SECTION BASED PROBABILISTIC PERFORMANCE PREDICTION FICTION OR FUTURE?

MSc.PhD. Alfred Weninger-Vycudil
PMS-Consult
Vienna, Austria
office@pms-consult.at

MSc. Georg Samek
Samek Engineering
Langenlois, Austria
g.samek@ib-samek.at

Thomas Rohringer
PMS-Consult
Vienna, Austria
trohringer@pms-consult.at

Abstract

Life-cycle-cost-analysis is a common method to assess maintenance strategies over a given period of time. Therefore performance prediction models are essential to evaluate the future condition of different attributes (rutting, cracking, etc.). Most pavement management systems use deterministic performance prediction models, which describe the future condition by a functional correlation between the condition attribute (technical parameter or index) and the descriptive variables (age, traffic load, temperature, etc.). For the practical application these models need detailed information about its variables but do not enable a conclusion about the quality of the prediction.

To include the distribution of the descriptive variables and to get a conclusion about the quality of the performance prediction the model development must be based on a more probabilistic approach. In the context of a Masters thesis, which was carried out at the Vienna University of Technology in 2004, section based probabilistic performance prediction models using Markov-chains were tested in practice in a small community in Austria for the first time. The models were applied on different assets (carriageway, sidewalk, curbs, etc.).

The objective of this paper gives an overview of the pros and cons, mathematical problems, and how section based probabilistic performance prediction models can be applied in practice on other road networks and assets in the future.

INTRODUCTION

Life-cycle-cost-analysis (LCCA) is a common approach for the assessment of maintenance strategies and treatments in modern Pavement Management Systems (PMS). One of the key elements in the context of LCC-analysis is the prediction of future condition through the use of so called "Performance Prediction Models". These models can be applied in practice in different ways and show a very high sensitivity to the results of the analysis.

Today, most PMS use deterministic performance prediction models, which describe the future condition by a functional correlation between the condition attribute (technical parameter or index) and the descriptive variables (age, traffic load, temperature, etc.). Deterministic models are generally applied on single "homogeneous" road sections and enable an estimation of the future pavement condition and possible treatment strategies.
For the practical application these models need detailed information about its variables but do not enable a conclusion about the quality of the prediction.

The first application of probabilistic performance prediction models can be found more than 20 years ago in the context of strategic or network analysis and results are published in a high number of papers at the International Conferences on Managing Pavements. Based on the distribution of the condition across the whole road network the future condition is predicted under different variables (e.g. budget). In these early implementations of probabilistic performance prediction models it was not possible to evaluate section based treatment strategies because of missing basic information. Only a general overview of the whole network was the output of these strategic level analyses.
As condition survey became more and more accurate, a new way of section based performance prediction evolved using similar methods and processes which were applied on strategic level in the past. The main objective of a probabilistic approach in comparison to a deterministic one can be summarized as follows:

- Conclusion about the quality of the results (e.g. calculation of quality scenarios);
- Increase of accuracy in the analysis and results;
- Consideration of distribution of variables, input parameters, and basic information (e.g. quality of maintenance treatments);
- Definition of "homogeneous" sections with "representative" values becomes less important.

In general new methods must be tested in practice and evaluated afterwards. The method described in this paper shows the pros and cons, mathematical problems, and the perceived complexity of statistical modeling methods from the engineering point of view.

PERFORMANCE PREDICTION MODELS

Categorization of performance prediction models

As already mentioned in the introduction, different types of performance prediction models can be used in the context of LCCA. The following categorization is based on an engineering allocation with respect to the field of application. The categorization can be different to a more mathematical or statistical application. Regarding the development processes, the models can be categorized as follows according to Molzer et al (2000):

- Analytical models: these models are based on the theoretical calculation of stresses and strains of the pavement under loading and climatic effects in combination with the use of performance based constitutive equations (often in combination with laboratory testing);
- Empirical models: these models are based on investigations and measurements finding a correlation between the performance of the pavement attribute and different variables describing the loadings and other variables.

The complexity and the high number of input parameters required has restricted the use of analytical models primarily to project level analysis and rendered them practically unsuitable for network level application at the moment. Thus, the long-term behavior of pavement condition is generally described through the use of empirical solutions. Empirical models can be subdivided into two groups subject to the statistical method of development and application:

- Probabilistic models: these models predict the probability distribution of pavement condition taking into account the variation of data and a required categorization of sections subject to the loadings and other variables (e.g. pavement construction type);
- Deterministic models: these models try to define a functional correlation between the pavement condition (expressed by technical parameters, e.g. rutting, cracking, etc. or an index) and a certain number of descriptive variables.

In the following two chapters a short overview of these two types of performance prediction models will be given. The mathematical background for the practical application of stochastic performance prediction model using Markov chains is discussed in a separate chapter (see Mathematical Digression about Markov Chains).

Deterministic Performance Prediction

Deterministic models are the most common method of performance prediction in PMS-analysis (see Haas et al, 1994). The advantages of these models can be seen in their relative easy of application and in the clear understanding of the output by engineers. A high number of research projects and activities have been carried out to develop this type of model during the last decades in many countries, so that many models and comprehensive experience are available. Furthermore, commercial PMS software products have been developed to use primarily deterministic functions for their performance prediction (e.g. HDM 4).

As already mentioned, deterministic performance prediction models define the deterioration or change of pavement condition by using a functional correlation between the condition attribute (technical parameter or index) and one or more regressors (descriptive variables), which includes, at least, one regressor holding a time dependent parameter (age, traffic, ESALs, etc.). The deterministic performance prediction model of a condition attribute *i* on a road section *j* can be defined as follows:

$$Y \in [a,b] \rightarrow R \quad Y_{i,j}(t) = f\left(X_1(t), X_2, \ldots, X_n\right) \tag{1}$$

$Y_{i,j}$........................ value of condition attribute i on section j at time t

X_i .... time dependent regressor
X_n......................... other regressor(s)

The development of deterministic performance prediction models is basically done by applying regression analysis. Thereby the information of the investigations and measurements accumulate the basis for a potential correlation between the output variable and the regressors. The simplest form of the statistical analysis is the linear regression with one single (time dependent) regressor. In the context of the linear regression, calculus is used to minimize the distance between the data points of the investigation or measurement and the linear regression curve. Beside linear regression there are more complex approaches using other regression functions and more than one regressor.

For the section based application of these “general” functions it is necessary to adapt them to local preconditions and settings (local input parameters, curve shifting, etc.).

In Austria, three research projects were carried out to define deterministic performance prediction models for the application in the Austrian PMS. A detailed description of the methods and the models resulting can be taken from Molzer et al (2000) and Molzer et al (2002).

Probabilistic Performance Prediction

Probabilistic models are perceived to be more complex than their deterministic alternatives and require that engineers using this type of performance modeling have some basic statistical training. Due to a relative lack of research activities in the field of probabilistic performance modeling, there dose not exist the same level of experience in comparison to the deterministic approach. Furthermore the development of these models needs special statistical software products and a large sampling of available data from different measurements or inspections from different years. Thus, probabilistic performance models are rarely found to be implemented in commercial PMS-software products.
On the other hand, probabilistic models enable an assessment about the quality of performance prediction and finally of all results which are based on the models. This means that the variation of certain input information in the context of LCCA (e.g. treatment behavior) can be taken into consideration.

In comparison to deterministic models, where the future condition will be defined by an exact value, the output of a probabilistic performance prediction model is the probability distribution of the condition attribute at a certain time point. The probability distribution is usually expressed by a vector, which is based on a necessary classification of the condition attribute. The probability of a condition attribute being in a certain condition class defines each single component of the vector. The start vector is the probability distribution of the current condition.

The performance prediction is done by changing the components of the vector over time through the use of so called “transition probabilities”. These transition probabilities define the changes of the probabilities of a condition attribute being in a certain condition class during one interval (e.g. year) of the analysis.

The approach described above can be mathematically defined by a stochastic process called "Markov chain". According to Molzer et al (2000), Haas et al (1994), Butt et all (1987), and Ningyuan et al (1997) Markov chains are appropriate for the practical application of performance prediction on road pavements. An overview about Markov chains can be taken from the following mathematic digression.

MATHEMATICAL DIGRESSION ABOUT MARKOV CHAINS

A Markov chain is a particular type of stochastic process, typically evolving in time. The goal is to evaluate the probability of future events. In this process, the future of the system depends only on the present (current state) and not on the past.

Suppose we observe some characteristic of a system such as the state of a machine at some discrete points in time ($t = 0,1,2,3,\ldots$). One set of characteristics of a system is the states of the system, $S = \{s_1,s_2,...,s_r\}$. If the state of a system can change in some probabilistic fashion in time, we have a stochastic process. An important feature of a stochastic process is the way in which the past behaviour influences the future behaviour. If such a process is 'memoryless' in the sense, that only the current state matters for the future and past states are inconsequential, we say the process has the Markov property and we call it a Markov chain. The process starts in one of the states and moves from one state to another. Each move is called a step. If the chain is currently in state s_i, then it moves to state s_j,at the next step, with a probability denoted by p_{ij}. This probability is dependent only on the present state; no additional information is given by the preceding states, before the current state was reached. The changes of states are called transitions, and the probabilities p_{ij} are called transition probabilities. The system may stay in the same state, and this occurs with probability p_{ii}. Additionally we require that probabilities are the same for any t, i.e. they do not change over time. This requirement is not a requirement for a Markov chain in general, but for a special kind of Markov chain, called homogenous Markov chain. Furthermore either the initial state of the system or the probability distribution of the initial state is known.

Example. In a far-off land there is usually good weather. A sunny day is 90% likely to be followed by another sunny day. If they have a rainy day, they have an even chance of having the same the next day. Now we can form a Markov chain as follows. We take as states the kinds of weather, R(ainy) and S(unny). Transition probabilities, which are determined by the information above, can be written in matrix form:

$$P = \begin{pmatrix} & S & R \\ S & 0.9 & 0.1 \\ R & 0.5 & 0.5 \end{pmatrix} \qquad (2)$$

The entries in the first row represent the probabilities for the kinds of weather following a sunny day; in a similar manner the entries in the second row represent the probabilities for the weather following a rainy day. Such a matrix is called a transition matrix. Notice that the rows of P sum to 1. The general form of the transition matrix is given by:

$$P = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,r} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r,1} & p_{r,2} & \cdots & p_{r,r} \end{pmatrix} \tag{3}$$

The weather on day 0 is known to be sunny. This is represented by a vector in which the 'sunny' entry is 100%, and the 'rainy' entry is 0%:

$$u^{(0)} = \begin{pmatrix} 1 & 0 \end{pmatrix} \tag{4}$$

In general, a vector

$$\vec{u}^{(n)} = \begin{pmatrix} u_1^{(n)} & \dots & u_r^{(n)} \end{pmatrix} \text{ with } 0 \le u_i^{(n)} \le 1 \text{ for } i = 1, \dots, r \text{ and } \sum_{i=1}^{r} u_i^{(n)} = 1 \tag{5}$$

is called a probability vector. Its entries represent the probabilities of the states of a system after *n* steps. Thus $u^{(0)}$ gives us the starting distribution of the system.

In order to predict the weather for the next few days we can proceed as follows:

$$\vec{u}^{(1)} = \vec{u}^{(0)} \cdot P = (1 \quad 0) \cdot \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.9 & 0.1 \end{pmatrix} \tag{6}$$

Thus, there is a 90% chance that day 1 will also be sunny. The weather on day 2 can be predicted in the same way:

$$\vec{u}^{(2)} = \vec{u}^{(1)} \cdot P = \vec{u}^{(0)} \cdot P^2 = (1 \quad 0) \cdot \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}^2 = \begin{pmatrix} 0.86 & 0.14 \end{pmatrix} \tag{7}$$

Hence we can easily guess the general rule:
Let *P* be the transition matrix of a Markov chain, and let $\vec{u}^{(0)}$ be the probability vector which represents the starting distribution. Then the probability distribution after *n* steps is represented by the vector

$$\vec{u}^{(n)} = \vec{u}^{(0)} \cdot P^n \tag{8}$$

In most cases transition probabilities have to be estimated on the basis of empirical data coming from condition measurements or investigations. Three suitable methods of estimation are mentioned here:

- Non-Linear Programming: its goal is to find transition probabilities, for which the difference between the observations and the predicted values is minimized.
- Maximum Likelihood Estimation: it aims to find the transition probabilities that make the observed data most likely based on a comparison of two or more measurement campaigns at different points in time.
- Expert opinion: estimation of transition probabilities based on a statistical evaluation of expert opinions.

COMPARISON OF PRACTICAL APPROACH

Figure 1 shows schematically the procedure for the practical application of deterministic and probabilistic performance prediction models in the form of a generalized and simplified comparison. It is based on the methods in which LCCA is used in the context of PMS applications in Austria and other European countries.

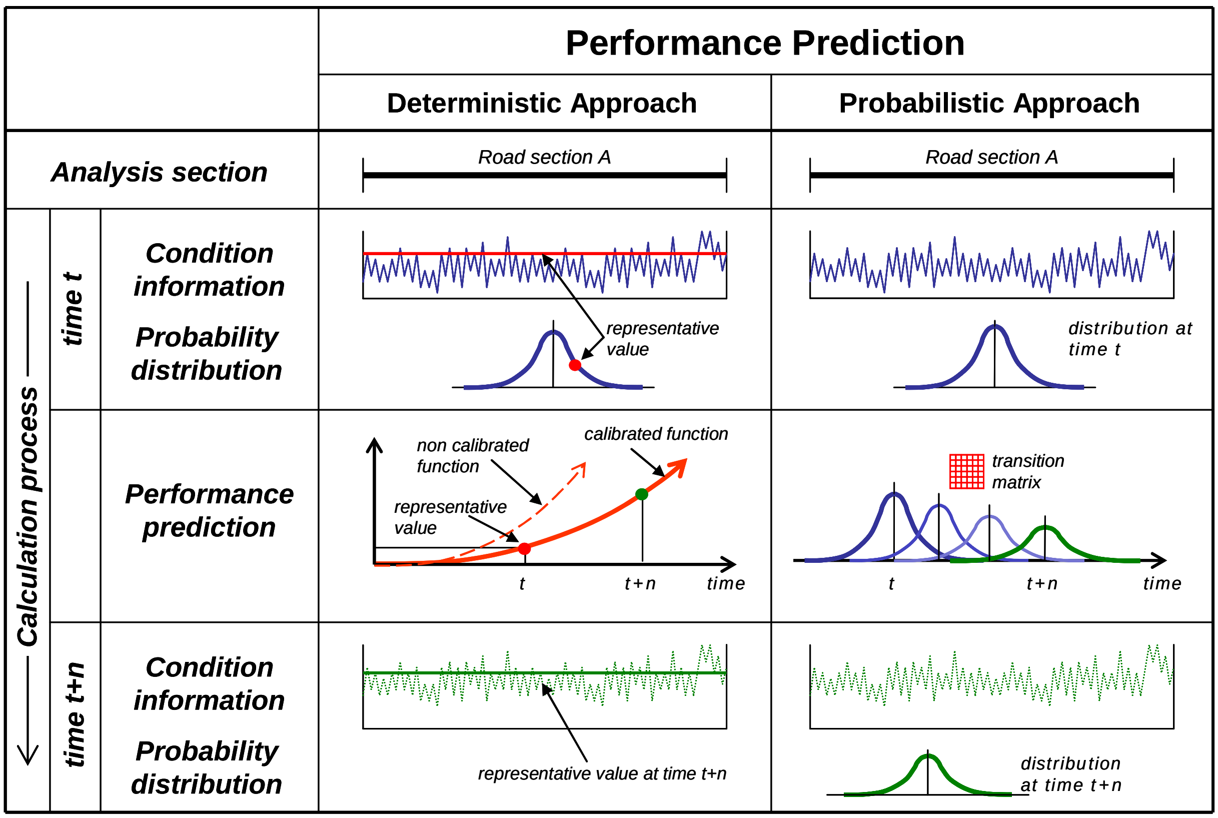

Figure 1: procedure for the application of deterministic and probabilistic performance prediction models in the context of LCCA (schematic)

The initial situation is defined by a "homogeneous" analysis section which holds a certain number of condition information (e.g. rutting) derived from measurements with high accuracy (measuring sections are much smaller than the homogeneous analysis section). Thus an empirical probability distribution of condition data is the input of both approaches.

In the context of the application of the deterministic model "representative values" will be calculated in a first step. The representative value is usually a statistical parameter (mean, median, mean plus standard deviation, etc.) which will be selected subject to the characteristic described by the condition data.
By using the representative value and other input parameters the "general" deterministic function will be calibrated to the local preconditions. According to (Weninger-Vycudil, 2003a) calibration is generally carried out in a three-stage process where stage 1 is the selection of the function, stage 2 is the application of the section based inputs (age, traffic load, etc.), and stage 3 is the

adaptation of the model to the representative value (regression analysis of time series using given function, curve shifting, etc.).
After the calibration process the performance prediction can be carried out by applying the function. At each single interval during the analysis period an exact value of the condition attribute will be available. A recharge to a probability distribution and finally to different representative values is not possible.
The condition distribution over the whole network will be expressed by the distribution of the representative value. It is different to the exact probability distribution, which is a result of the combination of the probability distribution of each single analysis section. There can be still a percentage of the network in a certain condition class although the condition distribution of the representative value shows a zero value.

Probabilistic performance prediction does not start with the calculation of a representative value because the given probability distribution is the input data of the start vector.
The section based calibration will be carried out on the one hand by the start vector itself and on the other hand by the selection of the adequate transition matrix (representative of the local preconditions, e.g. traffic, climate, pavement construction type, etc.). Through the multiplication of the transition matrix with the vector, the probability distribution will be available at each single interval.

The representative value or values (mean, median, percentile, etc.) can be selected individually by the user with respect to the desired quality and occurrence probability. It is possible to compare different treatment strategies according to a selected occurrence probability. The exact probability distribution of a condition attribute being in a certain condition class is available over the analysis period and can be used in the evaluation.

PRACTICAL APPLICATION OF PROBABILISTIC MODELING IN THE CITY OF KREMS

The first practical application of section based probabilistic performance prediction models was carried out in Austria in the context of a pilot project in the city of Krems. The results of this project were summarized in a Masters thesis (Samek, 2005) which was supported by the Institute for Road Construction and Maintenance of the Vienna University of Technology. The main objectives of the project can be summarized as follows:

- Practical application of section based probabilistic performance prediction models;
- Extension of the Austrian PMS to a community approach;
- Testing dTIMS™ (VIAPMS™) commercial PMS software for the practical application.

Beside road pavements additional assets (sidewalks, parking lanes, curbs, bicycle lanes, etc.) were taken into consideration and evaluated through the use of LCCA.

For the stochastic process simple, 5-dimensional transition matrixes were defined in cooperation with the road administration authority and other experts for all assets subject to the construction types (flexible, rigid, block, etc.), the condition attributes (structural condition index, comfort index, safety index, etc.), the loadings (traffic categories), etc. The assessment of all assets was based on the Austrian assessment scale, which shows 5 condition classes from “very good” to “very poor”, and was carried out in practice the context of a visual inspection.

The practical application of LCCA was carried out by using VIAPMS™ analytical asset management software (dTIMS™, Canadian origin), which can be adapted individually by the user to the given preconditions.
Also for this project the algorithm (system-setup) could be adapted easily to the stochastic process described above and extended to the requirements of a small community (e.g. treatment catalogue, combination procedures of condition attributes, optimization criteria, etc.). Only a higher calculation effort, which is strongly dependent on the dimension of the transition matrixes, could be detected during the analysis.

The output of the analysis can be split in results which represent on the one hand the whole network (condition distribution using representative values, treatment distribution, maintenance backlog, cost distribution, etc.) and on the other hand the situation on each single investigated road or analysis section respectively. The following Figure 2 shows the probability distribution of the Structural Condition Index for road pavements on one single section as an example for the section based probabilistic performance prediction.

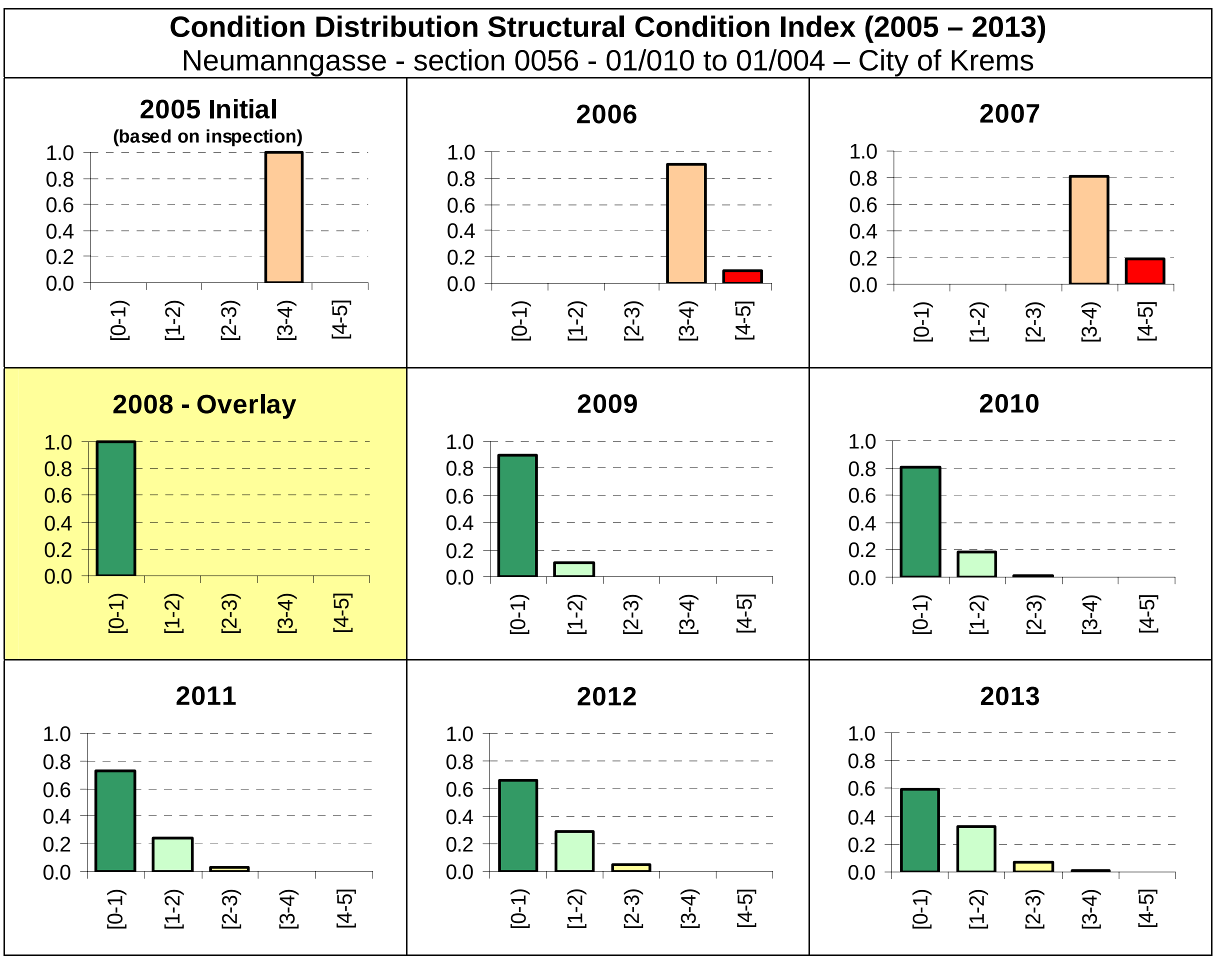

Figure 2: example of section base probabilistic performance prediction

CONCLUSION

Neither fiction nor future is the correct answer of the asked question in the title of the objective paper. The practical application of section based probabilistic performance prediction models is a fact. Especially the pilot project in the city of Krems could show that the application of a stochastic process for performance prediction provides a good basis for managing pavements and other assets. Furthermore the results of the analysis enable a much better assessment of the quality of LCCA and performance prediction in comparison to a deterministic approach.

On the other hand the probabilistic models are perceived to be more complex than their deterministic alternatives and require that engineers using this type of performance modeling have some basic statistical training. Due to a relative lack of research activities in the field of probabilistic performance modeling, new activities should take place during the next years. Primarily the Maximum Likelihood Estimation will be an easy solution for the calculation of transition probabilities if there are data of more than one investigation available.

REFERENCES

Book

Haas, R., Hudson, W.R., Zaniewski, J. (1994). Modern Pavement Management. *Krieger Publishing Company*, Florida, USA.

Proceedings

Butt, A., Shanin, M., Feighan, K., Carpenter, S. (1987) Pavement Performance Prediction Models Using the Markov Process. *Transportation Research Board, TRR 1123*, Washington D.C.

Mbwana, J., Meyburg, A. (1997). Practical Insights in Pavement Performance Modelling Using Markov Chains. *76th Annual Meeting Trasportation Research Board*, Paper No. 971300, Washington D.C.

Ningyuan, L., Haas, R. Xie, W. (1997). Investigation of the Relationship Between Deterministic and Probabalisitic Models in Pavement Management. *76th Annual Meeting, Transportation Research Board*, Paper No. 970171, Washington D.C.

Reports

Molzer, C., Felsenstein, K., Viertl, R., Litzka, J., Weninger-Vycudil, A. (2000). Statistische Methoden zur Auswertung von Strassenzustandsdaten (statistical methods for the evaluation of road condition data). *Bundesministerium für Verkehr, Innovation und Technologie*, Schriftenreihe Strassenforschung Heft 499; Vienna, Austria (in German).

Molzer, C., Felsenstein, K., Weninger-Vycudil, A., Litzka, J., Simanek, P. (2002). Statistische Auswertung der Zustandsdaten der messtechnischen Zustandserfassung 1999 und der visuellen Zustandserfassung 2000 auf dem Streckennetz der ASFINAG (statistical evaluation of condition data from 1999 and 2000 condition surveys on the ASFINAG road network). *Forschungsprojekt im Auftrag der ASFINAG*, Vienna, Austria (unpublished, in German).

PhD-thesis

Weninger-Vycudil, A (2003). Entwicklung von Systemelementen für ein oesterreichisches Pavement Management System. PhD-thesis, *Mitteilungen des Instituts für Strassenbau und Strassenerhaltung, Technische Universitaet Wien*, Heft 14; Vienna, Austria (in German).