Improving the reliability of performance modelling and harmonising specifications with the help of computational statistics

Dr Peter Kadar (ARRB Group Limited)

Asphalt properties and performance are commonly estimated by using averages of measured values. By using averages, we significantly reduce the information available and thus the reliability of the outcome. The disadvantages and consequences of forgoing the utilisation of the scatter of data may have profound consequences on all aspects of working and everyday life where we rely on data. In pavement engineering, these aspects range from meeting contractual requirements to pavement performance forecasts in pavement management systems. With the advent of cheap and fast computing, computational statistics are readily available for engineering applications. This is particularly relevant for performance forecasting – an area currently dominated by deterministic models. Yet, the future is always uncertain and uncertainty cannot be effectively addressed with the deterministic approach. The paper illustrates selected critical issues with real life examples ranging from construction practice to air travel through to laboratory testing.

In the past decades the pavement industry made large steps forward towards a better understanding of the working and failure modes of pavements. Linear and non-linear modelling of stresses and strains in the pavement has become everyday practice and it is embedded in several structural pavement design methods, including the Austroads method in use since 1989. Similarly large steps have lead to improved understanding of the behaviour of the constituent materials of asphalt mixes as well as those of the asphalt mix itself. Despite the composite and complex nature of the asphalt mixes, their behaviour and response to loading under a wide variety of conditions is now clearly understood and appreciated. Field trials and particularly the accelerated field trials have provided valuable information that allowed relating field performance to laboratory test results.

The accumulated knowledge was utilised and consolidated in models used for a wide variety of applications ranging from pavement management systems to developing specifications. The common thread between these applications was the intention to predict future performance. This intent is fully understandable and consistent with efforts made throughout mankind's history, though it challenges the conventional wisdom, of 'a well learned man knows what happened in the past, a well informed man may know what happens in the present, but no one can know what will happen in the future'. Statisticians frequently illustrate this point with the story of the black swan. Until the 17th century, the term black swan was used to describe something impossible or non-existing. This changed dramatically after the arrival of the first Europeans to Western Australia where they were confronted with the impossible and unexpected – the black swan. Ever since the warning of 'expect the unexpected' is expressed with the question: how about the black swan?

Knowing the future is presumed to hold the key to success, money and happiness among many other things. People have been looking at many possible places and means to learn about the future; gazing the stars, the winds, tea leaves just to mention a few. Engineering thinking turns towards better knowledge of the past to predict the future. Understanding the fundamental behaviour of materials is an obvious key to forecasting their future behaviour. Knowing the response to stresses and strains allows estimating future response to a range of loading. Besides the known cause-causative relationships unravelled by science, engineers also rely on historical data collected through observing pavement performance. Using this historical empirical data, many models have been developed and put to good use for product to budget development. Engineering thinking demands certainty, hence the overwhelming majority of the models is deterministic, i.e. yields one and only one answer. Yet, these models leave the question about the black swan open, usually not even contemplating the possibility of the existence of the black swan, practically denying any uncertainty. This approach starkly contrast with experience in other aspects of life – anyone exposed to the share market would confirm that past track record holds little promise let alone certainty for the future.

The earliest attempts to predict the future based on empirical observations were linked to gambling; mathematicians tried to work out the chances of winning for the most common games. Among the first to solve the riddle was Blaise Pascal in the 17^{th} century who used the term probability first (Bernstein 1996). Civil engineering practice does not accommodate words like probability lightly, though the term is a simple ratio of an event that occurs and the total number of events that may occur. When throwing a dice, it can land on only one face; yet it may land on any of the six faces. Consequently the ratio or probability of landing on any particular face is 1 in six, i.e. the probability of 0.166 or 16.66%. Pascal was also the first in history, who linked the two sides of the same issue by defining risk as the function of probability of an event occurring and the consequences of that event.

In the world of engineering we take a curiously ambiguous stance by avoiding the term probability by preferring deterministic forecasting models. Yet, design procedures usually include safety factors, whose primary role is to cater for unexpected events, such as variations in material properties and loading. For a realistic forecast we have to acknowledge uncertainties to be able to deal with them. Some of the uncertainties are known or can be anticipated, whilst others – the black swans – cannot. The recent floods in Queensland, and particularly their magnitude, could not be foreseen. Other events may be known, but some of their unintended consequences are yet to be fully understood. Such an event was the introduction of the new Austroads pavement design guide about a decade ago. The new pavement design guide acknowledges the uncertainties relating to pavements by introducing safety factors. It can be stated with a fair degree of certainty that pavements designed with the new procedure, built with contemporary technologies using new materials will perform differently – hopefully better – than earlier pavements. Yet, we calibrate our deterioration models to pavements built decades ago, designed and built according to earlier standards. So, whilst our models may accurately follow past history and extrapolate future performance, they carry an inherent uncertainty due to changes in the design and construction practices.

Uncertainty may also be introduced by not representing reality to its full extent. Typically the properties of a pavement or material are characterised by the average. The concept of average is probably the most used and abused statistical concept giving birth to the expression of 'lies, damn lies and statistics'. The proverbial man whose head is in the oven and feet are in an a bucket of icy water should be comfortable as the average temperature is the rather pleasant range. However, the extreme temperatures make his position less than pleasant. Similarly, the average rutting at network or sub-network level is comfortably low, yet we know that there are heavily rutted sections, particularly around intersections. Many specifications address this issue by taking the distribution of the results into account. In mathematical terms these requirements take the distribution of the data into account by determining a range or limit that includes 90, 95 or even 99% of all possible data points.

When average is used without describing the scatter of the data, it is very easy to misrepresent the underlying information. The NZ PSMC contracts indicated satisfactory performance when their conformance was measured in terms of average rutting and roughness (Kadar et al. 2006). However, when the scatter of the data was taken into account, it was clearly shown that whilst the average remained steady during the contract period, the condition of the bulk of the roads deteriorated. The average could be maintained by rehabilitating the worst roads at the expense of the overall condition.

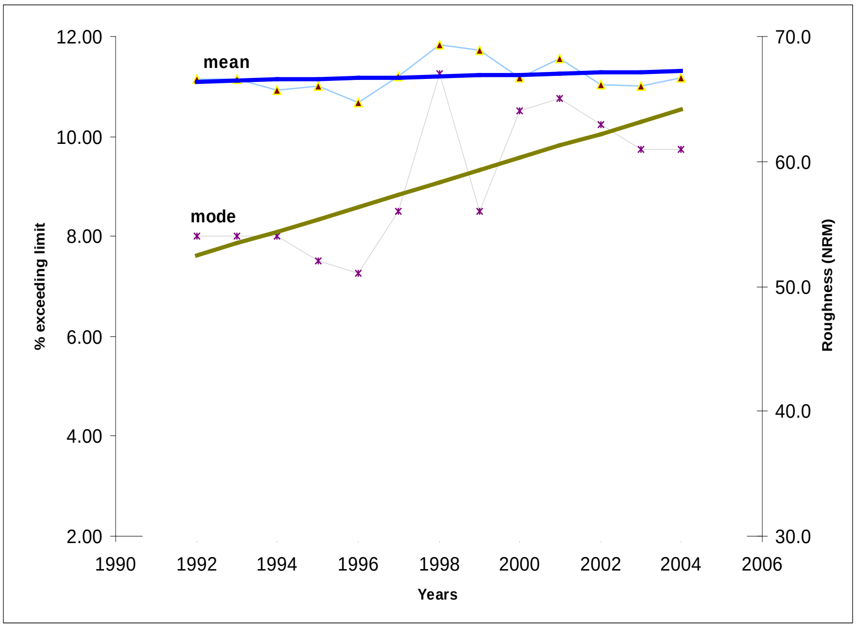

Figure 1: Mean and mode of roughness

The error committed by ignoring distribution of the data is aggravated when averages are combined in an algebraic formula. The sum of two numbers may not necessarily be the mathematical total; in other words, 2 + 2 is not necessarily 4. When the numbers represent a population distributed around the average in any manner, the sum of the averages may not be a valid representation of the whole data set. A simple example of building an asphalt pavement of two 50 mm layers illustrates this point. The layer thicknesses may vary within a range, following an even distribution, i.e. the thickness of each layer can vary between 45 and 55 mm. Whilst the sum of the averages of the two layer thicknesses may be 100 mm, any combination of the layer thicknesses within the distributions is possible. The resulting distribution of the total layer thickness is shown in Figure 2.

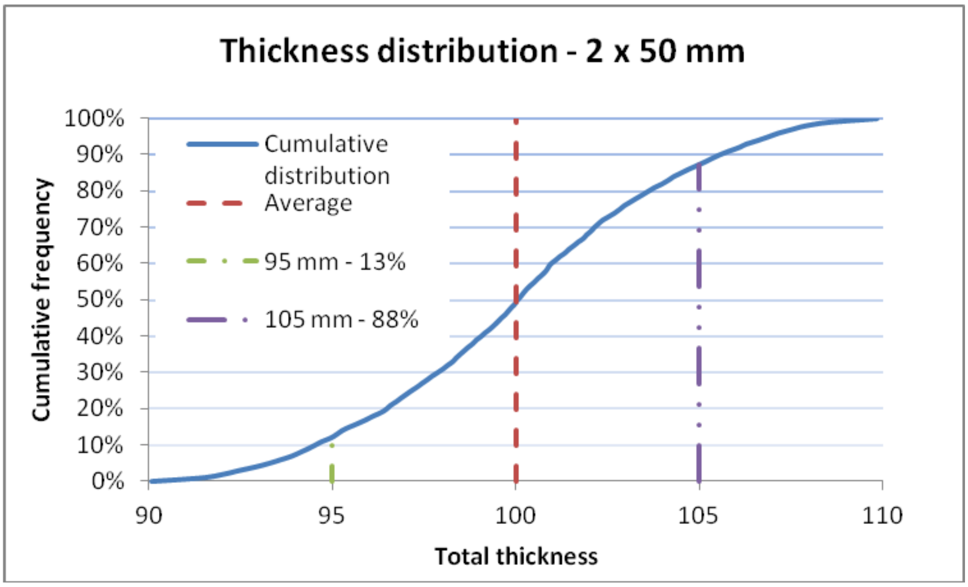

Figure 2: Thickness distribution of a 100 mm thick pavement built in 2 layers

The average may be a satisfactory 100 mm, but 13% of the pavement will have a total thickness less than 95 mm, or there is a 13% probability that the total thickness will be below 95 mm. Assuming that 95 mm thickness is the acceptable limit, the contractor might want to consider whether the company is prepared to weather the 13% probability of not meeting the contractual requirements.

Calculating the distribution of the aggregated data set involves sampling each data set. The more random samples are taken, the more accurate and refined the resulting distribution becomes; for generating Figure 2 3000 samples were used. The random sampling and the subsequent calculations can be done very quickly and efficiently by using Monte Carlo simulation.

The aggregation of the averages becomes even more critical when they are not independent from each other. This is the case for example when bitumen content, void content, compaction and VMA are specified. Assuming uniform and simple triangular distributions for the bitumen content, void content and compaction, the modulus estimated (Austroads 1992) will have an average of 3600 MPa, but even when keeping the input parameters within the specification limits, the modulus will vary between 2600 and 5000 MPa, with about 14% below 3100 MPa and 17% above 4100 MPa. This spread of modulus is predictable by using appropriate simple statistical procedures and by taking this spread into account the pavement's life can be estimated with increased reliability. It also may explain premature pavement failures as well as outstanding pavement performance on the same road.

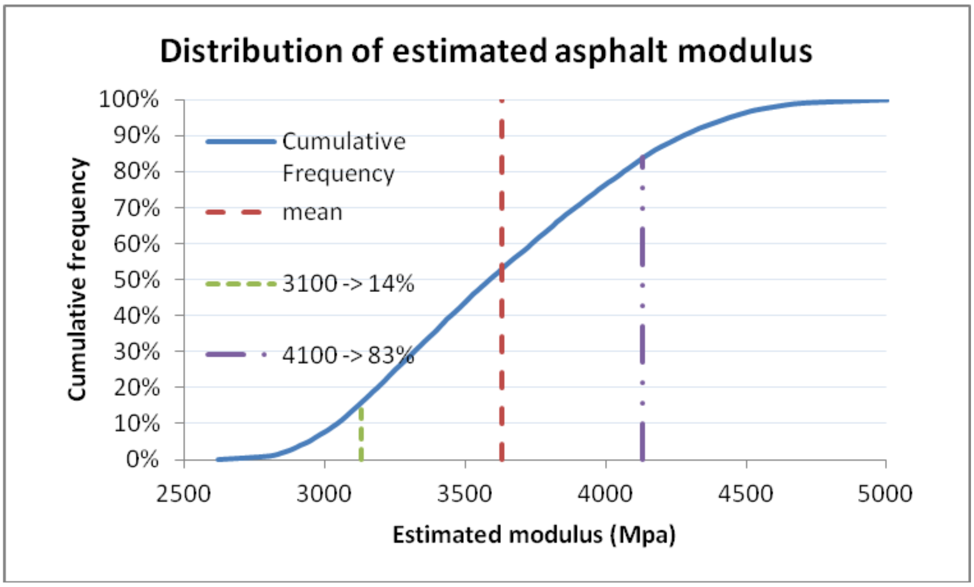

Figure 3: Distribution of estimated asphalt modulus

The examples so far assumed that the scatter of the data is known. This is not always the case, so alternative methods may need to be used for forecasting. The Markov Chain is such a tool that can be used for forecasting when very little or no historical data is available but sufficient empirical, qualitative experience is at hand. This method can be used to forecast both the extent and severity of a property. The broad applicability of the method is highlighted by the fact that it has been used to predict both deterioration of skid resistance and the changing of the condition of paint on the Sydney Harbour Bridge (Kadar et al. 2011). The 3200 steel components of the Sydney Harbour Bridge are regularly surveyed and their condition state and their extent are recorded. Four condition states were identified, ranging from condition 1 (no blemishes on the paint) to condition 4, in which the element needs renewal. Decisions about treatments are made on the basis on the size of the paint job and the condition, i.e. on the extent and condition together. (Figure 4)

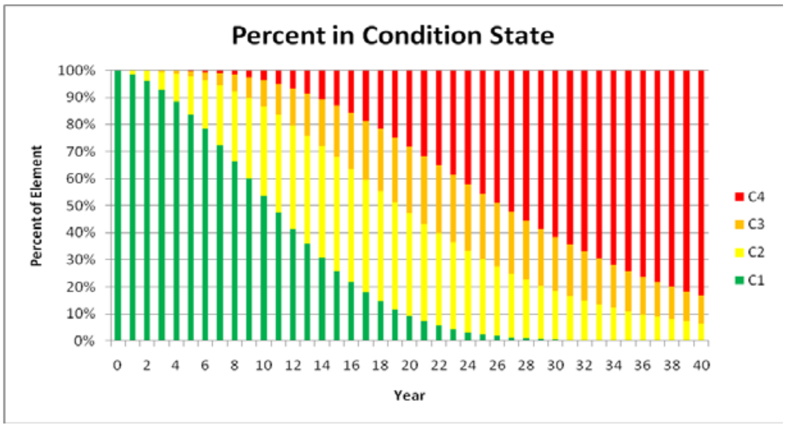

Figure 4: Condition states of the paint on the Sydney Harbour Bridge

The applied method is not entirely new – Markov Chain based solutions were used in several pavement management systems, among others in the RTA's FNOS system. The Markov Chain model has generally been replaced by deterministic models that have more explanatory power, but are less capable of handling uncertainties.

When using data for any purpose, the error inherent in any test method needs also to be considered. This issue becomes particularly critical, when the test results are used for making decisions. In real life there is rarely if at all 100% surety or 0% uncertainty; errors are made despite the best intentions. These errors can be both positive and negative, i.e. something that ought to be accepted is rejected, and something that ought to be rejected is accepted. The false positive – something rejected despite meeting all requirements - is the more critical one, as illustrated by the following example.

Assuming a test procedure that has a reliability of 90%, in 9 out of 10 cases the correct judgement will be made, but in 1 out of 10 cases the decision will be wrong, due to the limited (90%) reliability of the procedure. The consequences can be devastating as anyone travelling by air recently might have experienced. Assuming 10,000,000 passengers checked with a method having a reliability of 99%, and having only say 1000 undesirable travellers, the method will correctly identify 990 undesirables and will let 10 of them (i.e. 1%) slip through. At the same time, the method will identify 1% of all travellers incorrectly as undesirable – i.e. close to 100,000 travellers will be inconvenienced due to the false positive, as the procedure is having only 99% reliability (Savage 2009). The power of the false negative may affect everyone – not only in the asphalt industry but in our everyday life.

Summary and conclusions

Human thinking desires certainty and predictability and so does the engineering mentality. This desire is manifested in the clear preference of deterministic procedures, i.e. procedures that yield a single result without specifying its reliability or the range of scatter. This attitude leaves us defenceless when encountering a black swan – an event unexpected and considered impossible. The proverbial note that all models are wrong but some of them are useful underlines that if we want to have a realistic peek into the future, we must consider uncertainty. Uncertainty belongs to the realm of probabilistic modelling and cannot be efficiently addressed with deterministic models.

The examples used in the paper illustrate the everyday use of statistical methods. All examples in the paper utilise computational statistics as opposed to steam age statistics, i.e. from a time when it was necessary to describe distributions and other parameters with a closed – and usually complex – mathematical formula. The advent of fast and cheap computing made possible the simulation of distributions without actually defining them with a function. Computational statistics removes the complexity that seems to limit the use of statistical and probability calculations. It was illustrated that average may turn into a convenient untruth when the scatter of the data is ignored. The examples also illustrated that computational statistics can be used from contracting to refining specifications to the benefit of all industry stakeholders. It is strongly recommended to introduce computational statistical methods into engineering processes.

References

Guide to the Structural Design of Road Pavements (Austroads 1992)

Bernstein, Peter L., Against the Gods (John Wiley & Sons, 1996)

Kadar, P, Henning, T, Thew, C & Parkman, C 2006, 'Effectiveness of performance indicators in managing road networks', Transit New Zealand and New Zealand Institute of Highway Technology

Annual Conference, 8th, 2006, Auckland, New Zealand, New Zealand Institute of Highway Technology, Wellington, New Zealand,

Kadar, P., Zavitski, J., Manamperi, P. B., Mann, P., Optimised Maintenance Program for The Sydney Harbour Bridge; International Works Engineers Conference, Canberra 2011

Savage, Sam L., The Flaw of Averages (John Wiley and Sons, 2009)