optimization of data collection needs for manual and automated network-level pavement condition ratings based on transverse variability and neural networks

**Ahmed Shalaby and Alan Reggin**

**Abstract:** The paper deals with two approaches to optimizing pavement condition surveys for the urban pavement network of the City of Winnipeg, Manitoba. First, a nonparametric statistical test was applied to assess the transverse variability of the data. The test compared the ratings for one lane with those of all lanes of each segment. The test concluded that the medians of the two groups are equal at a 92% confidence interval and that there are observed biases in the data. The bias can be eliminated if the surveyed lane is selected randomly. The second approach was to predict visual survey scores from automated (laser-based) measurement of rut depth and international roughness index (IRI). A resilient back-propagation algorithm was selected, and the Kappa coefficient was used to examine the strength of the agreement. The results showed that only moderate agreement was achieved and that additional data elements are required to improve the predictive ability of the model.

*Key words:* international roughness index (IRI), rutting, cracking, spalling, pavement management system (PMS), Kappa coefficient, distress surveys.

**Résumé :** Le présent article aborde deux approches pour optimiser les relevés de l'état des chaussées pour le réseau de chaussées urbaines de la ville de Winnipeg, au Manitoba. Premièrement, un test statistique non paramétrique a été utilisé pour évaluer la variabilité transverse des données. Le test comparait les cotes d'une voie aux cotes de toutes les voies de chaque segment. Le test concluait que les médianes des deux groupes étaient égales à un intervalle de confiance de 92 %, et que certaines données sont biaisées. Ces biais peuvent être éliminés si la voie étudiée est sélectionnée au hasard. La deuxième approche consistait à prédire les cotes des relevés visuels à partir de mesures automatisées (au laser) de la profondeur des ornières et de l'indice de rugosité international (IRI). Un algorithme de rétropropagation résilient a été choisi et le coefficient Kappa a été utilisé pour examiner le degré de corrélation. Les résultats ont montré que seule une conformité modérée a été atteinte et que des données additionnelles sont requises pour améliorer la valeur de prévision du modèle.

*Mots-clés :* indice de rugosité international (IRI), formation d'ornières, fissuration, effritement, système de gestion routière, coefficient Kappa, relevés des dommages importants.

[Traduit par la Rédaction]

## Introduction

A pavement management system (PMS) can facilitate the efficient and cost-effective planning of maintenance and rehabilitation activities at both the network and project levels and encourage decision-making based on present and future predicted conditions and life-cycle cost. The system can generate alternate scenarios based on various levels of funding and different management approaches. Pavement condition surveys are the foundation of the PMS of the City of Winnipeg. The PMS makes use of visual ratings of surface condition that is reported as general condition scores, cracking scores, and joint spalling scores.

The City of Winnipeg road network consists of a regional road network of urban arterials and a residential street network of collectors and local streets. Both networks have been surveyed by rating crews for a number of years. One concern with the visual inspection of the regional street network is that it involves undesirable exposure to heavy traffic volumes operating at posted speeds of up to 80 km/h. Using two-person rating crews, the visual inspection of the regional street network requires four crews

Received 13 May 2005. Revision accepted 20 August 2006. Published on the NRC Research Press Web site at http://cjce.nrc.ca on 14 March 2007.

**A. Shalaby[1] and A. Reggin.[2]** Department of Civil Engineering, The University of Manitoba, 15 Gillson St., Winnipeg, MB R3T 5V6, Canada.

Written discussion of this article is welcomed and will be received by the Editor until 30 June 2007.

[1]Corresponding author (e-mail: shalabya@cc.umanitoba.ca).
[2]Present address: EBA Engineering Consultants Ltd., 9th Floor, 1066 West Hastings St., Vancouver, BC V6E 3X2, Canada.

 doi:10.1139/L06-126 

working for 8 weeks each to survey only representative segments of each road. The data are typically collected on an annual basis.

Many highway agencies are adopting automated laser-based surveys because they can be conducted more productively and without exposing the rating crews to heavy or high-speed traffic. The surveys are conducted from a vehicle travelling at highway speed and without interfering with traffic flows. Another benefit is that the entire road including all lanes of travel can be surveyed. The data are typically reported at 10–100 m intervals. The City of Winnipeg is examining the potential for replacing visual rating of regional streets with automated laser scanning surveys. The purpose of this research is to determine if the manual survey ratings can be inferred from the automated data. Although visual surveys can be biased as a result of differences between data collection personnel and varying lighting and weather conditions that affect the ability to perceive certain distresses, there is a large historical database of visual condition ratings, and it is desired to retain as much information as possible on cracking and joint spalling because they tend to drive maintenance programs and performance models.

Two methods of reducing the volume of collected data are evaluated. The first is to down-sample by collecting data on fewer lane-kilometres of the network. Sampling can often provide sufficient data for the network-level PMS but not for project-level analysis and design. The analysis of the transverse variability of pavement distresses is conducted to determine if a single lane can be taken as representative of all of the lanes of urban multilane segments, and therefore reduce the volume of collected data.

The second method of minimizing data collection needs is to collect fewer data elements. This is achieved by eliminating the collection of elements that are of little use and elements that can be reasonably estimated from other variables. Artificial neural networks (ANNs) were constructed to predict manual ratings of cracking and spalling based on the automated international roughness index (IRI) and rut depth data and to determine if the manual ratings could be predicted from laser-based automated data. Advancements have been made in the automated collection of surface distress data including spalling and cracking, and there may be the potential to directly collect those data elements with high-speed visual surveys using event keyboards. The intent of this research, however, is to determine if cracking and spalling data could be inferred from IRI and rut depth, which are not visually acquired and are not biased by time of day or lighting conditions.

## Data sources

Parallel automated and manual surveys of the regional street network were conducted in 2002. The automated survey was conducted by a consultant using laser-based equipment, and the city crews carried out the visual condition surveys. The regional street network comprises all of the arterial streets and several of the high-volume collector streets within the city limits. This 626 km of multilane street network utilizes three pavement types. The prevailing pavement type is termed asphalt over portland cement concrete (APC), which accounts for 72% of the network. APC is a composite pavement in which an old concrete pavement has been overlaid with asphalt concrete. Portland cement concrete (PCC) pavement makes up an additional 22% of the regional network and is generally jointed plain concrete pavement. The remaining 6% of the network is asphalt concrete (AC). This paper utilizes data collected on the APC and PCC segments only because of the limited volume of data on AC pavements. Thus, 94% of the network is included in the analysis.

## Visual rating data elements

Pavement surface condition ratings are collected by two-person crews according to a well-defined surface condition rating manual. Crews assign each segment three rating scores based on general condition, cracking, and spalling.

### General condition

The four categories of general condition are new, good, fair, and poor. New streets are those which are new or were recently overlaid. Good streets usually require minor maintenance, fair streets usually require major maintenance or rehabilitation, and poor streets need reconstruction and are thus only maintained to ensure the safety of the travelling public, and not to prevent further deterioration.

For cracking and spalling data, rating crews select a gauging area that is representative of the prevailing conditions of the control section. The gauging area is chosen so that items such as railway crossings, manholes, and utility patches are excluded unless they are prevalent over the segment. The rating crew conducts a detailed survey of 10 slabs per lane in the representative gauging area.

### Cracking and joint spalling data

The rating crew collects the number of cracks in each slab, the severity of cracks, and the number of cracked slabs in the gauging area. Cracks that extend across the full width or length of the slab are considered to be working cracks and are included in the rating, whereas shrinkage cracks are not included in the rating. Each slab is assigned a crack type based on the number of pieces the slab is broken into and a crack severity based on the size of the cracks and the condition of the sealant. Table 1 provides definitions of each crack type and severity. For example, a slab containing two unsealed cracks with an average width of 20 mm has a cracking label of BS (i.e., branch type with slight severity). The same rating system is used for APC and PCC pavements.

The extent of cracking and joint spalling in a segment is based on the number of cracked slabs and spalled joints in the segment as defined in Table 2. Each segment receives a cracking score and a joint spalling score based on the predominant distress and the extent. For example, a segment in which 60% of the slabs are cracked and most of the cracked slabs have two or three cracks of extreme severity has a cracking score of BX3. A segment in which 15% of the joints have some type of spalling and where the predominant spall has a slight severity has a joint score of S1.

**Table 1.** Definition of distress types and severities.

| Label | Definition |
|---|---|
| **Crack type** | |
| N (none) | No cracking |
| R (random) | Slab has cracked into two pieces |
| B (branch) | Slab has cracked into three or four pieces |
| P (pattern) | Slab has cracked into five or more pieces |
| **Crack severity** | |
| N (new) | Cracks that are sealed with 75% or more of the sealant in good condition |
| S (slight) | Cracks are not sealed or more than 25% of the sealant is damaged; width <25 mm |
| M (moderate) | Cracks are not sealed or more than 25% of the sealant is damaged; 25 < width < 40 mm |
| X (extreme) | Fully developed cracks are not sealed or more than 25% of the sealant is damaged; vertical or horizontal displacement >40 mm |
| **Joint spall severity** | |
| N (new) | Joints are tight and may have a few minor aggregate pop-outs adjacent to the joint |
| S (slight) | Parallel cracks adjacent to the joint with some of the cracks forming small pieces, which are tight, or aggregate pop-outs adjacent to the joint for more than 25% of its length |
| M (moderate) | Cracks have formed into a pattern and the pieces are loose or missing; the spall is less than 25% of the joint length |
| X (extreme) | Pieces are extremely loose or missing, and the full depth of the spall can be seen; the spall is more than 25% of the joint length |

**Table 2.** Crack and spall extent definitions.

| Extent | Lower (%) | Upper (%) |
|---|---|---|
| **Cracking** | | |
| 0 | 0 | ≤5 |
| 1 | 5 | ≤20 |
| 2 | 20 | ≤35 |
| 3 | 35 | ≤100 |
| **Joint spalling** | | |
| 0 | 0 | ≤5 |
| 1 | 5 | ≤20 |
| 2 | 20 | ≤40 |
| 3 | 40 | ≤100 |

## Automated data collection of international roughness index and rut depth

Roughness and rut data were collected for the entire length of one lane per direction in each segment. In all multilane segments, the second lane from the curb was surveyed and in this paper is termed the automated survey lane. An inertial profiler collected a longitudinal profile of each wheel path. Profile elevations are then processed and summarized as IRI, which was reported at 20 m intervals. Rut depth data were collected for each wheel path. The rutting index used was a simulated 1.8 m straightedge. Rut depth was mathematically calculated from a measured transverse profile and reported at the same 20 m intervals as IRI.

## Transverse variability of surface distress

Analysis of the transverse variability of surface distress is useful for determining data collection needs of manual and automated pavement surveys. In 2002, manual distress data were collected for all lanes of the regional street network, whereas the automated data were collected for only the second lane from the curb. The datasets compared in this analysis are the manually collected data for all lanes and the manually collected data for the second lane from the curb. The second lane from the curb was selected because it is the only lane that is also surveyed using the automated process.

An analysis of the transverse variability was conducted to examine the difference between the distresses in the two datasets. The objective of the analysis is to determine if a single lane can be representative of all of the lanes of multilane segments, and therefore reduce data needs.

## Sampling and analysis

The analysis sample includes 54 randomly selected sections with a total length of 49 km. Twenty-five sections were two-lane, 16 were three-lane, and 13 were four-lane. For each section and each survey (second lane from the curb and all lanes), the percentage of each type and severity of distress were calculated. Figure 1 shows a comparison of cracking data, and Fig. 2 a comparison of spalling data.

The data indicate that there is a slight shift in the type of cracking, with more slabs having no cracking on the automated survey lane and more slabs having pattern cracking on all lanes. There is also a shift in the crack severity, with more prevalent slight and extreme cracking on all lanes and more new cracks on the automated survey lane. There are more joints with moderate and extreme spalling on all lanes and more joints with new or slight spalling on the automated survey lane.

## Statistical hypothesis testing

A nonparametric statistical analysis was conducted to determine if the medians of the two populations are the same. Paired data samples consisted of the manual rating for

**Fig. 1.** Percentage of slabs with each type and severity of cracking.

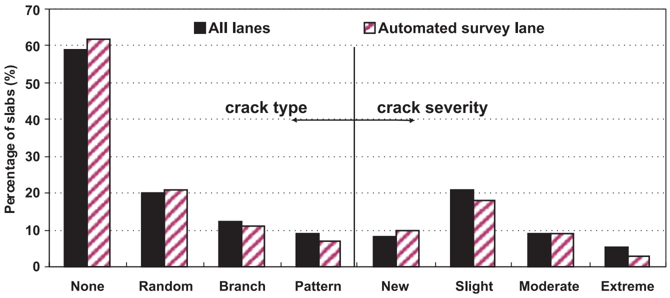
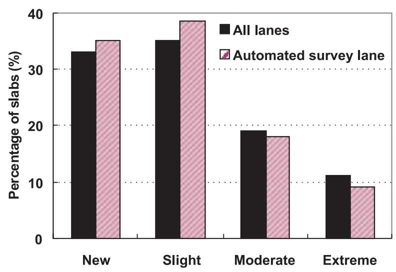



**Fig. 2.** Percentage of slabs with each severity of joint spalling.




all lanes and the manual rating for the second lane from the curb, the automated survey lane. A sample of the difference between the lanes is obtained by subtracting the percentage of each distress on the automated survey lane from the distress on all lanes. A positive difference indicates more distress on all lanes than on the second lane from the curb. If a confidence interval is entirely positive, then all of the lanes had more distress than the automated survey lane at the given confidence level. Conversely, if a confidence interval is entirely negative, then the automated survey lane had more distress than all of the lanes.

Two hypotheses were set up: the null hypothesis ($H_0$) that the two datasets have the same median amount of distress, and the alternative hypothesis ($H_a$) that the two datasets have different median amounts of distress. A binomial distribution can be used to calculate a confidence interval on the median of a population without any assumptions regarding the statistical distribution of the data (Salvia 1990). There is a 50% probability that any single randomly selected sample will fall below the median. For a sample size of 54, the probability that any number of samples lie below the population median is calculated using the binomial distribution as follows:

$$[1] \qquad f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where $f(x)$ is the probability that exactly $x$ number of elements are below the population median, $n$ is the sample size, $x$ is the number of elements below the median, and $p$ is the probability that a single element is below the median. For example, the probability that exactly 20 of the 54 samples lie below the median is calculated as

$$f(20) = \binom{54}{20} 0.5^{20}(1-0.5)^{54-20} = 0.0178$$

For a sample size of 54, there is 92.4% probability that the median of the population lies between the 21st and 34th elements. The 92.4% probability for this confidence interval is calculated as

$$[2] \qquad \text{probability} = 1 - \sum_{x=1}^{20} f(x) - \sum_{x=34}^{54} f(x)$$

Alternative confidence intervals that could have been used include the median lying between the 20th and 35th elements with 96.0% probability or the 22nd and 33rd elements with 86.6% probability.

## Influence of transverse variability on crack and joint score

The differences between the lanes on the aggregate level appear to be relatively small. Some individual segments had significant differences, however, which become evident when the rating scores are analyzed. Of the 54 randomly selected segments that were analyzed, 30 received the same crack score on all lanes as that from the manual rating of the second lane from the curb (automated survey lane). The differences in the 24 segments that had different scores are summarized in Table 3. It can be seen that the source of the difference is the dominant crack type in 12 cases, of which five segments had the automated survey lane rated as better and seven as worse than when all lanes were surveyed. Similarly, the table includes the differences in crack severity, crack extent, spall severity, and spall extent.

Whereas the conditions between adjacent lanes are the same to a statistically significant degree on an aggregate level, a detailed survey is required at the project level for all lanes. For network-level planning purposes, automated data collected for one lane are sufficient to represent the entire segment. Therefore, given that data for multiple lanes do not provide significantly more information and given the cost associated with these data, it is recommended that automated survey data only be collected for a single lane. It is further recommended that this lane be chosen randomly to eliminate the bias in the spalling scores, as spalling data are used to drive maintenance and rehabilitation programs for the APC and PCC roadways. This recommendation should apply only to multilane roadways where the traffic patterns do not differ significantly between lanes.

## Predicting manual ratings from automated surveys

Artificial neural networks (ANNs) are composed of many single units operating in parallel and are inspired by biological nervous systems. ANNs learn through a training process that utilizes training data to adjust the weights and biases of the connections between elements to achieve the desired output (Bishop 1995). Learning methods based on ANNs have been used extensively in transportation and pavement engineering applications. For example, Mohammadian and Miller (2002)

**Table 3.** Distress score differences between the automated survey lane and all lanes.

| Difference in score | Crack type | Crack severity | Crack extent | Spall severity | Spall extent |
|---|---|---|---|---|---|
| Similar condition | 12 | 10 | 11 | 5 | 4 |
| Automated survey lane was better | 5 | 6 | 8 | 7 | 8 |
| Automated survey lane was worse | 7 | 8 | 5 | 0 | 0 |
| Total | 24 | 24 | 24 | 12 | 12 |

**Note:** A total of 24 of the 54 sample segments did not have the same crack score, and 12 of the 54 sample segments did not have the same joint spalling score.

assessed the ability of ANNs to predict household vehicle choice. Fernandes et al. (2001) used neural networks for the programming of maintenance and rehabilitation of unpaved roads, and Abdallah et al. (2001) used ANNs in real time to predict pavement layer thicknesses and moduli based on deflection and seismic data.

In this paper, it is desired to use IRI and rut depth as indicators of various distresses that are collected manually in the current system. IRI and rut depth were selected because they are independent of each other, and the methods used for data collection are more reliable, more widely used, and less expensive than those required for collecting surface distresses. There are also substantial historical IRI and rut depth data in many agencies.

## Neural network architecture

Determining the neural network architecture requires significant effort, including the determination of input and output variables, number of hidden layers, and number of hidden neurons in each hidden layer. In practice, most neural network applications use only one hidden layer, and trial and error is employed to select the number of neurons for the hidden layer. Yang et al. (2003) found that processing time increases significantly when two hidden layers are used. In addition, one hidden layer is sufficient to simulate any polynomial function.

Multilayer perceptrons (MLP) were utilized to predict the classes of distress for segments based on IRI and average rut depth. The training of an MLP is usually accomplished by a back-propagation algorithm that involves two phases, the forward phase and the backward phase. During the forward phase the network weights and biases are fixed as the inputs are passed through the network and the outputs are calculated. The error is computed as the difference between the desired response and the actual response:

[3] $$e_i = d_i - y_i$$

where $e_i$ is the error for the $i$th element, $d_i$ is the desired response, and $y_i$ is the calculated network output. During the backward phase, $e_i$ is passed backward through the network and the network weights and biases are adjusted to minimize $e_i$, usually as represented by root mean square (RMS) error (Haykin 1994).

The MATLAB® (The MathWorks, Inc., Natick, Mass.) neural network processing toolbox (Demuth and Beale 2002) was used to train and test the multilayer perceptrons. The resilient back-propagation algorithm and sigmoid transfer functions are used in the selected network architecture, and both are suitable for pattern recognition problems. The initial bias and layer weights are randomly selected, which can lead to different decision boundaries each time the network is trained. To overcome this problem, the network was trained 100 times for 500 epochs each. An epoch is defined as a single iteration of the training process where the error is calculated and back-propagated to adjust the transfer functions. The error and the decision space (weights and biases of transfer functions) had stabilized after 500 epochs. Only the training run (of the 100 processed) that achieved the lowest RMS error was retained. A randomly selected 80% of the segments of each pavement type was used to train the network, and the remaining 20% of the data was used to test the effectiveness of the decision space.

The analysis uses neural networks with one hidden layer, with the number of neurons in the hidden layer selected to maximize the potential for learning while minimizing the risk of overfitting the model. A neural network with too few hidden neurons is incapable of sufficiently learning from the training set, and one with too many hidden neurons will attempt to memorize the training set and perform poorly for previously unseen data. The selection of a network with too many hidden neurons is analogous to the fitting of a polynomial of too high an order; overfitted models are unable to generalize to new data.

Two network architectures were used as shown in Fig. 3. Each neural network utilizes IRI and rut depth as inputs, five neurons in the hidden layer, and sigmoid transfer functions at each layer and is fully connected. Artificial neural network 1 (ANN1) uses four output neurons and thus an output in the form of a four-element vector. The target output vector is {1, 0, 0, 0} for new streets, {0, 1, 0, 0} for good streets, {0, 0, 1, 0} for fair streets, and {0, 0, 0, 1} for poor streets. Artificial neural network 2 (ANN2) uses a single output neuron. The potential disadvantage of this network is that there are far fewer weights to adjust between the hidden layer and the output layer; however, utilizing the ordinal nature of the data is an advantage; the classes of each category are ordered and so are the target outputs. For general condition, the classes are {new, good, fair, poor} and the targets for the classes are {0.125, 0.375, 0.625, 0.875}.

To avoid overfitting the neural network by using too many hidden neurons, a number of networks were trained, with varying numbers of hidden neurons, and the RMS error of the training and test data was calculated for each network. The results of these trials for the prediction of general condition of APC streets based on IRI and rut

**Fig. 3.** Neural network architectures: (*a*) neural network 1 (ANN1); (*b*) neural network 2 (ANN2).

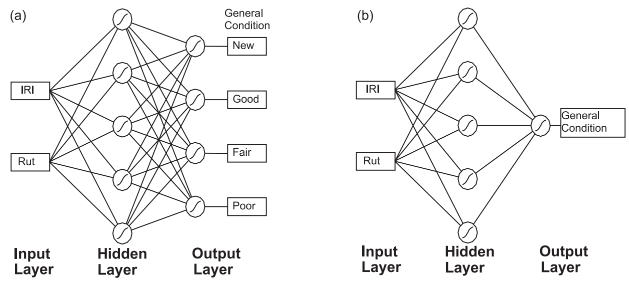



**Fig. 4.** Training and test RMS errors for the prediction of general condition of APC segments (ANN1).

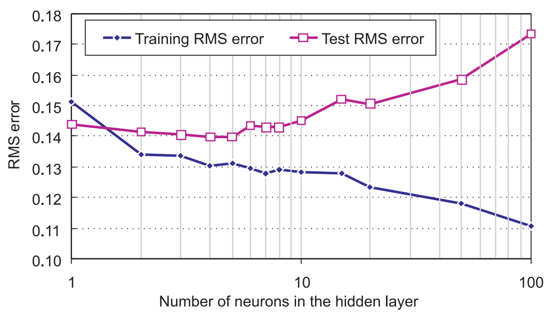



**Table 4.** Interpretation of agreement from the Kappa coefficient.

| Kappa | Strength of agreement |
|---|---|
| <0.00 | Poor |
| 0.00–0.20 | Slight |
| 0.21–0.40 | Fair |
| 0.41–0.60 | Moderate |
| 0.61–0.80 | Substantial |
| 0.81–1.00 | Almost perfect |

**Note:** From Landis and Koch (1977).

depth are shown in Fig. 4. In this figure, the error for the training set decreases with increasing numbers of hidden neurons as the network attempts to memorize the training data. The RMS error for the test data reaches a minimum at five hidden neurons, after which point the model becomes overfitted and the error for the test data increases. Subsequently, ANNs with a single hidden layer of five neurons were selected for this research.

## Strength of agreement between manual and automated ratings

The Kappa coefficient (Cohen 1960) is a method for assessing agreement that accounts for chance agreement. The Kappa coefficient was used to assess the agreement between the visual survey and the neural network output computed according to the following equation:

$$[4] \qquad \kappa = \frac{\text{observed accuracy} - \text{chance agreement}}{1 - \text{chance agreement}}$$

Kappa is interpreted according to the criteria developed by Landis and Koch (1977) and shown in Table 4. Landers et al. (2003) used the Kappa coefficient for quality assurance of manual pavement surveys by assessing the agreement between pavement raters. Goodman (2001) also used the Kappa coefficient to evaluate the variability between multiple visual pavement distress surveys.

Table 5 summarizes the neural network classification results for APC and PCC segments. For the classifications based on neural network 1, the Kappa coefficients ranged from 0.15 to 0.48 (agreement was 15%–48% better than chance), at best a moderate agreement between the ratings and the classification method. For the classification based on neural network 2, the Kappa coefficients ranged from –0.42 to 0.40. The negative Kappa coefficient represents disagreement between the two methods. A comparison of the agreement is shown in Fig. 5 for the APC segments and in Fig. 6 for the PCC segments. ANN1 performed slightly better in both cases; however, it is suggested by the authors that a substantial level of agreement (Kappa > 0.7) would be necessary to consider the replacement of the manual ratings with neural network based ratings.

The results suggest that manual and automated data are to some extent correlated but are not directly interchangeable. It is suggested that incorporating additional independent data elements, such as pavement age, material properties, and traffic, may improve the predictive power of the neural network model. Although this was not attempted in this research, it should also be noted that with the addition of new data elements the dimensionality of the model is also increased. The higher dimensionality has negative profound impacts on the model complexity and its learning ability, particularly if the variability is not explained by the data. The balance between the number of model inputs, size of the dataset, and model complexity is essential.

## Conclusions

The objective of this paper is to discuss the optimization of data collection needs by minimizing the quantity of data collected while maintaining data quality and consistency. An analysis of the transverse variability of surface distresses collected manually in 2002 for the automated survey lane and all lanes showed that a survey of one lane of a segment is representative of the entire segment to a statistically significant level. The nonparametric statistical analysis showed that the medians of the two datasets are equivalent with 92% confidence. The analysis also showed that the lane to be surveyed in future automated surveys should be randomly selected to eliminate bias in cracking and joint scores and that one lane is sufficient for network-level pavement management purposes. There was little overall difference in crack and joint scores; however, a few segments exhibited significant differences between the lanes, which affects decision making at the project level.

Artificial neural networks (ANNs) were used to predict the manual rating scores based on international roughness index (IRI) and rut depth only. A randomly selected 80% of the segments of each pavement type were used for training the ANNs. The remaining 20% of the data were used to test

**Table 5.** Neural network classification for APC and PCC segments.

| Parameter | Chance agreement (%) | Observed agreement (%) | Kappa coefficient | Interpretation of agreement |
|---|---|---|---|---|
| **APC ANN1** | | | | |
| General condition | 38 | 48 | 0.16 | Slight |
| Crack type | 40 | 59 | 0.32 | Fair |
| Crack severity | 31 | 44 | 0.18 | Slight |
| Crack extent | 55 | 76 | 0.48 | Moderate |
| Joint severity | 41 | 50 | 0.15 | Slight |
| Joint extent | 62 | 80 | 0.48 | Moderate |
| **PCC ANN1** | | | | |
| General condition | 21 | 57 | 0.45 | Moderate |
| Crack type | 39 | 64 | 0.41 | Moderate |
| Crack severity | 39 | 50 | 0.18 | Slight |
| Crack extent | 32 | 54 | 0.32 | Fair |
| Joint severity | 29 | 46 | 0.25 | Fair |
| Joint extent | 46 | 68 | 0.40 | Fair |
| **APC ANN2** | | | | |
| General condition | 38 | 59 | 0.34 | Fair |
| Crack type | 40 | 52 | 0.20 | Slight |
| Crack severity | 31 | 41 | 0.14 | Slight |
| Crack extent | 55 | 35 | –0.42 | Poor |
| Joint severity | 41 | 56 | 0.26 | Fair |
| Joint extent | 62 | 61 | –0.02 | Poor |
| **PCC ANN2** | | | | |
| General condition | 21 | 39 | 0.23 | Fair |
| Crack type | 39 | 43 | 0.06 | Slight |
| Crack severity | 39 | 50 | 0.18 | Slight |
| Crack extent | 32 | 32 | 0.00 | Poor |
| Joint severity | 29 | 57 | 0.40 | Fair |
| Joint extent | 46 | 61 | 0.27 | Fair |

**Fig. 5.** Strength of agreement for classification of APC segments based on IRI and rut depth.

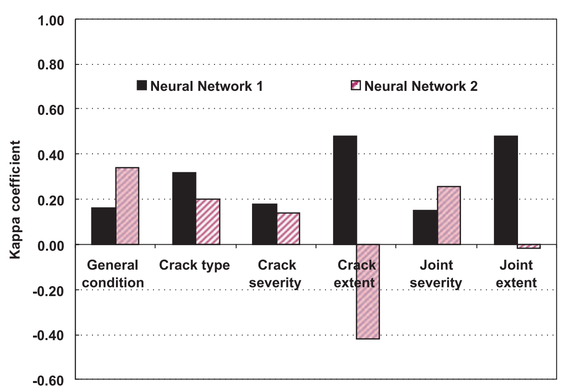
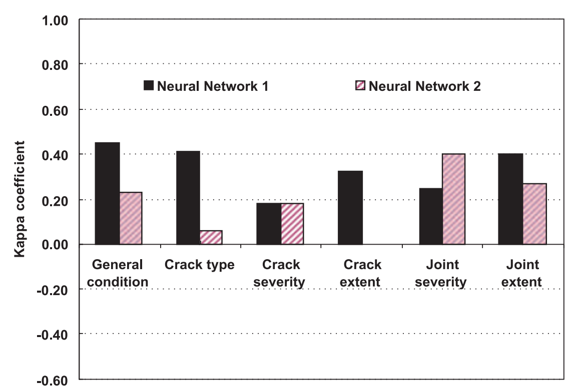



**Fig. 6.** Strength of agreement for classification of PCC segments based on IRI and rut depth.




the ability of the classifier to predict manual ratings based on IRI and rut depth. Agreement between the manual ratings and the predicted ratings was assessed with the Kappa coefficient. The Kappa coefficients showed at best moderate agreement for ANN1 and ANN2.

Although the results indicate that there exists a relationship between manual and automated ratings, ANN classification based on IRI and rut depth alone is not recommended. The classification may require the addition of other independent data elements such as pavement age, structural adequacy, or traffic indicators to improve the predictions.

## Acknowledgements

The authors gratefully acknowledge the financial and technical support of the Public Works Department, City of

Winnipeg. The authors are grateful for the support of the Natural Sciences and Engineering Research Council of Canada through a discovery research grant to the first author and a postgraduate scholarship to the second author. The Transportation Association of Canada is acknowledged for the graduate scholarship awarded to the second author.

## References

Abdallah, I., Yuan, D., and Nazarian, S. 2001. Integrating seismic and deflection methods to estimate pavement moduli. Transportation Research Record, **1755**: 43–50.

Bishop, C.M. 1995. Neural networks for pattern recognition. Oxford University Press Inc., New York.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, **20**: 37–46.

Demuth, H., and Beale, M. 2002. Neural network toolbox user's guide. The MathWorks Inc., Natick, Mass.

Fernandes, J.L., Oda, A.S., Rodgher, S.F., Fabbri, G.T.P., and Soria, M.H.A. 2001. Program of maintenance and rehabilitation activities for unpaved roads using artificial neural network. *In* Proceedings of the 29th Annual Conference of the Canadian Society for Civil Engineering, Victoria, B.C., 30 May – 2 June 2001. *Edited by* D. Noakes. [CD-ROM] Canadian Society for Civil Engineering, Montréal, Que.

Goodman, S.N. 2001. Assessing the variability of surface distress surveys in the Canadian long-term pavement performance program. Transportation Research Record, **1764**: 112–118.

Haykin, S. 1994. Neural networks: a guided tour. *In* Soft computing and intelligent systems: theory and applications. *Edited by* N.A. Sinha and M.M. Gupta. Academic Press, San Diego, Calif. pp. 71–80.

Landers, S., Bekheet, W., and Cowe-Falls, L. 2003. Cohen's weighted kappa statistic in quality control – quality assurance procedures for network-level contract pavement surface condition surveys in British Columbia, Canada. Transportation Research Record, **1860**: 103–108.

Landis, J., and Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics, **33**: 159–174.

Mohammadian, A., and Miller, E.J. 2002. Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance. Transportation Research Record, **1807**: 92–100.

Salvia, A.A. 1990. Introduction to statistics. Saunders College Publishing, Philadelphia, Pa. pp. 211–261.

Yang, J., Lu, J.J., Gunaratne, M., and Xiang, Q. 2003. Forecasting overall pavement condition with neural networks: application on Florida highway network. Transportation Research Record, **1853**: 3–12.