

ROLLING STRAIGHTEDGE SAMPLING PLAN SIMULATION AND SPECIFICATION DERIVATION

Richard M. Weed, Division of Research and Development,
New Jersey Department of Transportation

The rolling straightedge is a device designed to measure large-scale surface irregularities, thus providing a means for evaluating the riding quality of a pavement. In this paper, it is used as an example to illustrate the derivation of a statistically based specification. A Monte Carlo simulation procedure is used to determine the component of variance attributable to the actual pavement roughness and the manner in which it is sampled. This is combined with the variance associated with the precision of the device to obtain the overall variance of the measurement process. The desired quality levels are selected and appropriate producer's and consumer's risks are chosen to derive the necessary specification acceptance limits for both bituminous and concrete pavement. A table is included that gives the variability associated with several possible sampling plans. Operating characteristic curves are presented that illustrate the capability of the resultant specifications. A discussion of the various ramifications of this method is included to provide assistance to those wishing to apply this general approach to other measurement processes to derive specifications with balanced producer's and consumer's risks.

•TO use a measurement process effectively, it is necessary to know how precise it is. In statistical terms, the variance is a measure of the repeatability and, therefore, the precision of the measurement. When the measurement is to be used as the basis for accepting or rejecting some item of construction (pavement in this example), the variance must be known in order to develop an acceptance procedure that will distribute the producer's and consumer's risks in an equitable manner. Ideally, the producer's risk (rejection of satisfactory work) and consumer's risk (acceptance of unsatisfactory work) should be zero, but this is often impossible or impractical to achieve. Alternatively, it is desired that both risks be equal and as small as possible. This paper describes the determination of the variance for a particular measurement process (rolling straightedge) and the development of a specification that satisfies these risk requirements.

BACKGROUND

The rolling straightedge, shown schematically in Figure 1, is a mechanical device designed to measure large-scale surface irregularities. It is pushed by hand along the pavement and automatically dispenses a dye to mark areas that deviate from a perfectly flat surface by more than some specified amount (usually $\frac{1}{8}$ in. within a length of 10 ft). From these dye marks it is possible to record the number of defects per unit length of pavement; measure the total length of the defects per unit length of pavement; or, if the depths or heights of the defects are noted, calculate an integrated value that accounts for both frequency and severity of these deviations.

The second option, which is termed the "percent defective length" of the pavement, has been recommended for quality assurance applications (1, 2). Correlation tests (1) with a BPR roughometer have demonstrated this to be a valid approach for the determination of pavement riding quality. Straightedge data in terms of the defective length parameter are currently being used for riding quality control in several states and are endorsed by the FHWA (2).

For a typical bituminous pavement in New Jersey, the percent defective length is about 1.0 percent. For New Jersey concrete pavement (expansion joints at regular intervals), the percent defective length may typically be as large as 9.0 percent.

The FHWA has recommended the use of the rolling straightedge for final acceptance of bituminous pavements with a graduated penalty schedule being applied for varying degrees of noncompliance (2). A knowledge of the precision of the measurement process is of obvious importance for this function and is also required for the establishment of the acceptance limits.

STATISTICAL CONCEPTS

There are two primary components of variance to be considered when making measurements with the rolling straightedge. These are the variability related to the precision of the instrument (σ_i^2) and the variability associated with the actual pavement roughness and the manner in which it is sampled (σ_s^2). Using the principle that variances of independent factors are additive, the total variance (σ_T^2) may be expressed as follows:

$$\sigma_T^2 = \sigma_i^2 + \sigma_s^2$$

In order to determine σ_T^2 , it is first necessary to find σ_i^2 and σ_s^2 . Of these, σ_i^2 is the easier to obtain since it may be calculated from several repeat readings on the same section of pavement. Strictly speaking, the value of σ_i^2 determined in this manner may contain a small component of variance associated with the roughness of the pavement. This is so because, if the operator strays off the intended line of travel, the pavement surface at that point may be of a slightly different roughness level and may produce a different (and thus more variable) reading. Since it is relatively easy to guide the rolling straightedge along the desired line, it is believed that this "pavement component" is a very small part of σ_i^2 . For the purposes of this study, it can be ignored because the repeat runs are typical of those made when evaluating an actual job. Therefore, whatever value of σ_i^2 is obtained can be expected to apply when future jobs are evaluated.

A previous study has shown that the standard deviation associated with the precision of the instrument is influenced to a small degree by the general roughness level of the pavement. On bituminous pavement, which is comparatively smooth, the instrument standard deviation was found to be approximately $\sigma_i = 0.30$ percent defective length. On concrete pavement, which is substantially rougher than bituminous pavement, a typical value for the instrument standard deviation is $\sigma_i = 0.40$ percent defective length.

DETERMINATION OF SAMPLING PLAN VARIANCE

Determination of the component of variance associated with the sampling plan (σ_s^2) is more involved, due in part to the several possible sampling plans that might be used. If 100 percent sampling were employed, there would be no sampling error (i.e., error due solely to fractional sampling) and the total variance would consist only of the variance due to the precision of the instrument. However, depending on the availability of both equipment and manpower, it is probably neither practical nor necessary to require 100 percent sampling.

Various fractional sampling plans must be explored to determine which is the most

appropriate. In the reference cited, the FHWA recommends continuous longitudinal sampling with the provision that the transverse location be chosen at random every 300 ft. Since it is our practice to make rolling straightedge measurements only at the approximate locations of the wheelpaths, there are then 2 possible transverse locations per lane that may be randomly selected. For pavement 2 lanes wide, this plan would result in a sampling fraction of 25 percent, since 1 of 4 possible wheelpaths would continuously be sampled. For the purposes of this study, sampling rates of 50 percent and 12.5 percent will also be investigated.

A prohibitive amount of work would have been required to perform these tests on actual pavement with the rolling straightedge. If this method had been used, approximately 200 man-days would have been required to obtain the data for the 8 tests listed in Table 1. By comparison, it required only 3 man-days to simulate these tests using Monte Carlo techniques. This illustrates the tremendous savings in both time and expense that can be realized by the use of Monte Carlo simulation.

Plots of the locations of defects on several New Jersey pavement sections proved to be especially useful for the simulation procedure. A typical plot is shown in Figure 2. The 4 closely drawn parallel lines represent the 4 wheelpaths of 2 adjacent lanes. Each run was $\frac{1}{4}$ mile long and was plotted as four 330-ft sections, with the stations indicated throughout each section. The defects were measured and plotted to the nearest foot, with no distinction being made between high and low readings.

These $\frac{1}{4}$ -mile plots thus represent roadways of known roughness. To test any sampling plan, the locations for sampling are determined as prescribed for that particular plan, and the number of defects is counted directly from the plot. Since it is possible to count the defects exactly, this procedure excludes instrument error and isolates the variance due solely to the sampling plan. Each sampling plan was repeated enough times so that a reliable determination of this variance could be made.

The particular roughness plots chosen for this study were selected so that several average levels of roughness would be represented. Each is a graphical representation of a section of an actual pavement in New Jersey judged to be reasonably typical of its type, whether bituminous or concrete. This selection was subjective but was based on the judgment of experienced engineers who have made roughness measurements on many miles of pavement within the state.

Although the FHWA report recommends that the wheelpath be randomly chosen at 300-ft intervals, this was approximated by using 330-ft intervals because it greatly simplified the use of the plot shown in Figure 2. A starting station was selected, a ruler was placed on the plot at this station perpendicular to the lines depicting the wheelpaths, and the defects were counted along the randomly selected wheelpaths using the ruler to mark the starting and stopping points. This use of 330-ft intervals greatly speeded the gathering of data and was assumed to have no significant effect on the results.

There are several methods by which the starting stations and wheelpaths could be randomly selected. In the cases studied, there were 4 possible wheelpaths (2 each in 2 lanes) and almost exactly thirteen 100-ft sections (actually 1,320 ft) in a run. Therefore, a deck of cards was used in conjunction with a 2-digit random number table. Selection with replacement from the deck of cards was used with suits determining the wheelpaths and face values determining the first digits of the starting stations. Selections from the random number table then provided the remaining 2 digits for the starting stations.

In all cases except one, the sampling plans were performed randomly and were repeated between 60 and 78 times to obtain reasonably accurate estimates of σ_s . For the case in which sampling was started at the beginning of the job rather than at a random starting point, it was possible to compute an exact numerical value for the sampling plan variance. Since there were four 330-ft sections, each with 4 wheelpaths, the total possible number of different samples was $4^4 = 256$. These were tabulated and used to calculate an exact value of σ_s for this particular case.

Although they are not shown in this report, histograms were plotted for each data set. In every case they were very nearly normal. This result was expected and served to confirm the validity of the computation of σ_s and its use in the procedures that followed.

Figure 1. Schematic drawing of a rolling straightedge.

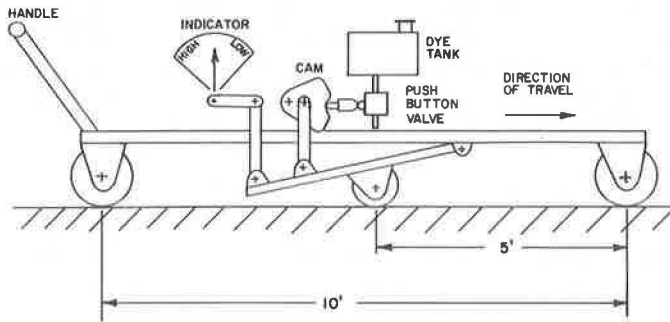
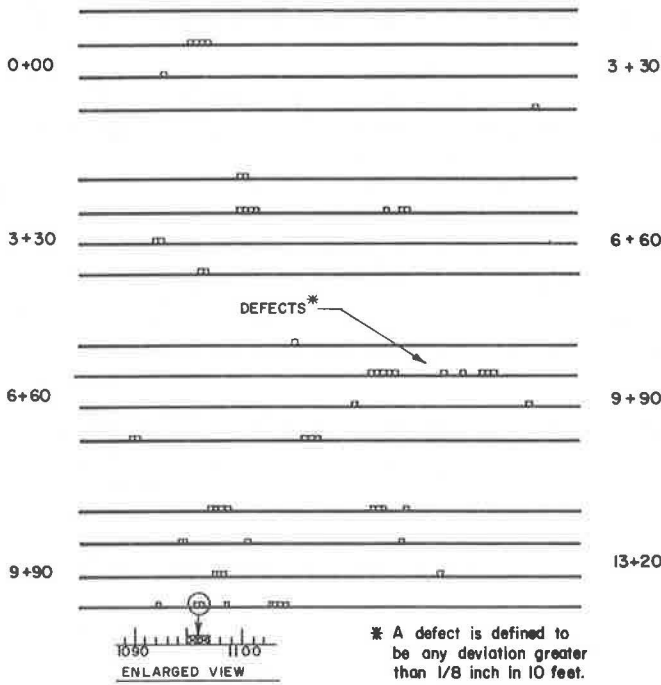


Figure 2. Typical graphical representation of a roadway of known roughness.



* A defect is defined to be any deviation greater than 1/8 inch in 10 feet.

Table 1. Rolling straightedge sampling plan simulation tests.

| Test | Pavement Type | Number of Replicate Runs | Sampling Plan | | True Mean Percent Defective | Sample Mean Percent Defective | σ_s Due to Sampling Plan |
|------|---------------|--------------------------|---------------|--|-----------------------------|-------------------------------|---------------------------------|
| | | | Percent | Details | | | |
| 1 | Bituminous | 78 | 25 | Random starting station, random selection of wheelpath every 330 ft | 0.38 | 0.42 | 0.24 |
| 2 | Bituminous | 72 | 25 | Same | 0.64 | 0.64 | 0.37 |
| 3 | Bituminous | 72 | 25 | Same | 2.44 | 2.49 | 0.72 |
| 4 | Concrete | 60 | 12.5 | Random starting station, 2 (only) successive randomly selected 330-ft sections | 8.21 | 8.08 | 1.99 |
| 5 | Concrete | 60 | 25 | Random starting station, random selection of wheelpath every 330 ft | 8.21 | 8.34 | 1.20 |
| 6 | Concrete | 60 | 25 | Start at beginning of job, random selection of wheelpath every 330 ft | 8.21 | 7.94 | 1.17 |
| 7 | Concrete | 256 | 25 | Same (exact numerical simulation) | 8.21 | 8.23 | 1.09 |
| 8 | Concrete | 60 | 50 | Start at beginning of job, 2 different randomly selected wheelpaths from each 330-ft section | 8.21 | 8.18 | 0.65 |

Table 1 gives the results obtained with various sampling plans and varying levels of roughness. (The characteristically higher percent defective level of concrete pavement is attributed to the expansion joints that occur at regular intervals of approximately 80 ft.) Four general observations can be made from the data in this table:

1. σ_s increases as the level of roughness increases, although the increase is not linear. The approximate relationship between σ_s and the percent defective length is shown in Figure 3. Values near the origin were obtained from bituminous pavement whereas the higher values were obtained from concrete pavement. This combination of the 2 types of data is felt to be appropriate because the rolling straightedge measures surface irregularities without regard to the nature of the surface. It would appear from the continuity of the curve in Figure 3 that this assumption is valid.
2. For a pavement that is quite rough, σ_s is strongly influenced by the level of the sampling fraction. This is shown in Figure 4 for a typical New Jersey concrete pavement. It is expected that this effect would be less pronounced on bituminous pavement, which typically has a much lower average level of roughness.
3. Essentially the same value for σ_s was obtained by starting the sampling at the beginning of the job as was obtained with a randomized starting station (1.17 versus 1.20).
4. The single exact numerical determination of σ_s provided a close check with the value obtained by the Monte Carlo simulation procedure on the same section of pavement (1.09 versus 1.17).

DEVELOPMENT OF THE SPECIFICATION

The components of variance associated with both the instrument and the various sampling plans are now reasonably well determined. These values, combined with engineering judgment concerning the acceptable quality level (degree of roughness), are sufficient to establish an appropriate specification. The steps are as follows:

1. Determine what constitutes satisfactory work and unsatisfactory work. The practical approach (assuming no other information is available) is to make a statistical survey of many jobs that are judged to represent both satisfactory and unsatisfactory workmanship. The parameters thus obtained can be used to establish appropriate criteria for the evaluation of future work.
2. Establish specification limits that, if complied with, will yield the desired results. Consideration must be given to the risks that can be tolerated by both the producer and the consumer. This will involve engineering judgments such as, "We can afford to accept up to 10 percent of the work below some specified level," or what is the same thing in the long run, "We can afford to run a 10 percent risk that we will accept work below this specific level." Another requirement might be, "We want no work to fall below some minimum level." In terms of producer's risk, an additional requirement could be stated, "We want any contractor who does satisfactory work to run no more than a minimal risk for having his work rejected."
3. Select a measurement process and a sampling plan that will determine whether the desired objectives are being achieved.

These steps will now be followed to illustrate how a rolling straightedge specification and sampling plan may be derived. For illustration purposes, suppose the engineering requirements are as follows:

1. Producer's risk—Based on historical data, a contractor who constructs a bituminous pavement with a percent defective value of 0.75 or less is doing good work and should run essentially a zero risk for nonacceptance.
2. Consumer's risk—Based on the previously cited correlation with the BPR roughometer and a study of historical data, a percent defective value greater than 2.5 is judged to be totally unacceptable for bituminous pavement. The risk for acceptance of

Figure 3. Relation between σ_s and percent defective length for 25 percent sample fraction.

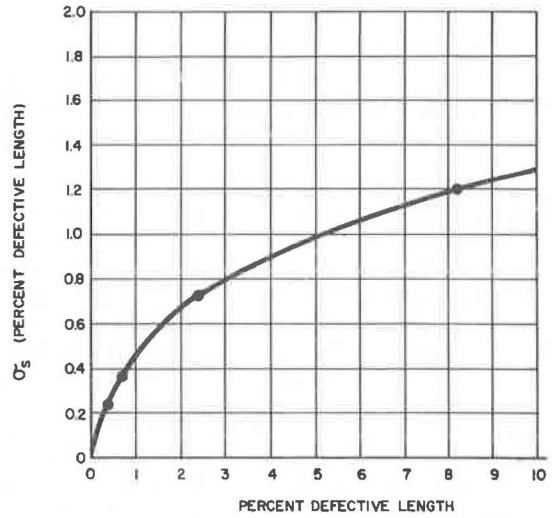


Figure 4. Relation between σ_s and size of sample fraction for concrete pavement with 8.21 percent defective length.

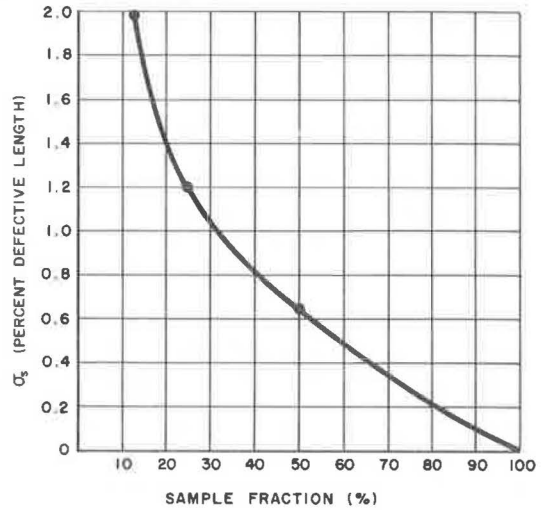
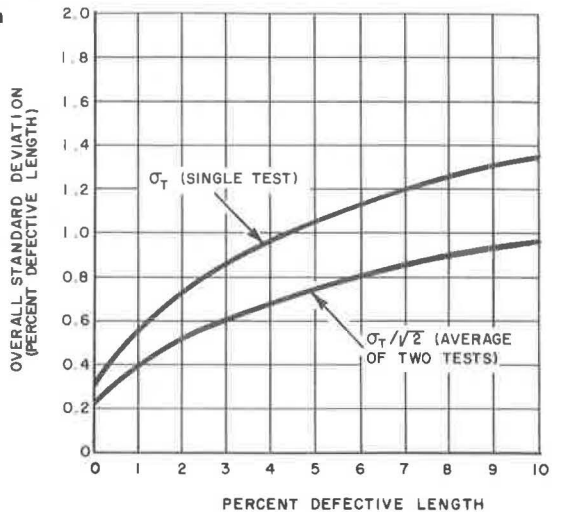


Figure 5. Relation between overall standard deviation and percent defective length for 25 percent sample fraction.



a pavement of this quality should be essentially zero.

At this point, in order to formulate the appropriate specification requirements, it is necessary to do some trial-and-error work with operating characteristic curves. An operating characteristic curve is simply a graphical representation of the probability of acceptance for all possible quality levels of the work. To construct such a curve, the standard deviation of the process must be known and an acceptance limit chosen.

Normally, the standard deviation is assumed to remain constant throughout the range covered by an operating characteristic curve. However, as can be seen from Figure 3, this will not be true for this example. Although the instrument standard deviation is assumed to remain constant ($\sigma_1 = 0.30$) throughout the normal percent defective range for bituminous pavement (0 to 3.0), the sampling plan standard deviation (σ_s) is seen to increase rapidly with increasing average roughness of the pavement. The overall standard deviation (σ_T) will vary accordingly since it represents a combination of both instrument and sampling variability. For convenience in constructing the operating characteristic curves, the σ_T values for varying average levels of roughness are plotted in Figure 5.

For example, at a percent defective length of 3.0, σ_s is found from Figure 3 to be 0.80. Since σ_1 is approximately 0.30 in this range of roughness, σ_T is calculated from the expression

$$\sigma_T = \sqrt{\sigma_1^2 + \sigma_s^2} = \sqrt{0.30^2 + 0.80^2} = 0.85$$

and is plotted at a percent defective level of 3.0 in Figure 5. Because σ_1 is known to be approximately 0.40 at higher levels of percent defective length, this value is used to calculate σ_T for the upper part of the curve. Because it will later be seen to be useful, $\sigma_T/\sqrt{2}$ is also plotted in this figure. This represents the standard deviation associated with the average of 2 tests.

Before discussing the results, it may be worthwhile to describe how the operating characteristic curves are derived. A trial-and-error procedure is required, the objectives being to balance the risks while keeping both the risks and the sampling rate at reasonably low levels. If extremely low risks are insisted on, then the sampling rate will be unnecessarily high. On the other hand, if a low rate of sampling is arbitrarily selected, then the risks (both to the producer and the consumer) may be too great. This turns out to be a situation that cannot be optimized but, instead, is one in which judgment must be used to select the plan that accomplishes all objectives to a satisfactory degree. The plans presented herein are felt to achieve this purpose, but it should be understood that they are by no means unique. Similar plans could be designed that might be considered equally good.

The operating characteristic curves illustrating the capability of the acceptance procedure for bituminous pavement are shown in Figures 6 and 7. Figure 6 shows the separate curves for the first test and (when required) a retest. Figure 7 shows the single operating characteristic curve that represents the combined probability of the acceptance procedure, i.e., the probability of acceptance by either the first or second test.

The plan shown in Figure 6 requires a sample fraction of 25 percent and involves 2 distinct acceptance limits: 0.9 maximum percent defective for the first test and 1.2 maximum percent defective for the average of 2 tests. In actual practice, if a value of 0.9 or less is obtained on the first test, the work is accepted. If a value larger than 0.9 is obtained on the first test, the complete test procedure is repeated (including the random selection of wheelpaths) and the average for the 2 tests is then required to be less than or equal to 1.2 to be accepted. The average of the 2 tests is taken as the final result, and no further tests are permitted. If a reduced payment schedule is to be applied, it should be based on this average result, and the onset of penalties should begin just above a percent defective level of 1.2.

Figure 6. Individual operating characteristic curves for a 25 percent sample fraction on bituminous pavement.

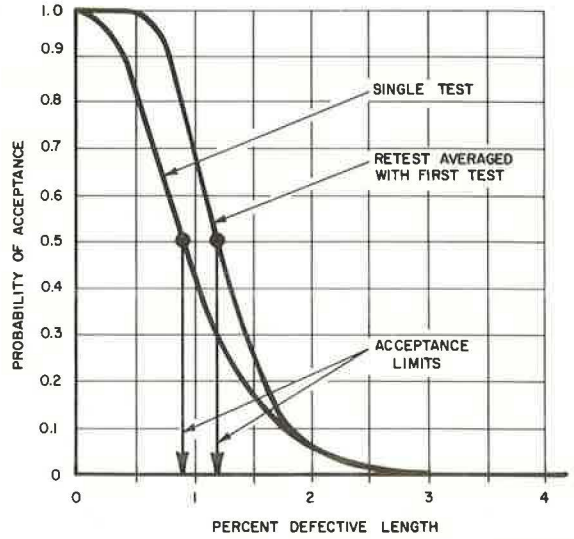


Figure 7. Operating characteristic curve for the complete test procedure with a 25 percent sample fraction on bituminous pavement.

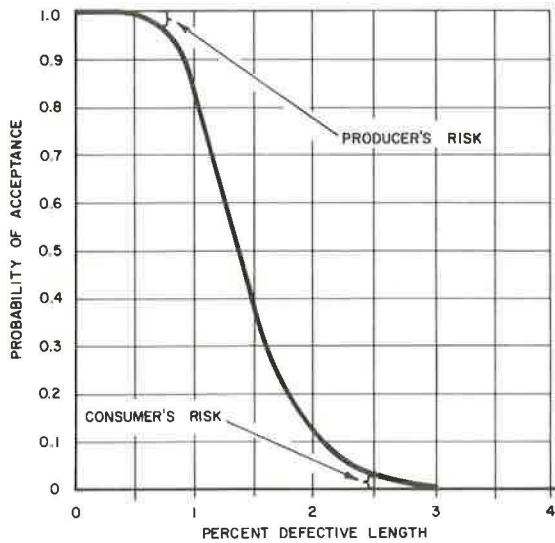


Figure 8. Operating characteristic curve for a 25 percent sample fraction on concrete pavement.

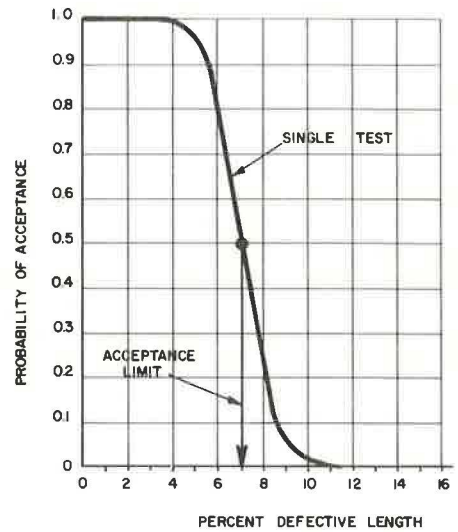


Table 2. Summary of the specification.

| Item | Bituminous Pavement | Concrete Pavement |
|---|---|---|
| Upper limit, which desirably is never accepted | 2.5 percent defective length | 11.0 percent defective length |
| Lower limit, which desirably is never rejected | 0.75 percent defective length | 4.0 percent defective length |
| Sampling procedure and calculation required for first test | Continuous measurement starting at beginning of job, random selection of wheelpath every 300 ft, calculate percent defective length | Continuous measurement starting at beginning of job, random selection of wheelpath every 300 ft, calculate percent defective length |
| Acceptance criteria for first test | Percent defective length must be 0.9 percent or less | Percent defective length must be 7.0 percent or less |
| Sampling procedure and calculation required for second test | Repeat first test procedure (again choosing random wheelpaths) and average the results of both tests | Second test not permitted |
| Acceptance criteria for second test | Average percent defective length must be 1.2 percent or less | Second test not permitted |

Note: In order to simplify the illustration, this specification applies only to pavement that is 2 lanes (four wheelpaths) wide. To satisfy the same requirements at the same risk levels for pavements of a different width, it may be necessary to change the sample fraction or the acceptance criteria or both. For a more complete treatment of this aspect of the subject, the reader is referred to the work of Croteau (1).

To illustrate how the operating characteristic curves are actually plotted, consider the "single test" curve of Figure 6. At a percent defective length of 2.0, for example, it is desired to know the probability that the pavement will be accepted (i.e., the probability that the testing procedure will measure a value of 0.9 or less). For a single test, σ_T is found from Figure 5 to be 0.73 at a percent defective length of 2.0. By using a table of the cumulative normal distribution, the probability of acceptance is found as follows:

$$Z = \frac{X - \mu}{\sigma} = \frac{0.9 - 2.0}{0.73} = -1.51$$

$$\alpha = 0.066$$

The probability of acceptance is then 6.6 percent, which is plotted at a percent defective length of 2.0 in Figure 6.

The operating characteristic curve in Figure 7 is obtained in a different manner. Here it is desired to know the probability of acceptance for the test procedure as a whole. Since the pavement could be accepted by either the first test or the second test, this curve is obtained by adding the probabilities of these 2 events. If the probability of passing the first test is designated P_1 and the probability of passing the second test is designated P_2 , then the combined probability of acceptance (P_c) can be calculated from the relationship

$$P_c = P_1 + (1 - P_1) P_2$$

For any particular level of percent defective length, the values for P_1 and P_2 are read directly from Figure 6, the calculation is performed, and the result is plotted in Figure 7.

Since Figure 7 represents the application of the test procedure as a whole, it is here that we check to see if our objectives have been achieved. It can be seen that the risks are balanced and small, being approximately equal to 3.5 percent at the critical levels of 0.75 and 2.5 percent defective length.

One final remark is appropriate in regard to this sampling plan. Recognizing that the 2 distinct acceptance levels of 0.9 and 1.2 do not constitute a unique solution, one might wonder if it would be possible to find a single value for both acceptance levels that would permit the risks to be balanced and, if so, what the risks would be. The answer is yes, it can be done, and the result is a plan whose slightly higher risks might be considered by some to be justified by the simplicity of the single acceptance level for both tests. The value for the single acceptance level is 1.07 and the risks are balanced at about 5 percent. If this acceptance level is rounded off to 1.1, the risks go slightly out of balance, being approximately 4 percent and 5 percent for the producer and consumer respectively.

The same general approach is followed to develop the test procedure for concrete pavement. Because our concrete pavements are not only rougher but also more variable, it is somewhat difficult to select a specific roughness that separates acceptable and unacceptable work. For illustration purposes, let us define a level of 4.0 as definitely acceptable and a level of 11.0 as definitely unacceptable. A 4.0 percent defective pavement should almost always be accepted whereas an 11.0 percent defective pavement should almost always be rejected.

In this case, the resulting specification will be different from that obtained for bituminous pavement. The 2 levels at which a risk requirement is imposed are far enough apart so that a single test with a sample fraction of 25 percent is capable of satisfying both requirements. An acceptance level of 7.0 should be specified and only a single

test would be permitted. The operating characteristic curve for this test is shown in Figure 8.

SUMMARY

This study illustrates the use of a Monte Carlo simulation procedure to determine the component of variance associated with the sampling plan for a measurement process. For this particular example (rolling straightedge), the sampling plan variance was found to be a substantial part of the overall variance. The overall variance, combined with acceptable quality levels and appropriate risk levels, was used to derive a statistically based specification, which is summarized in Table 2. It was also indicated by the simulation tests that the starting location for sampling may always be the beginning of the job, if desired, with no appreciable effect on the precision of the method.

Although the results obtained in this study apply directly to New Jersey pavement and the particular model of rolling straightedge used, this work should serve as a useful guide to anyone who plans to use a device of this type. It may be of still further use to anyone who might wish to apply this general approach to other measurement processes to derive specifications with balanced producer's and consumer's risks.

REFERENCES

1. J. R. Croteau. Pavement Riding Quality. New Jersey Department of Transportation, Sept. 1973.
2. Improved Quality Assurance of Bituminous Pavements. Federal Highway Administration, New Jersey Project Report, Region 15, Jan. 1973.