Reliability in Pavement Design: Issues, Concepts, and Significance

OLGA J. PENDLETON

Reliability as both a concept and a methodology has been an integral part of scientific research for centuries. As a methodology, however, it has not been widely embraced by the pavement design research community. A historical summary of the evolution of reliability methods in other disciplines, specifically medicine, is provided. Several issues relating reliability to pavement design and evaluation are addressed. Applications and misuses of reliability methods as related to pavement performance modeling are then presented. The object is to encourage the appropriate use of reliability methods in building, evaluating, and validating pavement performance models, thus dissipating the clouds of skepticism and distrust that have surrounded this well-founded and powerful methodology.

Reliability. The word itself conjures thoughts of dependability, trustworthiness, and credibility. But when preceded by the adjective *statistical*, reactions may vary from skepticism to fear. Yet statistical reliability is nothing more than a measurement device, a yardstick, by which one can scientifically access a process in an objective, unbiased manner. So why the mixed reviews? In this paper several issues relating reliability to pavement design and evaluation are addressed in the hopes of clarifying misconceptions and dissipating the clouds of skepticism and distrust that have surrounded this well-founded and powerful methodology.

When searching for a definition of statistical or mathematical reliability the most common definitions are

- A methodology concerned with random occurrences of undesirable events (*1*).
- The probability that a system, when operating under stated environmental conditions, will perform its intended function (*1*).
- The study of the proper functioning of equipment and systems (*2*).
- The study of a random variable that represents the lifetime or time to failure of a unit (*3*).
- The probability that a unit survives until a fixed time (*4*).

So which is it: a study or a probability? Mathematically, it is a probability, namely

$$1 - P\text{ [failure]}$$

or, simply put, the probability that something will *not* fail. Specifically, in the area of pavement performance, AASHTO defines it as

> the probability that the pavement system will perform its intended function over its design life and under the conditions (or environment) encountered during operation. . . . the probability that any particular type of distress will remain below or within a permissible level . . . during the design life. (*5*)

Yet in reading textbooks on the subject, one finds reliability methods covering everything from survival analysis to least-squares regression modeling, from hypothesis testing to the formulation of prediction intervals, from normal probability distributions to Weibulls. In other words the term *reliability*, depending on the setting, may refer to something far more general than a single probability. The origin of this more global interpretation of the term probably stems from the grammatical definition of reliability.

Webster's defines "reliability" as "the state or quality of being dependable, sound, irrefutable, unquestionable, conclusive, incontestable, and infallible." The thesaurus provides antonyms such as "undependable" and even "dangerous". Probability theory, statistical inference, and the scientific method are all synonymous with methodologies that attempt to arrive at conclusions that are reliable. So it is quite logical that the term "reliability" would take on a broader meaning than merely a probability of not failing. Regardless of personal preferences as to the meaning of the term, it is important that the more general context of the word be recognized to understand the vast array of methodologies available to achieve reliability in scientific experimentation.

Historically, reliability has always been met with some skepticism. Early in the developmental history of the scientific method the recognition of variability within populations being studied brought out natural questions about the mechanisms causing the variability. These mechanisms were, unavoidably, the chance mechanisms that are the subject of probability theory. One of the oldest disciplines to first embrace the union of probability concepts with scientific experimentation was medicine. However, when Francis Galton and Karl Pearson, founders of the science of biometry, first proposed the concept at the turn of the century, the Royal Society of London (the British Academy of Science) strongly opposed it. How incredulous to recommend the mixing of mathematics and biology!

Although biometry is now a well-ingrained part of the medical field, it would appear that other sciences are still at the turn of the century with regard to acknowledging the importance of probability concepts in drawing scientific conclusions. The field of pavement design and evaluation would appear to be among these sciences. There appears to exist some reluctance toward accepting and implementing reliability theory into the modeling process. This is most unfortunate because not only has much of the difficult theory been developed but computer software is also readily available for straightforward implementation and interpretation. Many of the biomedical computer packages could be applied to the pavement area, avoiding the reinvention of the wheel, so to speak.

Texas Transportation Institute, TTI/CE Building, Suite 508, College Station, Tex. 77843.

In 1973 several scientists recognized this link between reliability and biometry and held a symposium from which proceedings were published (*6*). In reply to the question "What do reliability and biometry have in common?", the proceedings' preface states:

> Both disciplines apply statistical methods to predict, estimate, and hopefully extend lifelength. In reliability, the lifelength under study is generally that of an engineering system; in biometry, the lifelength is that of a living creature. However, aside from the presence or absence of animation . . . , the two disciplines are remarkably alike in their main goals and the tools that can be used for reaching these goals. And yet, in looking back over the history of the two subjects, one is struck by the fact that the development of the two fields has proceeded largely independently by separate groups (*6*).

Although this was acknowledged nearly 20 years ago, it would appear that the two disciplines have not made much progress toward recognizing these similarities and taking advantage of the "tools" developed by the other.

Many of the difficult problems that have resulted in complex theoretical developments are held in common by both groups. One such example is the problem of censoring. In the pavement area one is faced with the dilemma of what to do about pavements that had not yet failed up to some point in time but then were changed in some way that affected their failure time. In medical clinical trials, some patients are lost to follow-up either because they did not properly follow their treatment regimen, dropped out of the study, or died as a result of something unrelated to the disease, like a car accident. Yet there is still some valuable information about these patients in that their condition is known up to the time that they were dropped from the trial. Similarly, there is valid information in knowing that the changed pavements had not failed up to the time that the change was made. These data are generally termed "censored data," and methods exist for incorporating the valid information about these data without classifying them as "failed."

Similarly, in medicine and pavement evaluation there is the situation of "complex repairable systems." In pavements the "repair" may consist of maintenance procedures performed over the life of the pavement. In medicine the "repair" may be biological, as in the case of infant mortality as a result of some disease. As the infant matures the body internally "repairs" itself through the immune system. This is a factor that cannot be controlled for, but it can be modeled as a nonhomogeneous Poisson process. This same model would appear to be relevant to modeling pavement failure rates, including the maintenance process.

Many of the problems addressed in biometry are difficult because they cannot be controlled through experimental design. This is also true of pavement evaluation. The point is that many of these problems have been resolved in biometry and could easily be applied to pavements. The time has come for the pavement design discipline to take advantage of these developments and emerge from the cloud of skepticism and fear. It is hoped that this paper will result in a significant first step toward this end.

RELIABILITY MODELS

Reliability models are built on the assumption that there are three inherent sources of variability in the model:

- Variability in the output variables,
- Variability in the explanatory or input variables, and
- Sampling variability or experimental error.

Figure 1 represents a schematic depicting this modeling process. The input portion of the model may consist of known equations and relationships such as physical or mechanistic laws of nature. Or the input model may be a known or hypothesized probability model. It may even, in fact, be a black box in which nothing is known about either the mathematical form of the functions or even the input variables that belong in the box. But regardless of what is or is not known about the input box, a reliability model acknowledges that there is always an error associated with any experimental modeling process. Of course, the objective is to determine what form of the input box will minimize the amount of error in the model and thus produce reliable, predictable outputs.

As a simplistic example consider the relationship between the amount of stress required to deflect a beam to some degree and the physical and material properties of the beam. The exact relationship is known by using the physical and mechanistic laws of nature. Yet for a sample of beams there may be some variation from this known relationship because of differences in environmental conditions at the time of experimentation (humidity, temperature), physical properties (manufacturing differences in production of the beams' dimensions), or material properties. In this example many of the factors contributing to this variability could be controlled, especially in a laboratory setting. However, now suppose that these samples of beams have been buried underground for different periods of time and at different locations. Now the environmental factors cannot be controlled. It may be that data can be collected and these factors recorded retrospectively, in which case they go into the input box and are taken out of the error box. Suppose also that because of their submersion in the earth their physical and material properties have changed from what they were initially. A sample might have to be drawn and the extent of change in properties might need to be measured and put into the input box, again reducing the magnitude of the error box, or the uncertainty. Now, suppose this example is not about a beam at all but a section of pavement.

This, in a nutshell, is the objective of reliability modeling: to reduce the magnitude of the error box and, in so doing, to explain the dynamics of the input box. Reliability theory and methodology help to get a handle on the uncertainty or error in the models in a scientific, objective, and unbiased way. And an essential part of doing this involves understanding variability and its sources.

The importance of variability and covariability in describing pavement performance is essential and critical in designing pavements. Reliability methods provide a means of incorporating variability and covariability in developing pavement designs. More important, these methods provide an unbiased and scientifically accepted procedure for developing deformation models at acceptable confidence levels.

Some reliability models differ from other types of statistical models (i.e., regression models) in that they incorporate variability inherent in the input as well as the output variables. This is done

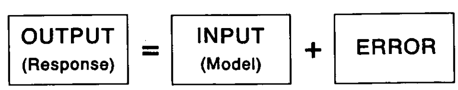

FIGURE 1 Schematic depicting reliability modeling.

through the assumption that these input variables are themselves random rather than fixed quantities and as such follow some probability distribution defined by some shape and scale parameters. The most common such distributions are the normal, lognormal, Gumbel, and Weibull distributions. The methodology exists for handling any of these four models, and tests exist for determining which model best describes the system on the basis of a sample of representative data.

Reliability theory plays a critical role in any scientific experiment from beginning to end or modification. In the beginning reliability in the design of the experiment is critical if the end product is to be optimal. The use of the concepts of probability and variability helped to establish minimum sample size requirements. Controlling the levels of certain key factors can contribute to reduced experimental variability. In addition experimental design conditions can sometimes reduce the number of samples needed by optimally sampling at the extremes or where the variability is greatest. This is design reliability.

During the course of the experiment or study reliability in management and testing (estimation) and model building are important. And, finally, reliability concepts provide a means of assessing the goodness and reproducibility of the model (reliability demonstration), which in turn allow modification of the model. Once modified, it may be desirable to redesign the experiment and begin the process again until the optimal model is obtained. So reliability plays a dynamic, not stochastic, role in any experimental process. Like the testing of a product, such as a motor vehicle, which consists of many component processing steps, the entire system, not just the end product, must be tested. So, too, must reliability methodology be used at every juncture of a scientific experiment.

COMMON ERRORS IN RELIABILITY MODELING

In this section some common errors and misapplications of the concepts of reliability and probability in engineering are illustrated. One such misconception is that a level of reliability can be generated from plotted data. The appropriate probability distribution can be found and statistically tested for goodness of fit by using the data, but ultimately it is this distribution that theoretically determines the level of any degree of reliability.

Another misconception is that model inputs are known and without error. These inputs or model components are random variables that follow some probability distribution and that have some variability, and, subsequently, coefficients of variations, that must be accounted for in the development of reliable designs and models for pavement performance. These are not fictitious or unrealistic concepts and can easily be applied to pavement performance modeling.

Specifically, the following example is presented. It shows that the variance of the log of traffic (N) is reduced by the inclusion of the (positive) correlation between material fatigue parameters. Since N is a function of these parameters, if they are treated as random variables with some bivariate probability distribution, then N, and hence the variability of N, must depend on their mean values as well as their variability and correlation. Likewise, a negative correlation among the input variables will result in a larger estimate of variability, and exclusion of this correlation produces an underestimate of the true variability. Another example of this can be seen in modeling the relationship of stress and strain to estimate expected pavement life.

Another error that is often committed is the use of the variance of means in place of the variance of individual observations. The central limit theorem states that, regardless of the distribution of a random variable, a sample of means of the variable will be normally distributed and the variability of these means is less than the variability of the random variable by a factor of $1/n$ because n, the sample size, tends to infinity. Thus, the use of means in place of individual observations is wrong and assumes a much smaller variance than what is really in effect.

Still another area of concern is in the description of the relationship between cracking and load applications. Here it is essential that the variability inherent in both of these random variables be considered. In the assumed relationship of Figure 2, for example, had a least-squares solution been used to develop this relationship using this "cloud of points," the resulting model would have assumed that the explanatory variable N was a fixed variable with no error. This would only be true if somehow in measuring cracking (C) one could specify various load applications, that is, go out and find roads that had N_1, N_2, etc., traffic loads and measure the cracking on those roads. In practice, one cannot do this. A sample of roads is selected and then whatever values of N and C result are taken. In this way, N must be treated just like C, that is, as a random variable that had some probability of occurrence in the sample and that is variable; that is, if another experiment was performed (drew another sample), there is a good likelihood that, for the same cracking level, one would observe a different load application from the previous one. Similarly, on repeated sampling for the same value of load application one would in all likelihood observe a different degree of cracking. This is what the "normal" distributions mean in Figure 2.

Another disconcerting fact is that if one were to reverse the roles of cracking and load application—that is, put load application on the y-axis and cracking on the x-axis and fit a least-squares regression model—the model and reliability levels would be different for the reversed case. The reaon for this is that cracking has now been assumed to be fixed without error and only the variability in load application is accounted for. The only way that the same model would result regardless of the roles of these variables in the least-squares modeling procedure is the very restricted case in which both random variables have exactly the same variabilities.

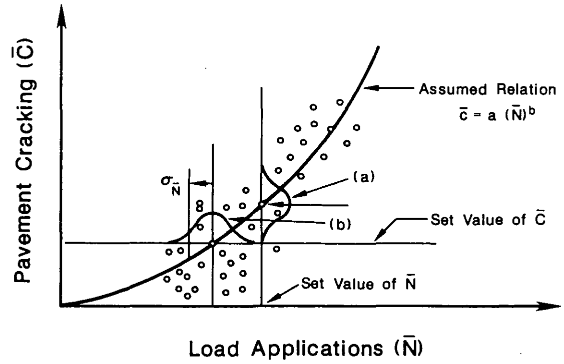

FIGURE 2 Pavement cracking versus load application.

Another common yet erroneous practice is the use of *means* rather than the observed individual cracking values as datum points. The means, as discussed earlier, have a much smaller variability, and hence the relationship appears to be better (more reliable) than it really is. The greater the number of pavements sampled, the worse this error becomes and the model appears to be better than it is. In fact a designer could arbitrarily make a model look good using this trick of modeling means.

Some researchers have recommended that rather than fit a model to the ''cloud'' of datum points a separate curve be established for each pavement using a model form that satisfies the necessary limits of the variables (7). By selecting an arbitrary but small lower value and then finding the curve parameters that will fit the observed point and the assumed lower point, a separate curve can be established for each pavement. This modeling method offers the advantage of controlling for other outside factors that make the pavements behave differently from each other. That is, suppose the cloud of points shows no relationship, yet the sample consists of pavements of one type that are known to have very steep curves and pavements of another type that are known to have relationships witha very low slope. Combining these pavements together as a cloud and trying to find one model that represents them all will result in a very poor model that does not represent any of them. The sketch in Figure 3 is an attempt to describe this situation. If separate models were established for each of the two types of pavements, this problem could be circumvented. In reality, however, one does not have knowledge of the factor type or any of a number of other factors that could result in similar obscurities. By letting each pavement stand alone, models that control for these unknown confounding effects are built.

Reliability can be established either on the basis of the number of load applications, that is, the probability that the traffic will not exceed a maximum level, or on the basis of cracking, that is, the probability that the cracking areas will not exceed a maximum acceptable level. For either case even if the probability distributions are not known, any of the four most common ones can readily be applied and tested to see which one is the best.

Another common misconception is that the material properties used in construction specifications to control quality can be considered to control reliability. Controlling the construction quality by monitoring test values of a single variable does not guarantee a specified design reliability. Design specifications are generally set arbitrarily. Yet implicit in these specifications are factors that affect design reliability, namely, the means and variances or coefficients of variation. To control the reliability knowledge of and experience with these coefficients of variation are essential. To set specifications without this knowledge is a dangerous practice.

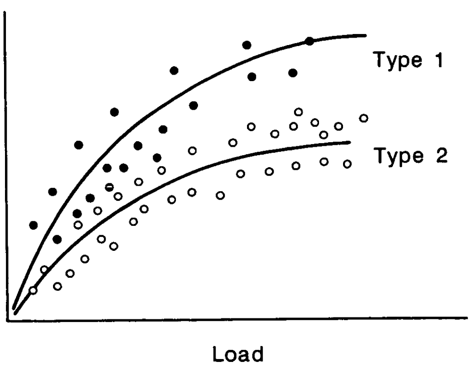

FIGURE 3 Example of mixing pavements of different types.

Returning to the schematic of Figure 1, another popular practice is the adoption of the ''black box'' approach when, in fact, there is knowledge of some of the interrelationships and interactions that take place in the input box. This ''blind'' approach in the face of knowledge makes absolutely no scientific sense. In fact examples can be constructed to show how futile it is in some cases to ever arrive at the known, true relationship if this ''naive'' approach is used.

CONCLUSIONS

Both empirical and nonempirical approaches should be merged in any modeling effort to produce a reliable design. Empirical relationships based on data should serve as guidelines in pavement design and performance modeling but cannot stand alone as conclusive and absolute indicators of the reliability of the models. The reason for this is the error box, the inherent error owing to variability of the random variables being observed. Each datum point represents only one realization in many, and it is the variability of the many that dictates the reliability of the models developed. Knowledge of the distribution of the random variables, which is available through the nonempirical mechanistic design relationships, can be used to incorporate this variability and hence establish reliability levels for the resulting pavement performance models.

In many areas of science little if any knowledge exists regarding the true relationships of observed phenomena. Investigators in these scientific disciplines have no choice but to blindly follow the observed data dictates and make decisions in the absence of known relationships. The result is often disappointing, and many replications and redesigns of experiments are required before progress, if any, can be achieved. The discipline of pavement performance modeling appears to have the advantage of a vast field of knowledge in the form of known physical laws and properties to serve as foundations on which to build models using empirical evidence along the way. It is truly a step backward in the scientific pursuit of truth to blindly pursue the naive, empirical approach and forsake all known engineering concepts.

The purely empirical approach has the additional, most dangerous potential of introducing bias and unfair practices in the inferential process. The statement ''you can prove anything with statistics'' emerged as a result of this all too inevitable bias. The fact is, it is not the ''figures (statistics) that lie but the liars that figure,'' and this is causing the problem. It is the subtle and perhaps even unintentional use of noncomparable models or hidden, invalid assumptions that allows bias to creep into the scientific method.

In short what is currently being proposed by this and other papers in this Record is the application of a standard, well-accepted mathematical methodology, along with the use of the principles of mechanistic design, to develop, in a scientific and unbiased way, pavement performance models that will be comparable and free of bias. It is appalling to think that an area as critical and costly as the designing of highway pavements has received such little attention from the standpoint of sound mathematical and scientific treatment, especially in light of the fact that both the mathematics and engineering principles have been in existence for nearly a century or more and are used routinely in so many other scientific arenas. It is not too late to begin implementing these procedures. The tools for accomplishing this have been refined over the years and are available and within practicable reach. One needs only to step forward, outside the cloud of skepticism and fear, to see them.

REFERENCES

1. Shooman, M. L. *Probabilistic Reliability: An Engineering Approach.* McGraw-Hill Book Co., New York, 1968.
2. Kotz, S., and N. L. Johnson. *Encyclopedia of Statistical Science,* Vol. 8. John Wiley & Sons, Inc., New York, 1988.
3. Kapur, K. C., and L. R. Lamberson. *Reliability in Engineering Design.* John Wiley & Sons, Inc., New York, 1977.
4. Bain, L. J. *Statistical Analysis of Reliability and Life-Testing Models.* Marcel Dekker, Inc., New York, 1978.
5. *AASHTO Guide for Design of Pavement Structures.* AASHTO, Washington, D.C., 1986.
6. Proschan, F., and R. J. Serfling. *Reliability and Biometry.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1974.
7. Lytton, R. L., and D. G. Zollinger. Modeling Reliability in Pavements. Presented at 72nd Annual Meeting of the Transportation Research Board, Washington, D.C., 1993.

Publication of this paper sponsored by Committee on Flexible Pavement Design.