Expert Systems with Applications xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Modeling asphalt pavement overlay transverse cracks using the genetic operation tree and Levenberg–Marquardt Method

Machine Hsie [a,*], Yueh-Feng Ho [a], Chih-Tsang Lin [a], I-Cheng Yeh [b]

[a] Department of Civil Engineering, National Chung Hsing University, No. 250, Kuo Kuang Rd., Taichung 402, Taiwan
[b] Department of Information Management, Chung Hua University, No. 707, Sec. 2, Wufu Rd., Hsinchu 300, Taiwan

ARTICLE INFO

Keywords:
Levenberg–Marquardt
Genetic operation tree
Asphalt pavement cracking

ABSTRACT

The Artificial Neural Network (ANN) and the nonlinear regression method are commonly used to build models from experimental data. However, the ANN has been criticized for incapable of providing clear relationships and physical meanings, and is usually regarded as a black box. The nonlinear regression method needs predefined and correct formula structures to process parameter search in terms of the minimal sum of square errors. Unfortunately, the formula structures of these models are often unclear and cannot be defined in advance. To overcome these challenges, this study proposes a novel approach, called "LMGOT," that integrates two optimization techniques: the Levenberg–Marquardt (LM) Method and the genetic operation tree (GOT). The GOT borrows the concept from the genetic algorithm, a famous algorithm for solving discrete optimization problems, to generate operation trees (OTs), which represent the structures of the formulas. Meanwhile, the LM takes advantage of its merit for solving nonlinear continuous optimization problems, and determines the coefficients in the GOTs that best fit the experimental data. This paper uses the LMGOT to investigate the data sets of pavement cracks from a 15-year experiment conducted by the Texas Departments of Transportation. Results show a concise formula for predicting the length of pavement transverse cracking, and indicate that the LMGOT is an efficient approach to building an accurate crack model.

1. Introduction

Asphalt overlays have been widely used to prolong the service life of pavement, and recover both structural and functional performance of existing pavement. However, the amount of pavement cracking increases as the service time increases, allowing water to penetrate the pavement. This decreases the bond between the overlay and underlying layer in the pavement, strips asphalt mixtures, and softens the base and sub-grade layers (Chen & Hong, 2010; Hong & Chen, 2009).

Typical pavement cracks include fatigue cracking, block cracking, edge cracking, longitudinal cracking, and transverse cracking (SHRP, 1993). This study focuses on the transverse crack, which is the most common type of crack.

Four factors control the development of transverse cracking: time served, surface preparation (mill/no mill), material used (virgin/reclaimed asphalt pavement) and thickness of overlay (Hong & Chen, 2009). Designers usually count the allowed length of transverse cracks to determine the service life of pavement. Hence, building the relationship between these four factors and the length of transverse cracks is crucial.

Previous researchers built a model for pavement cracks using the Artificial Neural Network (ANN) (Lou, Gunaratne, Lu, & Dietrich, 2001). However, the ANN has been criticized for incapable of providing clear relationships and physical meanings, and is usually regarded as a black box (Baykasoglu, Oztas, & Ozbay, 2009; Chang, Hung, & Chen, 2006; Lou et al., 2001; Yeh & Lien, 2009).

Non-linear regression can also build crack models (Chen & Hong, 2010; Hong & Chen, 2009). For example, Hong and Chen (2009) used non-linear regression to find the coefficients of a preconceived sigmoid function to match the SPS-5 experimental data. In their formulas, the dependent variable Y is the total length of the transverse crack, and the independent variables x_i represent time served, surface preparation, material used, and thickness of overlay. Eq. (1) shows their formula, which has 7 coefficients ($\beta_1 \sim \beta_7$) and is quite complicated.

$$Y = \frac{\beta_1}{1 + e^{[\beta_2 + (\beta_3 + \beta_4 \times ML + \exp(\beta_5 \times RAP) \times (\beta_6 \times TH_B + \beta_7 \times TH_C)) \times T]}} \quad (1)$$

where

Y: transverse cracking length (ft or m);
ML: dummy variable for surface preparation, $ML = 1$ for mill, and 0 for no mill;

* Corresponding author. Tel.: +886 4 2285 5647.
E-mail address: mchsie@gmail.com (M. Hsie).

0957-4174/$ - see front matter
doi:10.1016/j.eswa.2011.10.005

Please cite this article in press as: Hsie, M., et al. Modeling asphalt pavement overlay transverse cracks using the genetic operation tree and Levenberg–Marquardt Method. *Expert Systems with Applications* (2011), doi:10.1016/j.eswa.2011.10.005

RAP: dummy variable for overlay material, *RAP* = 1 for reclaimed asphalt pavement (RAP), and 0 for virgin;
TH_B: variable for Type B overlay thickness (inch or mm);
TH_C: variable for Type C overlay thickness (inch or mm);
T: time served after overlay construction (year);
$\beta_1 \sim \beta_7$: non-linear regression parameters to be estimated.

Because the architecture of this formula must be presumed in advance, only the coefficients are searched based on the least mean square error. However, two questions remain to be answered: (1) Is the presumed formula structure correct? (2) What can we do if the structure of the formula is totally unknown?

In summary, the ANN is only a black box, and non-linear regression requires a correct preconceived formula structure. To build a model without knowing a clear formula structure, some researchers have used the genetic operation tree (GOT) to generate the model structure, and iteratively find the formula that best fits the experimental data (Koza, 1992; Sette & Boullart, 2001; Yeh & Lien, 2009).

The GOT is widely used in the fields of engineering, medicine, finance, and image recognition (Azamathulla, Ghani, Zakaria, & Guven, 2009; Baykasoglu et al., 2009; Chang, Chen, Chen, & Liu, 2008; Chen, 2003; Etemadi, Rostamy, & Dehkordi, 2009; Fonlupt, 2001; Potvina, Sorianoa, & Vallee, 2004; Worzel, Yu, Almal, & Chinnaiyan, 2009). For example, Baykasoglu et al. (2009) applied the GOT to high performance concrete, and determined the relationship between the concrete compressive strength and the amount of cement, fly ash, blast furnace, slag, water, super-plasticizer, coarse aggregate, and fine aggregate.

Based on the genetic algorithm (GA) approach, the GOT generates operation trees (OT) that represent the structures of the formulas. In each generation, elite genes with less error are more likely to reproduce the genes of the next generation. The traditional GOT evolution procedures also generate the coefficients (continuous values) based on the concept of GA. Unfortunately, the GOT usually has difficulty reaching convergence. This is primarily because the GA is good at organizing the structure of the formula, but is inept at searching the coefficients of continuous values. The traditional GOT process discards many good OT structures simply because their coefficients are not properly decided, which greatly increases inefficiency.

This study proposes an innovative approach, called "LMGOT," that integrates the Levenberg–Marquardt optimization technique with the GOT. The GA organizes structures of formulas represented as OTs, while the Levenberg–Marquardt Method determines the coefficients of the OTs. Results show that the LMGOT can efficiently build an accurate crack model. This study uses the LMGOT to investigate the data sets of pavement cracks from a 15-year experiment conducted by the Texas Departments of Transportation (TXDOT). The proposed approach produces a very concise formula for predicting the length of pavement transverse cracking.

2. Levenberg–Marquardt genetic operation tree

The LMGOT includes three parts: operation tree (OT), genetic algorithm (GA), and Levenberg–Marquardt optimization (LM).

2.1. Operation trees

The OT is a binary tree describing the architecture of an explicit formula. Each node on the OT has either zero or two children. A node with no children is a leaf, and must be a variable or a constant; if a node has two children, it is a mathematical operation ($+, -, \times, \div$, power, log (natural logarithm), etc.). For example, Fig. 1 shows an OT that represents the following formula (Sette & Boullart, 2001; Yeh & Lien, 2009): $x_1 \times ((x_2 - 2.1) + (x_3 \times x_4^{1.5}))$.

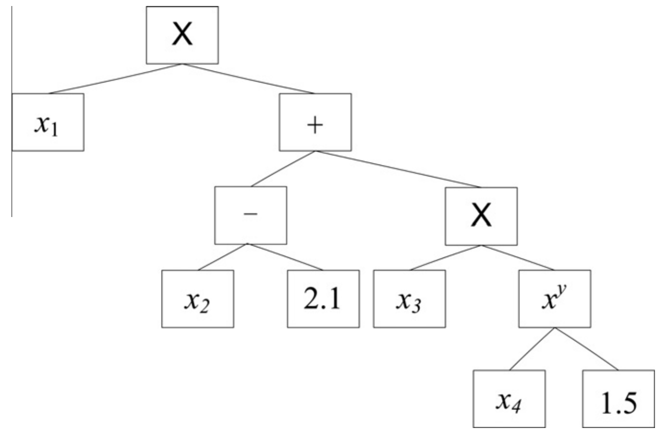

Fig. 1. An example operation tree.

Given a set of values for the variables on the OT, the operation tree (formula) can deduce a function value. With a set of data on pavement crack length, producing an OT whose formula best fits the data is an optimization problem. Determining the structures of OT is a discrete problem, while determining the coefficients of OT is a continuous problem. These are discrete and continuous hybrid optimization problems, and are therefore very difficult to solve. The GA (Davis, 1991; Goldberg, 1989) is a new optimization paradigm that mimics the natural evolution mechanism. The framework of the GA naturally corresponds to discrete optimization problems. On the other hand, Levenberg–Marquardt Method (LM) is good at solving continuous optimization problems. Therefore, this study integrates both methods to produce the OT (formula) that best fits the pavement crack length data.

2.2. Genetic algorithms

The GA generally uses five components to solve a problem (Davis, 1991; Goldberg, 1989):

(1) A genetic representation of solutions to the problem.
(2) A way to create an initial population of solutions.
(3) An evaluation function that provides their "fitness values".
(4) Genetic operations that effect the composition of children during reproduction.
(5) Parameters that the GA uses (e.g., population size, crossover rate, mutation rate).

Two basic concepts of GA are described as follows (Davis, 1991; Goldberg, 1989):

2.2.1. Reproduction

During each generation, the fitness of each solution is evaluated, and elite solutions are selected for reproduction based on their fitness. Selection embodies the principle of "survival of the fittest." Good solutions are more likely to reproduce, while bad solutions are not: the probability of selection is directly proportional to fitness. The selected solutions then undergo recombination under the action of the crossover and mutation operators (Davis, 1991; Goldberg, 1989). To improve the performance of the algorithm, the best solution of the current generation passes directly to the next generation. This strategy is also called an elitist strategy.

2.2.2. Recombination (crossover and mutation)

The crossover process combines the features of two parent structures to form two or more similar offspring. It operates by swapping segments of a string corresponding to the elite parent solutions. On the other hand, the mutation process randomly alters one or more components of a selected structure; this increases the variability of the population. Each part of each solution vector in the population undergoes a random change with a probability equal to the mutation rate.

Each execution cycle of the GA is usually called a generation. This cycle consists of three steps (Davis, 1991; Goldberg, 1989):

(1) Reproduction: From the existing population of structures, build a new mating pool of elite parents by selecting structures using the fitness evaluation function.
(2) Recombination: Apply crossover and mutation to pairs randomly selected from the mating pool to create offspring structures for the next population.
(3) Replacement: Replace the existing population with the new population.

However, the OT coefficients of the traditional GOT are very difficult to determine using the genetic operators of crossover and mutation. This is because the GA is an algorithm for solving discrete optimization problems, and is therefore relatively inefficient at optimizing continuous parameters. To overcome this drawback, the proposed approach integrates the classical Levenberg–Marquardt Method into the traditional GOT.

2.3. Levenberg–Marquardt optimization

This study proposes the Levenberg–Marquardt optimization technique to identify the coefficients of an OT, as the following example describes:

Suppose the GOT generates Eq. (2), $E(T, TH, RAP, p_1, p_2, p_3, p_4)$, which contains 4 coefficients: (p_1, p_2, p_3, p_4). Given m experimental data: $\{(y_1, T_1, TH_1, RAP_1), (y_2, T_2, TH_2, RAP_2), \ldots; (y_m, T_m, TH_m, RAP_m)\}$, try to find which of p_1, p_2, p_3, p_4 makes the $E(T, TH, RAP, p_1, p_2, p_3, p_4)$ best fit the dataset. The $f_i(p_1, p_2, p_3, p_4)$ in Eq. (3) represents the errors between the function and the experimental data. The F in Eq. (4) represents the summation of square errors. Trying to minimize the objective function F by searching the independent variables (p_1, p_2, p_3, and p_4) is an optimization problem.

$$E(T, TH, RAP, p_1, p_2, p_3, p_4) = p_1 \times \log\left[\frac{(T - p_2)}{TH^{(p_3 - RAP) \times p_4}}\right] \quad (2)$$

$$f_i(p_1, p_2, p_3, p_4) = y_i - E(T_i, TH_i, RAP_i, p_1, p_2, p_3, p_4), \quad i \in \{1, 2, 3, \ldots, m\} \quad (3)$$

$$F(p_1, p_2, p_3, p_4) = \sum_{i=0}^{m} (f_i(p_1, p_2, p_3, p_4))^2 \quad (4)$$

This study uses an iteration approach to identify the optimal solution ($\boldsymbol{P}^* = [p_1^*, p_2^*, p_3^*, p_4^*]$) starting from $\boldsymbol{P}_0 = [0,0,0,0]$ until a termination criterion is met (Madsen, Nielsen, & Tingleff, 2004).

$$F(\boldsymbol{P}_{K+1}) = F(\boldsymbol{P}_K + \boldsymbol{h}) = F(\boldsymbol{P}_K) + \boldsymbol{h}^T \times F'(\boldsymbol{P}_K) + O\|\boldsymbol{h}\|^2 \quad (5)$$

$$\boldsymbol{h} = \boldsymbol{P}_{K+1} - \boldsymbol{P}_K \quad (6)$$

Note that $\boldsymbol{h}$ is a vector representing the length of step to the next iteration. Eq. (5) is the Taylor expansion of F, and $O\|\boldsymbol{h}\|^2$ is the tail of the Taylor expansion, which is negligible here. The calculation of $\boldsymbol{h}$ could involve numerical approaches, such as: (1) Steepest Descent method, (2) Newton method, (3) Gauss–Newton method, and (4) Levenberg–Marquardt Method.

The **Steepest Descent method** appears in Eq. (7). The major drawback of this method is that the results often fluctuate near the optimal solution (Madsen et al., 2004).

$$\boldsymbol{h} = -F'(\boldsymbol{P}_K). \quad (7)$$

Eqs. (8) and (9) show the **Newton method**, which takes the derivative of Eq. (5), and omits its higher order tail. If $(\boldsymbol{P}_k + \boldsymbol{h}) = \boldsymbol{P}^*$, then $F(\boldsymbol{P}_K + \boldsymbol{h}) = 0$.

$$0 = F'(\boldsymbol{P}_K) + \boldsymbol{h}^T \times F''(\boldsymbol{P}_K) \quad (8)$$

$$\boldsymbol{h} = \frac{-F'(\boldsymbol{P}_K)}{F''(\boldsymbol{P}_K)} \quad (9)$$

The **Gauss–Newton method** improves the efficiency of the **Newton method** by Eq. (10). However, the drawback of this method is slower convergence when it is far from the optimal solution.

$$\boldsymbol{h} = \frac{-f'(\boldsymbol{P}_K)^T \times f(\boldsymbol{P}_K)}{f'(\boldsymbol{P}_K)^T \times f'(\boldsymbol{P}_K)} \quad (10)$$

To avoid the drawbacks of the Steepest Descent method and the Gauss–Newton method, this study uses an improved method called the **Levenberg–Marquardt Method**. The Levenberg–Marquardt Method works more like the **Steepest Descent** method when the parameters are far from their optimal values, yet works more like the **Gauss–Newton** method when the parameters are near their optimal values. This study applies damping parameters a damping parameters μ, shown in Eq. (11), where "I" is a unit matrix. When $\mu > 0$, the coefficient matrix will be positive definite, ensuring the $\boldsymbol{h}$ always points to a decreasing direction. When μ is relatively large, Eq. (12) is close to Eq. (7). Yet, when μ is small, Eq. (12) is close to Eq. (10) (Bazaraa, Sherali, & Shetty, 2006; Madsen et al., 2004).

$$\boldsymbol{h} = \frac{-f'(\boldsymbol{P}_K)^T \times f(\boldsymbol{P}_K)}{f'(\boldsymbol{P}_K)^T \times f'(\boldsymbol{P}_K) + \mu I} \quad (11)$$

$$\boldsymbol{h} \cong -\left(\frac{1}{\mu}\right) \times F'(\boldsymbol{P}_K) \quad (12)$$

Proper adjustments of μ will make $\boldsymbol{h}$ converge to the objective function faster. Let $S = F(\boldsymbol{P}_{k+1}) - F(\boldsymbol{P}_k)$. When $S \geqslant 0$, μ increases in ten-fold; when $S < 0$, μ decreases in ten-fold. In this study, μ has an initial value of 0.001.

3. Procedures of Levenberg–Marquardt genetic operation tree

Fig. 2 shows the procedures of the Levenberg–Marquardt genetic operation tree (LMGOT). The LMGOT is an innovative approach that integrates OT, GA, and Levenberg–Marquardt optimization. The major breakthrough of this approach is that the searching of coefficients no longer uses GA, but instead uses Levenberg–Marquardt optimization, which makes the LMGOT very efficient.

This procedure includes the following steps:

(1) Setup the parameters, such as population size, crossover and mutation rate, and operation tree depth.
(2) Initialization: Randomly generate the initial population of OTs.
(3) Optimization: Apply the Levenberg–Marquardt Method to optimize the coefficients of each OT in the generation.
(4) Evaluation: Evaluate the fitness of each OT. The better the predicted values of OT fit the actual values in the training dataset, the higher the fitness.
(5) Reproduction: Reproduce OTs based on their fitness.
(6) Recombination: Apply genetic operations (crossover and mutation) to generate the next generation of OTs.
(7) Replacement: Replace the existing population with the new population.
(8) Repeat Steps 3–7 until a stop criterion is met.
(9) Output the optimum OT with the highest fitness.

Please cite this article in press as: Hsie, M., et al. Modeling asphalt pavement overlay transverse cracks using the genetic operation tree and Levenberg–Marquardt Method. *Expert Systems with Applications* (2011), doi:10.1016/j.eswa.2011.10.005

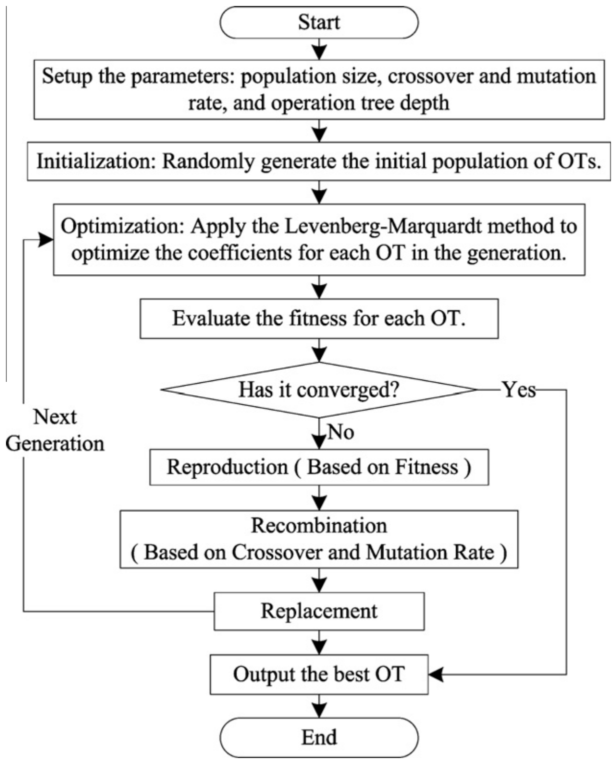

Fig. 2. The LMGOT framework.

4. Model and data analysis

This study develops a model of transverse cracks for overlay pavements using the proposed LMGOT. Data was collected from a long term experiment conducted by the Texas DOT. The following section describes this procedure:

4.1. Data collecting

In an effort to determine the performance of pavement after long-term service (the Long-Term Pavement Performance; LTPP), the United States of America (USA) began the Strategic Highway Research Program (SHRP) in 1987 to conduct ongoing experiments. The LTPP program conducts two types of pavement performance experiments: the General Pavement Study (GPS), and the Specific Pavement Study (SPS). The GPS includes 9 experimental categories (GPS-1 to GPS-9) to develop a comprehensive national database of pavement performance. The SPS includes 9 experimental categories (SPS-1 to SPS-9), but focuses on specially constructed, maintained, or rehabilitated pavement sections. This study focuses on data of the SPS-5, entitled "Rehabilitation of AC Pavements" (Hong & Chen, 2009).

This study uses crack data from asphalt overlay pavement experiments conducted by the USA Texas DOT SPS-5 A502-509 on experimental road sections. Eighty-eight sets of data were recorded, and one dataset was discarded as an outlier, leaving 87 datasets in the final sample. The data was shuffled using a random sampling, and divided into 70 training sets and 17 testing sets. Both groups were used to evaluate the fitness of each individual (OT) in the population. Specifically, the fitness of training data was used in the reproduction process, while the fitness of testing data was used to evaluate the generalization of LMGOT.

The Texas SPS-5 experiment was conducted on US highway 175 to determine the length of cracks in Asphalt Concrete (AC) overlay pavement. This experiment began in 1992 on 8 road sections. Each section consisted of AC overlay pavement measuring 500 feet (152.4 m) long and 12 feet (4 m) wide, with a 100-foot (30.3 m) transitional area. Fig. 3 shows the existing pavement structure (before overlay) and the new structure after overlay. The experiment examined three major factors, mill, material, and thickness (Hong & Chen, 2009), which are described as follows.

4.1.1. Surface preparation (Mill)

Some researchers believe that overlay performance is related to the surface preparation of the overlay. For example, removing existing distress likely contributes to better overlay performance. In this study, surface preparation means milling the existing pavement before placing the overlay. To compare the effect of milling on overlay performance, the experimental design included both milled and non-milled sections. On four of the eight sections, 2 inches (50.8 mm) of the existing surface AC was milled. Before the overlay placement, the milled thickness was replaced by the same material used in the overlay (a mill-and-fill operation). Fig. 3 shows that for the milled sections, the design thickness did not include the replaced asphalt mixture. For the non-milled sections, the overlay material was placed directly on the existing pavement.

4.1.2. Overlay thickness

As Table 1 shows, the SHRP SPS-5 experimental plan involved several overlay thicknesses, ranging between 2.2 inches and 7.1 inches, on eight road sections.

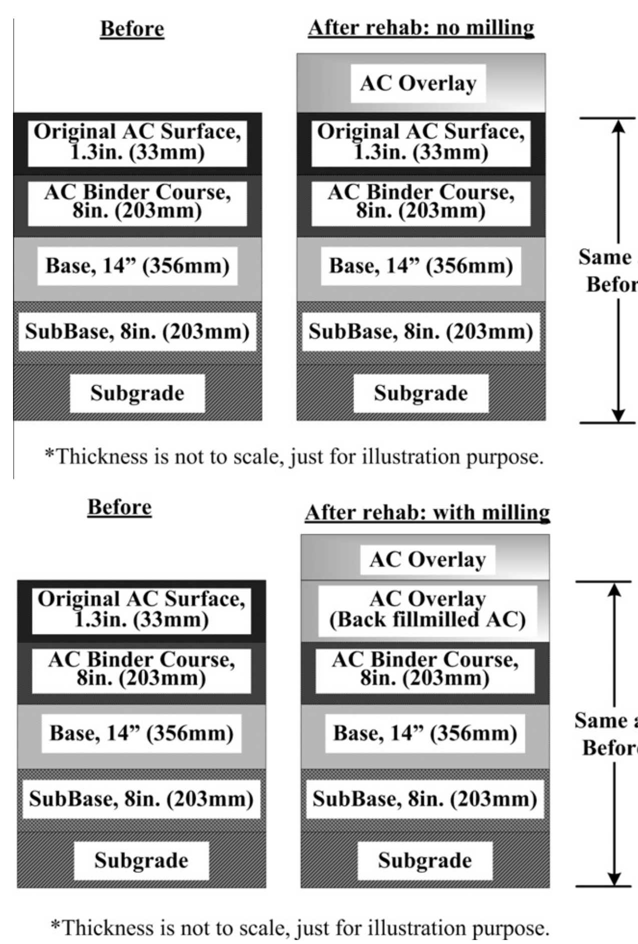

Fig. 3. Existing pavement structure before and after rehabilitation (Hong & Chen, 2009).

Table 1
Experimental design factors at Texas SPS-5 sections.

Section ID	Construction preparation	Observation time (years)	Overlay thickness (inches)	AC material
A502	No mill (0)	15	2.2	35% RAP (1)
A503	No mill (0)	15	5.1	35% RAP (1)
A504	No mill (0)	15	5.2	Virgin (0)
A505	No mill (0)	15	2.0	Virgin (0)
A506	Mill (1)	15	4.3	Virgin (0)
A507	Mill (1)	15	7.0	Virgin (0)
A508	Mill (1)	15	7.1	35% RAP (1)
A509	Mill (1)	15	4.2	35% RAP (1)

Table 2
Genetic codes of mathematical operators.

Code	1	2	3	4	5	6
Operator	+	−	×	÷	x^y	ln

Table 3
Genetic codes of variable and particular coefficients.

Code	7	8	9	10	11
Variable	*T*	*ML*	*RAP*	*TH*	*p*

4.1.3. Overlay material (RAP)

Environmental concerns and the rising cost of asphaltic materials have emphasized the necessity of exploring the use of reclaimed asphalt pavement (RAP). Foreseeing this trend, the SHRP incorporated RAP as one of the major experimental factors in the SPS-5 study. In engineering practices involving RAP, new and used asphalt concrete are mixed, in certain percentages, to produce a "combined" mixture for pavement construction. Four of the eight sections included RAP. Based on current pavement construction practice, this study adopts the maximum RAP content of 35% recycled AC. The overlay AC material in the other four sections not containing RAP is referred to as "virgin" material.

Table 1 shows the relevant data, designating the surface preparation *ML* as 0 or 1: *ML* = 0 (no mill) means directly overlaying the AC on the original pavement, while *ML* = 1 (mill) means removing the original pavement and then overlaying pavement. There are two types of material: virgin AC or recycled AC (RAP): *RAP* = 0 means virgin AC, while *RAP* = 1 means reclaimed AC.

4.2. Gene coding and operation tree rules

The OT nodes include mathematical operations, variables, and constants. Mathematical operations include +, −, ×, ÷, x^y (power), and log (natural logarithm), designated as codes 1–6 in Table 2.

This study examines the following variables and constant coefficients: time (*T*), mill/non-mill (*ML*), RAP/virgin (*RAP*), thickness (*TH*), and particular coefficients (*p*), which are coded as 7–11 and listed in Table 3.

The OT in Fig. 4 consists of 7 layers, and contains up to 127 codes. The following rules are used:

(1) The gene N_1 on the top layer is restricted to be mathematical operations. Therefore, it can be integers 1 through 6 to represent the different mathematical operations shown in Table 2.
(2) The middle layers, from 2 to 6 layer ($N_2 \sim N_{63}$), could be any mathematical operations or variables, using integers 1 through 11 to represent the different mathematical operations in Table 2 and the different variables or particular coefficients in Table 3.
(3) The genes N_{64} to N_{127} in the bottom layer are restricted to be variables or constants, using integers 7 through 11 to represent the different variables or particular coefficients in Table 3.
(4) When the node is a mathematical operator: log (natural logarithm), only the left sub-node is operated.
(5) When the nodes are variables or constants, their sub-nodes are all discarded.

5. Results

5.1. LMGOT results

The proposed LMGOT method sets the GA population size: *N* = 100, crossover rate = 0.9, mutation rate = 0.001, and the evolution generation = 1000. Using the root of mean squared error (RMSE) as the fitness values, the LMGOT iteratively searches for the OT that best fits the data. The LMGOT effectively builds a concise OT (formula) shown in Fig. 5 for the crack length of the in-service overlay pavement. Four parameters p1 to p4 shown in Fig. 5 are determined with Levenberg–Marquardt Method. Fig. 6 shows the final model for asphalt pavement overlay transverse cracks which also can be represented as a formula in Eq. (13). Eq. (13)

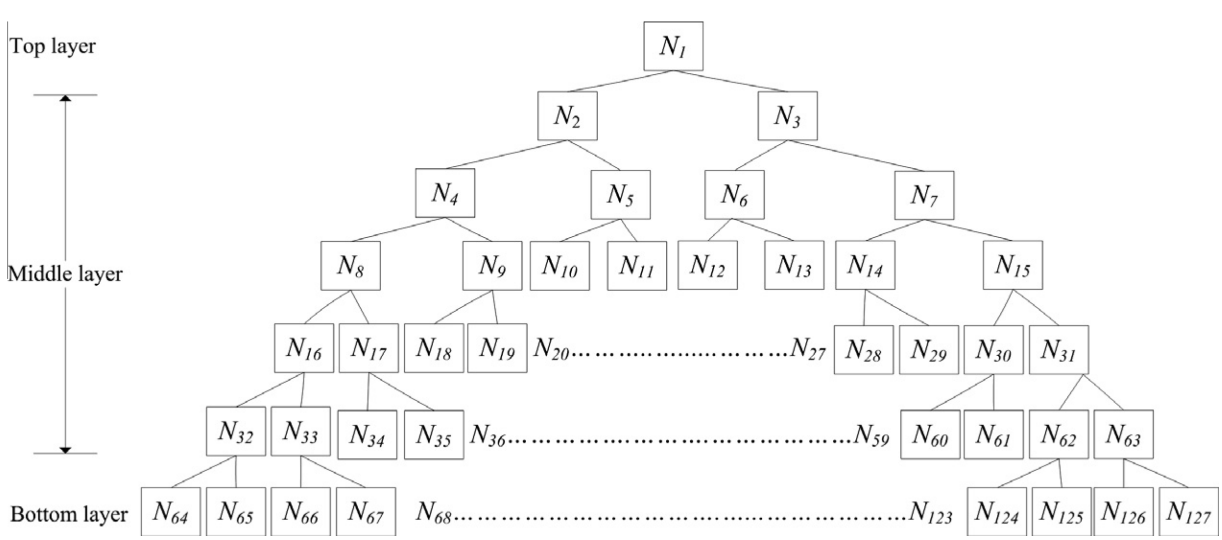

Fig. 4. A seven-layered OT model.

Please cite this article in press as: Hsie, M., et al. Modeling asphalt pavement overlay transverse cracks using the genetic operation tree and Levenberg–Marquardt Method. *Expert Systems with Applications* (2011), doi:10.1016/j.eswa.2011.10.005

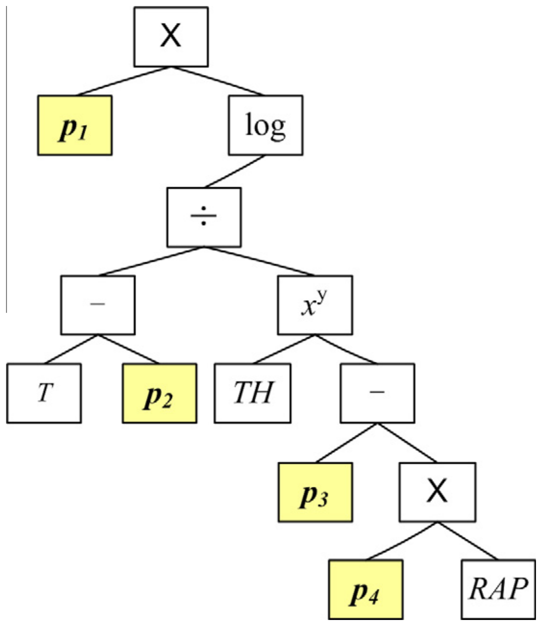

Fig. 5. The OT before applying the Levenberg–Marquardt optimization technique.

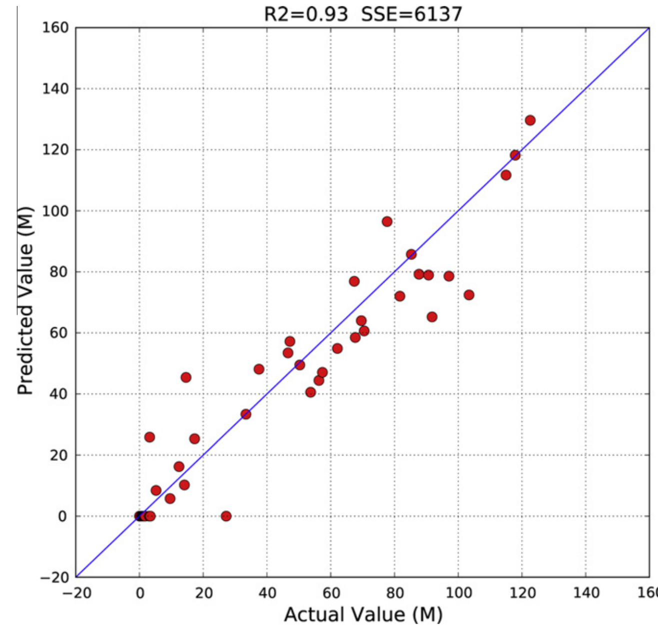

Fig. 7. Scatter chart of training data from LMGOT.

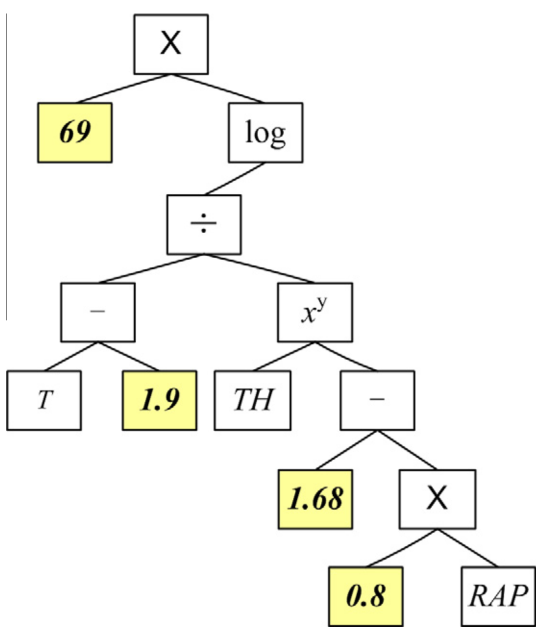

Fig. 6. The final model for asphalt pavement overlay transverse cracks.

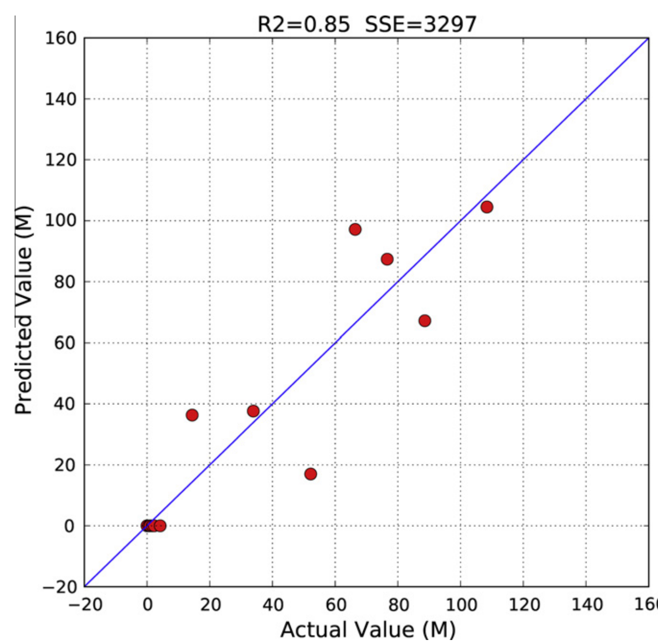

Fig. 8. Scatter chart of testing data from LMGOT.

is relatively concise compared with Eq. (1) as developed by Hong and Chen (2009). Figs. 7 and 8 show the scatter charts for training sets and testing sets, respectively.

$$Y = 69 \times \log\left[\frac{T - 1.9}{TH^{(1.68 - 0.8 \times RAP)}}\right]. \tag{13}$$

Let

$$f = \frac{T - 1.9}{TH^{(1.68 - 0.8 \times RAP)}}. \tag{14}$$

If

$$f > 1, \quad \text{then} \quad Y = 69 \times \log(f). \tag{15}$$

Else $Y = 0$, where

Y: transverse cracking length (m);
RAP: dummy variable for overlay material, $RAP = 1$ for reclaimed asphalt pavement (RAP), and 0 for virgin;
TH: overlay thickness (inch);
T: time served after overlay construction (year).

Since the length of cracks must be non-negative, the f value in Eq. (14) must be no less than 1. Thus, the formula was trimmed by adding the following condition: if $f \leqslant 1$, then set $\log(f) = 0$ to ensure that the length of cracks is non-negative.

When $RAP = 0$ indicates virgin AC, the formula can be further simplified to Eqs. (16) and (17). Meanwhile, when the $RAP = 1$ for Reclaimed AC, the formula can be simplified to Eqs. (18) and (19).

Please cite this article in press as: Hsie, M., et al. Modeling asphalt pavement overlay transverse cracks using the genetic operation tree and Levenberg–Marquardt Method. *Expert Systems with Applications* (2011), doi:10.1016/j.eswa.2011.10.005

Case 1. *RAP* = *0*

$$f = \frac{T - 1.9}{TH^{1.68}} \tag{16}$$

If

$$f > 1, \quad \text{then} \quad Y = 69 \times \log(f) \tag{17}$$

Else $Y = 0$

Case 2. *RAP* = 1

$$f = \frac{T - 1.9}{TH^{0.88}} \tag{18}$$

If

$$f > 1, \quad \text{then} \quad Y = 69 \times \log(f) \tag{19}$$

Else $Y = 0$

Since the LMGOT automatically screens out the mill/non-mill factor, the crack length is not related to surface preparation. This formula derives the following points:

(1) The crack length is directly proportional to the time in service (T), and is inversely proportional to thickness of the overlay.
(2) The virgin AC ($RAP = 0$) in Eq. (16) provides the resistance to cracking with $TH^{1.68}$, while the recycled AC ($RAP = 1$) in Eq. (18) provides the resistance to cracking with only $TH^{0.88}$. The recycled AC significantly reduces the effectiveness of thickness.
(3) The scatter chart illustrates that the coefficient of determination R^2 is 0.938, showing that this formula has high confirming with the training set. The scatter chart for the testing data, with an R^2 value of 0.85, shows that the formula still has a certain level of generalization.

5.2. Surface preparation variables comparison based on the model

Many researchers have discussed the effect of milling on AC overlay performance in terms of cracking. The proposed model automatically screens out the factor of surface preparation, implying that the 2 inch (50.8 mm) milling contributes little to resisting the transverse cracking. These results are consistent with new research results, which suggest that existing cracking can occur not only in asphalt layer, but also in base layers, sub-base layers, and sub-grades. In this experiment, the $ML = 1$ (mill) surface treatment did not completely remove the underlying cracks. Thus, transverse cracking still existed after milling, and propagated upward into the AC overlay (Hong & Chen, 2009).

5.3. Material variables comparison based on the model

Based on the overlay pavement design specification, the required service life is 8 years. The formula shows that the material

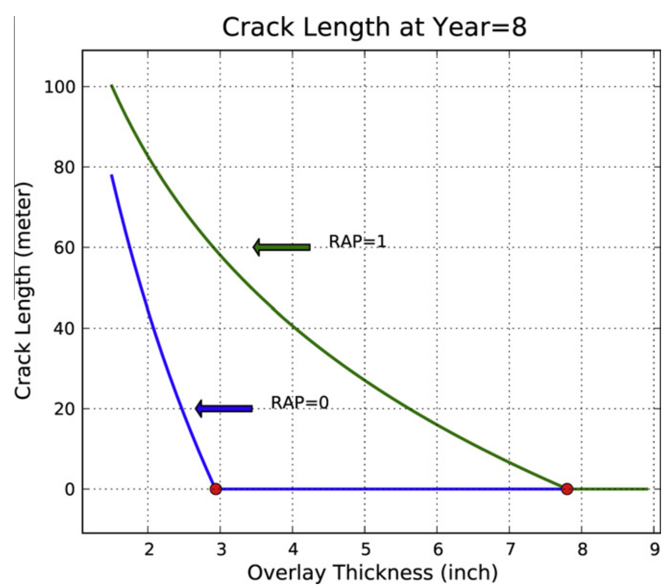

Fig. 10. The predicted pavement overlay performance focused on the required service life of 8 years.

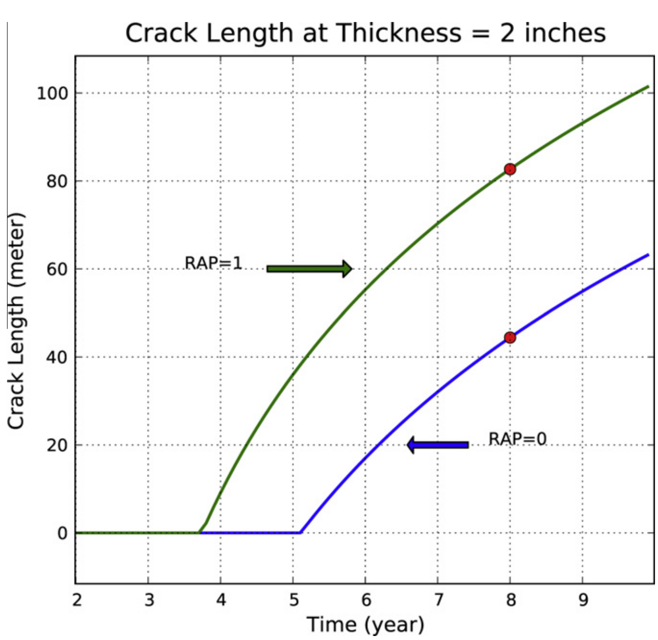

Fig. 9. The predicted pavement overlay performance focused on 2 inch thick overlay.

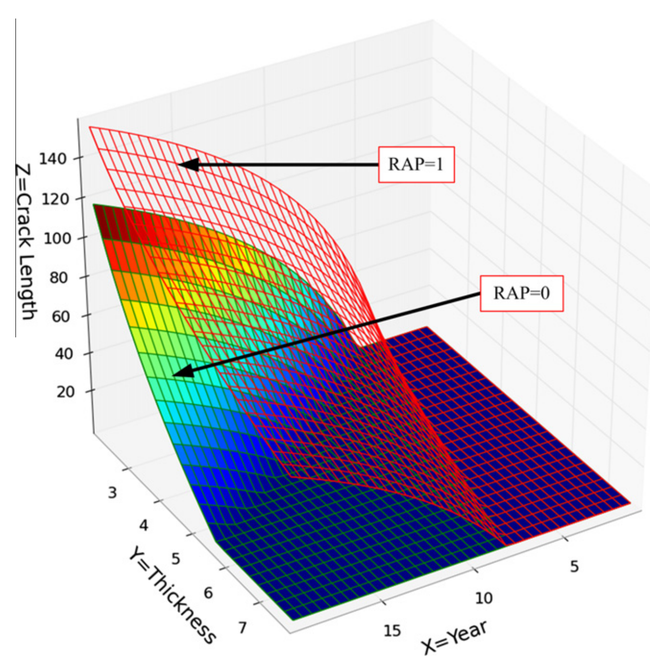

Fig. 11. The 3D mesh-plot for the length of transverse crack in RAP type and virgin type.

variable is an important factor. Fig. 9 shows that if the overlay is 2 inches thick, using Eqs. (16) and (17) for Virgin AC and Eqs. (18) and (19) for Reclaimed AC, the Virgin/ RAP tended to cause crack lengths of 44.4 and 82.7 m, respectively, at the end of required service life 8-year. The reclaimed AC accelerated the aging the pavement and caused cracks approximately twice as long as those in virgin AC. This can be explained from an asphalt material property viewpoint. The asphalt binder in RAP usually demonstrates higher viscosity due to the aging effect. Thus, the overlay with RAP is less resistant to cracking than that consisting of virgin AC.

5.4. Thickness variables comparison based on the model

Eq. (15) shows that the crack length is inversely proportional to the overlay thickness. To ensure that the overlay pavement has no significant cracks (set crack length $Y = 0$) in 8 years. Fig. 10 shows that the required thickness of pavement is 2.9 inches for Virgin AC and 7.8 inches for Reclaimed AC, respectively.

The 3D plot in Fig. 11 illustrates the inter-relations between the crack length and three factors: time, thickness, and *RAP*. The lower solid surface is for $RAP = 0$ and the upper hollow mesh is for $RAP = 1$. These results clearly illustrate that the $RAP = 0$ (virgin AC) has better resistance to cracking than the reclaimed AC.

6. Conclusions

This paper proposes an innovative approach, LMGOT that integrates the OT, GA, and Levenberg–Marquardt Method. This study applies the LMGOT to Texas DOT, SPS-5 data on transverse cracks of overlay pavements, and produces a concise and convincing formula. The following conclusions are derived:

(1) Even without predefined structures of the formulas, the LMGOT can still build a concise model for the crack length of overlay pavement.
(2) The proposed model automatically eliminates the surface preparation factor of 2 inches of milling, indicating that the LMGOT is superior at self-organizing the OT and screening out unrelated factors.
(3) The crack length formula is a log (natural logarithm) function, showing that cracks are proportional to the log of time and inversely proportional to the log of thickness. This confirms the relationship with clear physical meaning between these factors.
(4) With the same 2 inch overlay, the crack lengths of virgin AC and reclaimed AC at the end of the required service-life of 8 years predicted to be 44.4 and 82.7 m, respectively. This implies that the cost-effectiveness of reclaimed AC requires further investigation.

In summary, the LMGOT can produce a self-organized formula structure and optimize the coefficients in the formula without preconceived formula architectures. This feature is very useful when the formula structure is unknown. Because the structures of formula of many material behaviors are unknown, the LMGOT may be a better choice than the non-linear regression in modeling material behaviors. Thus, the LMGOT can help researchers develop a compact formula with physical meaning.

References

Azamathulla, H. M., Ghani, A. A., Zakaria, N. A., & Guven, A. (2009). Genetic programming to predict bridge pier scour. *Journal of Hydraulic Engineering, 136*(3), 165–169.

Baykasoglu, A., Oztas, A., & Ozbay, E. (2009). Prediction and multi-objective optimization of high-strength concrete parameters via soft computing approaches. *Expert Systems with Applications, 36*(3), 6145–6155.

Bazaraa, M. S., Sherali, H. D., & Shetty, C. M. (2006). *Nonlinear programming: Theory and algorithms*. New Jersey: John Wiley and Sons Inc..

Chang, J. R., Chen, S. H., Chen, D. H., & Liu, Y. B. (2008). Rutting prediction model developed by genetic programming method through full scale accelerated pavement testing. In *ICNC'08, Fourth International Conference on Natural Computation, Jinan, China* (Vol. 6, pp. 326–330).

Chang, J. R., Hung, C. T., & Chen, D. H. (2006). Application of an artificial neural network on depth to bedrock prediction. *International Journal of Computational Intelligence Research, 2*(1), 33–39.

Chen, L. (2003). Study of applying macroevolutionary genetic programming to concrete strength estimation. *Journal of Computing in Civil Engineering, 17*(4), 290–294.

Chen, D. H., & Hong, F. (2010). Lessons learned from rap sections with 17 years of service. *Journal of Testing and Evaluation, 38*(4), 482–493.

Davis, L. (1991). *Handbook of genetic algorithms*. New York: Van Nostrand Reinhold.

Etemadi, H., Rostamy, A. A. A., & Dehkordi, H. F. (2009). A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert Systems with Applications, 36*(2), 3199–3207.

Fonlupt, C. (2001). Solving the ocean color problem using a genetic programming approach. *Applied Soft Computing, 1*(1), 63–72.

Goldberg, D. E. (1989). *Genetic algorithms in search optimization and machine learning*. MA: Addison-Wesley Publishing Company.

Hong, F., & Chen, D. H. (2009). Effects of surface preparation, thickness, and material on asphalt pavement overlay transverse crack propagation. *Canadian Journal of Civil Engineering, 36*(9), 1411–1420.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.

Lou, Z., Gunaratne, M., Lu, J. J., & Dietrich, B. (2001). Application of neural network model to forecast short-term pavement crack condition: Florida case study. *Journal of Infrastructure Systems, 7*(4), 166–171.

Madsen, K., Nielsen, H. B., & Tingleff, O. (2004). *Methods for non-linear least squares problems*. Tech. Report, Informatics and Mathematical Modeling, Technical University of Denmark.

Potvina, J. Y., Sorianoa, P., & Vallee, M. (2004). Generating trading rules on the stock markets with genetic programming. *Computers and Operations Research, 31*(7), 1033–1047.

Sette, S., & Boullart, L. (2001). Genetic programming: Principles and applications. *Engineering Applications of Artificial Intelligence, 14*(6), 727–736.

SHRP (1993). *Distress identification manual for the long-term pavement performance project*. Strategic Highway Research Program, Project 338, Washington, DC National Research Council.

Worzel, W. P., Yu, J., Almal, A. A., & Chinnaiyan, A. M. (2009). Applications of genetic programming in cancer research. *The International Journal of Biochemistry and Cell Biology, 41*(2), 405–413.

Yeh, I.-C., & Lien, L.-C. (2009). Knowledge discovery of concrete material using genetic operation trees. *Expert Systems with Applications, 36*(3), 5807–5812.

Please cite this article in press as: Hsie, M., et al. Modeling asphalt pavement overlay transverse cracks using the genetic operation tree and Levenberg–Marquardt Method. *Expert Systems with Applications* (2011), doi:10.1016/j.eswa.2011.10.005