Variability in the Testing and Production Of Bituminous Mixtures

J. HODE KEYSER, Chief Engineer, Control and Research Laboratory, Department of Public Works, Montreal, Canada; and
P. F. WADE, Associate Director, Management Consulting Services, Price Waterhouse and Co., Montreal, Canada

•LIKE most production processes, the production of bituminous mixtures is subject to variation. This variation can broadly be attributed to two major sources related to (a) mixing, composition and the characteristics of the constituents, and (b) sampling and testing.

From a practical standpoint, a statistical evaluation of the sources and magnitude of the variability is almost essential to:

1. Establish the capability of a given testing procedure.
2. Establish the capability of the production process with respect to each property.
3. Refine the production process where necessary and practicable.
4. Set realistic specifications.
5. Provide statistical measures to control the process and evaluate experimental data.

Figure 1 shows how these two major sources can be broken down into components. The supplier-to-supplier variation is related to factors such as uniformity of composition, the characteristics of the materials, and the efficiency of the process. Different plants have different process capabilities, which are reflected in the following:

1. Components related to mixing, composition and characteristics of the constituents:
   (a) Day-to-day variation in the process efficiency and the uniformity of the mixture.
   (b) Batch-to-batch variation in uniformity of the material.
2. Components related mainly to sampling and testing:
   (a) Sample-to-sample variation of the uniformity of the material within a batch and the representativeness of the samples.
   (b) Laboratory-to-laboratory variation.
   (c) Operator-to-operator variation.
   (d) Day-to-day variation due to the changes in testing conditions on different days.
   (e) Briquet-to-briquet variation, expected between determinations performed by a single operator on the same sample using the same apparatus on the same day.

Odasz and Nafus (8) and Shook (9) studied plant-mix variation, whereas Corbett (12) and Vokac (11) studied laboratory mixtures. The scope of these four investigations is compared in Table 1, from which it is evident that they differ in many respects. The statistical measure of variation quoted in each case reflects these differences (materials, mixing procedures, etc.) and also differences in the basic variation components, which were combined in the estimate. This latter difference arises partly as a result of the method of statistical analysis, given in the last column of Table 1.

The complexity of the variation problem and the importance of clear definitions for the scope of any statistical studies are apparent from the foregoing.

The purpose of this paper is to illustrate by practical examples further applications

Paper sponsored by Bituminous Division, Department of Materials and Construction.

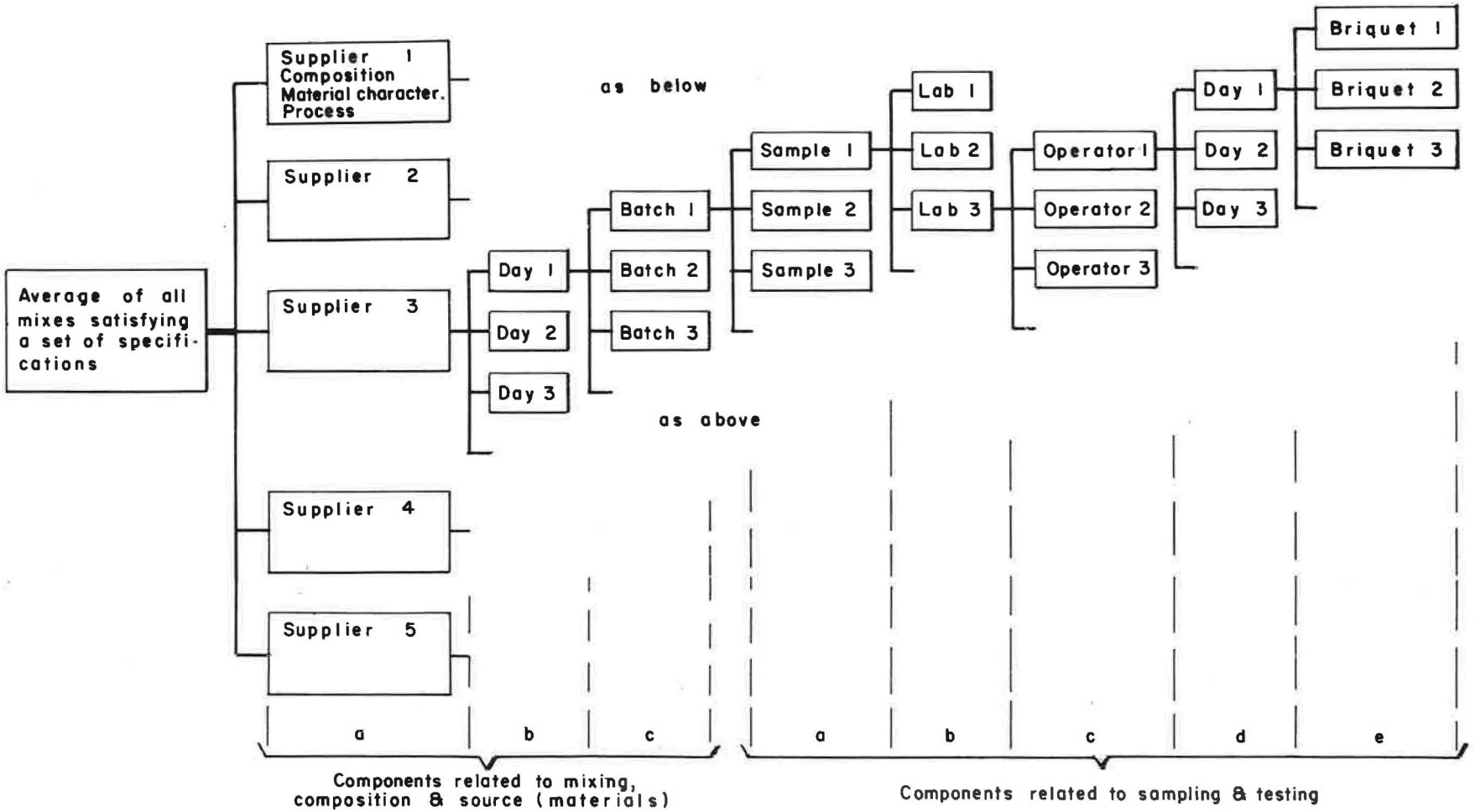


Figure 1. General sources of variation.

TABLE 1

LIMITATION AND SCOPE OF FOUR VARIATION STUDIES

Reference	Type of Materials	Size Distribution	Description of Sample and Test	Component of Variance[1]
Odasz and Nafus (8)	Plant mix: 100-120 pen. asphalt; crushed limestone agg.; natural sand.	1-in. max., 57% pass No. 8, 15% pass No. 200	Two samples obtained from separate batches in one truck; hand compaction, 50 blows per face; one briquet per sample; one briquet per run.	$\sqrt{\sigma^2_{BB} + \sigma^2_{res}}$ (plant)
Corbett (12)	Lab. mix: 85-100 pen. asphalt; (a) crushed stone screening, blank sand; (b) slag, screening, blank sand	½-in. max., 43% pass No. 10, 5% pass No. 200	New batch for each briquet; hand compaction, 50 blows per face; one briquet per run.	$\sqrt{\sigma^2_{BB} + \sigma^2_{res}}$ (plant)
Shook (9)	Plant mix: 85-100 pen. asphalt, crushed dolomitic limestone coarse agg.; natural coarse and fine sand; mineral filler	1-in. max., 46% pass No. 10, 5.2% pass No. 200	Two samples obtained from separate batches in one truck; mechanical compaction, 50 blows per face; two briquets per run.	$\sqrt{\sigma^2_{BB} + \sigma^2_{res/2}}$ (plant)
Vokac (11)	Lab. mix; 85-100 pen. asphalt; gravel; natural sand; Ottawa sand; limestone filler	⅜-in. max., 54% pass No. 8, 6% pass No. 200	Factorial design; 4 samples per day, 4 days per week, 4 consecutive weeks; new batch for each briquet; kneading compactor; 2 samples per run.	σ_{res}

[1]BB = between batches; res = residual.

of statistical methods. The study covers a laboratory and a field investigation of two mixtures (surface and base) produced by two plants. The study is reported in the following four parts:

1. An analysis of the repeatability of the Marshall stability and density tests and Rice's maximum density test. The limitations of the Marshall test and Rice's test were determined by analysis of variance studies made on test results obtained from laboratory-prepared mixtures. The scope and limitations of the experiment are shown in Figure 2.
2. An analysis of the variations occurring within a well-controlled production process. The extent of production variation was assessed by the statistical analysis of test results from controlled production samples. The scope of this work is shown in Figure 3.
3. A discussion of the influence of unavoidable process variations on mix design and the setting of specifications.
4. A discussion of the use of statistical control charts.

No attempt was made to study the technical validity of the Marshall test in assessing the in situ quality of the material.

REPEATABILITY OF MARSHALL AND RICE'S MAXIMUM DENSITY TESTS

Strictly speaking, testing variation or precision is the variability in results which can occur between replicate tests made on the same test piece or specimen over a number of days using a variety of operators and testing machines. However, in the case of the Marshall test, because the test is destructive, a different briquet must be used each time and a certain amount of unmeasurable and unremovable variability due to sampling and mixing (lack of homogeneity) is introduced.

Because of limitations on the number of operators and testing machines available for this study, the work was necessarily restricted to only one operator and one set of compaction and testing equipment. The influence of different operators and test equipment must not be neglected, however, where interlaboratory comparisons are being made, and these can be assessed through round robin tests (1).

This and other studies (2) have shown that, for a given equipment-operator combination, the magnitude of the "testing" variation of the Marshall tests may vary with each "design formula-supplier" combination. This is demonstrated in Table 2, which gives the "between duplicate briquet" variability for material supplied by two contractors to

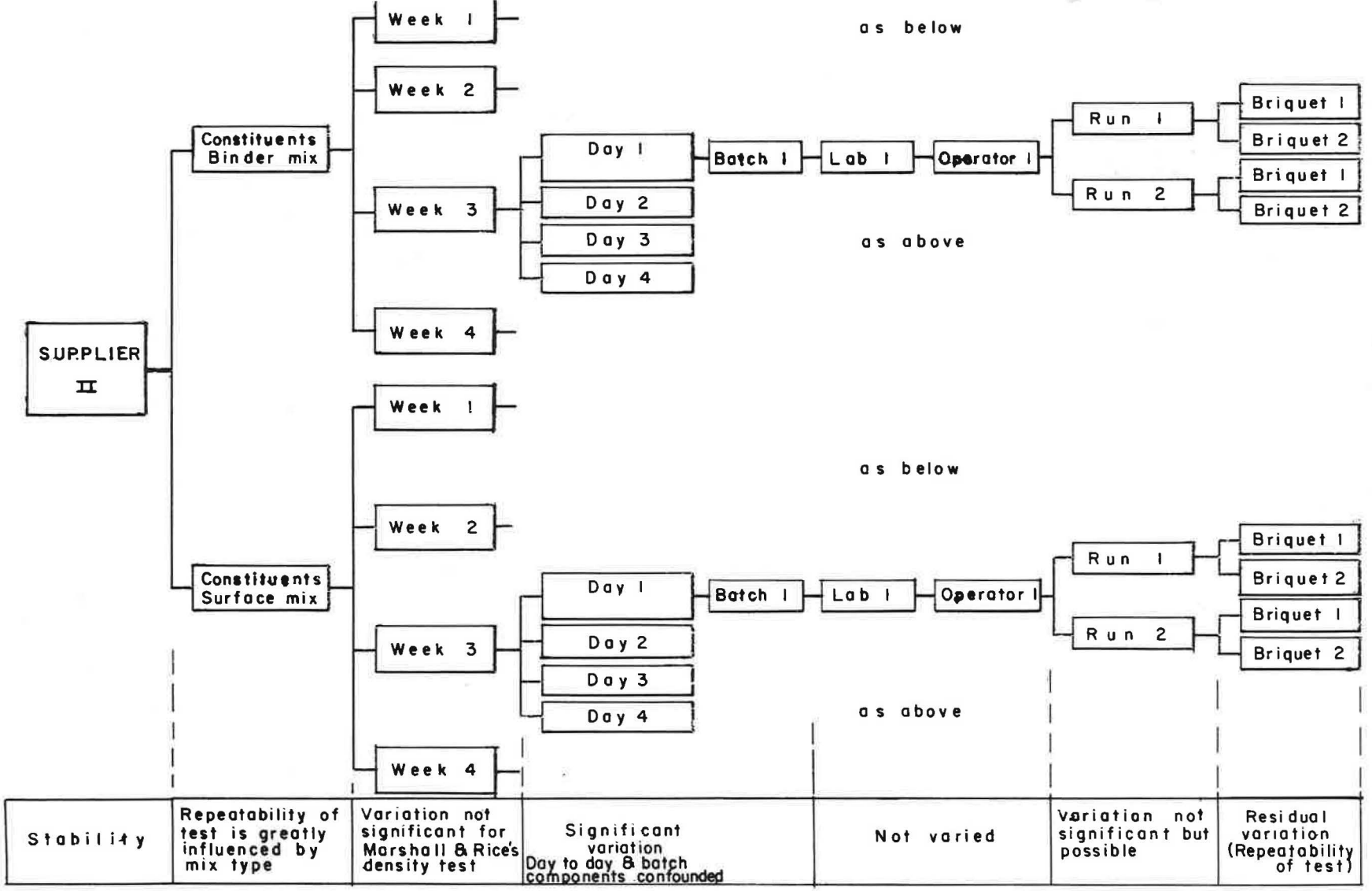


Figure 2. Analysis of variance, laboratory investigation.

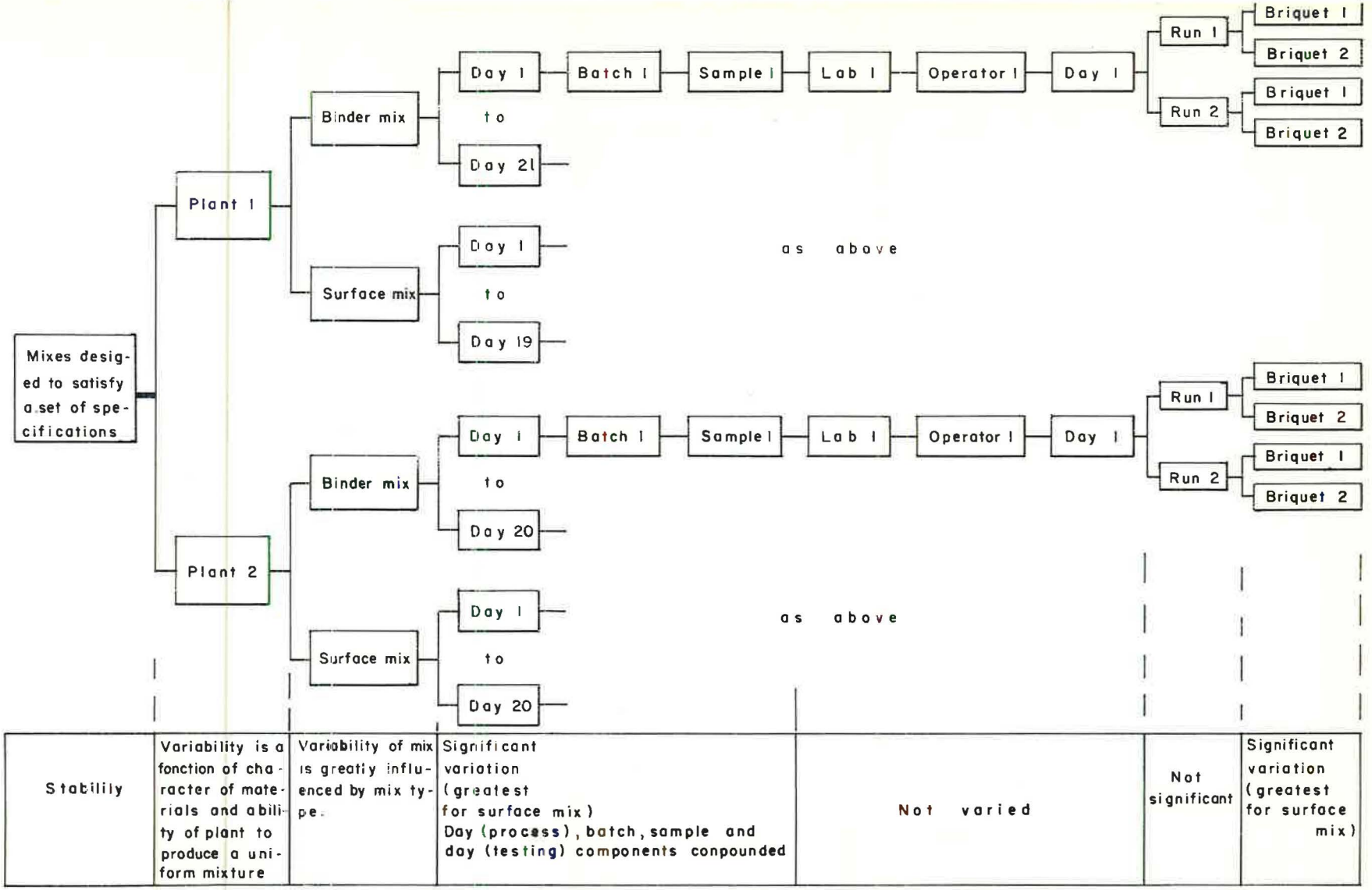


Figure 3. Analysis of variance, field investigation.

TABLE 2

MARSHALL TESTING VARIABILITY OF SIMILAR BITUMINOUS MIXTURE FROM TWO CONTRACTORS[1]

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Components of Variance[2] Estimated by Mean Square
(a) Analysis of Variance, Contractor C				
Between days	16	2,858,702	178,669	$2\sigma^2_{BD} + \sigma^2_{WR}$
Within day (within a run)	17	692,825	40,754	σ^2_{WR}
	$\hat{\sigma}^2_{WR}$ = 40,800 lb			$\hat{\sigma}^2_{BD}$ = 69,000 lb
(b) Analysis of Variance, Contractor D				
Between days	17	1,274,481	74,969	$2\sigma^2_{BD} + \sigma^2_{WR}$
Within day (within a run)	18	207,338	11,518	σ^2_{WR}
	$\hat{\sigma}^2_{WR}$ = 11,500 lb			$\hat{\sigma}^2_{BD}$ = 31,700 lb

[1] Original data taken from routine production test results; two Marshall briquets, one sieve analysis, and one extraction test made on each sample.

[2] BD = between days; WR = within run; F test: $F_{(17, 18)} = \frac{40,754}{11,518} = 3.5$. This is statistically significant at 0.01 level.

approximately the same design specifications. Contractor C's "testing" variability, $\hat{\sigma}^2_{WR}$, is significantly higher than that of contractor D at the 0.01 level of probability.

Thus, it must be recognized that any particular study of the precision of the Marshall test can rarely be applied directly to other laboratories or to material from other suppliers. In addition, the generality of the components of variance reported in the literature is restricted by the mixing methods used to prepare the material for the briquets. For example, in one study (11) each briquet was made from a small batch mixed separately in the laboratory and tested hot, whereas in the investigation reported here four briquets were produced from a 20-lb sample of the contractor's material.

To evaluate some sources of variability of the Marshall and the Rice's vacuum saturation tests, two statistically designed experiments were carried out.

Experimental Conditions

Experimentation was done on two types of mixtures: a 1-in. nominal size binder mix and a ½-in. nominal size surface mix. The experiments were designed to evaluate the effects of four factors, the results to be presented, where possible, as components of variance. Other known factors were controlled by the experimental techniques.

The factors studied were as follows:

1. The Marshall compaction hammers. A two-hammer compaction machine was used. The hammers were designated A and B.
2. The run-to-run variation. Two briquets were compacted per run and two runs were made successively each day.
3. The day-to-day variation. Four consecutive days in each week were devoted to making briquets, the testing being done on the day following.
4. The week-to-week variation. The experiment was carried out over four successive weeks.

The following were closely controlled:

1. Gradation, quality and character of the coarse and fine aggregates and the mineral filler.
2. Grade and character of the asphalt cement.
3. Mixture composition.
4. Preparation and testing technique.

Test Specimens and Testing Technique

Materials.—The mixtures were prepared from three sizes of crushed limestone coarse aggregate, a crushed limestone screening, a natural siliceous coarse sand, a mineral filler, and an 85-100 penetration grade asphalt.

Preparation of Batch Mixes and Compaction.—An 8,000-g batch was prepared each day for each of the two types of mixture in order to provide specimens for four briquets, two Rice's vacuum saturation tests, and one sieve analysis. The procedures used for the preparation of batch mixes and compaction are similar to those described in the

TABLE 3

COMPONENTS OF VARIANCE OBTAINED FROM MARSHALL STABILITY, MARSHALL DENSITY AND RICE'S MAXIMUM DENSITY[a]

Determination	Mix Type	Est. Component of Variance, $\hat{\sigma}^2$							
		Within Run		Between Runs[a]		Between Days		Between Weeks	
		Deg. of Freedom	Value	Deg. of Freedom	Value	Deg. of Freedom	Value	Deg. of Freedom	Value
Marshall stability	Binder	31	36,000	15, 31	21,800[b]	12, 15	7,000	3, 12	0
	Surface	31	3,500	15, 31	550	12, 15	5,000[c]	3, 12	0
Marshall density	Binder	31	33×10^{-6}	15, 31	8×10^{-6}	12, 15	17×10^{-6}	3, 12	0
	Surface	31	20×10^{-6}	15, 31	0	12, 15	21×10^{-6c}	3, 12	5×10^{-6}
Rice's max. density	Binder	-	-	-	19×10^{-6}	12, 16	54×10^{-6}	3, 12	0
	Surface	-	-	16	9×10^{-6}	12, 16	26×10^{-6c}	3, 12	1×10^{-6}

[a]Within a day.
[b]Significant at 0.05 level.
[c]Significant at 0.01 level.

Asphalt Institute's "Mix Design Methods for Hot Mix Asphalt Paving," except that (a) the mechanical compactor was used and 60 blows were applied on both top and bottom of the specimen (equivalent to 75 blows of the hand-operated Marshall hammers), and (b) two briquets were molded at the same time (or on one run).

Testing Methods.—The testing methods used were (a) for the Marshall Test, ASTM D1559 60T Tentative Method of Test for Resistance to Plastic Flow of Bituminous Mixtures Using Marshall Aparatus (the four briquets were processed together and the tests were made in order of compaction); and (b) for the maximum density test, Rice's vacuum saturation procedure (13).

Test Results and Conclusions

A sample set of the original data from the experiment and specimen analysis of variance calculations are given in the Appendix. The components of variance are summarized in Table 3, from which the following conclusions can be made for the Marshall stability test:

1. Individual test results appeared to approximate a normal distribution as regards "within-run" variability. This was demonstrated by the following statistical test: If the individual results are normally distributed, the within-run differences (range with sign) will also be normally distributed, with mean 0 and standard deviation $\sigma_d = \sqrt{2}\,\sigma_{WR}$, assuming there is no bias between the first and second briquet within each run. These differences were plotted on a control chart and found to fall between the prescribed limits. The cumulative frequencies approximated a straight line when plotted on normal probability paper.
2. There was no statistically significant difference due to hammer positions for either the binder or the surface mix.
3. There was a significant difference, or bias, between the first and second pair of briquets, the first pair being lower by an average of 200 lb for the binder mix and 80 lb for the surface mix. Further investigation failed to locate any cause for this bias. However, as it did not appear in other test series with the same equipment and operator it was concluded that this was nonrepresentative and that no routine correction factor was necessary to adjust for this effect. The operator was told to check routine data for similar signs, but none were observed. In the discussions which follow, this bias is ignored.
4. If "repeatability" is defined as a statistical measure (1σ) of the variation possible between single determinations on a single sample performed by a single operator using the same equipment and technique on different runs on a single day, the repeatability under the conditions prevailing at the City of Montreal Control and Research Laboratory with the specific mixes studied is

Mix Type	$\sqrt{\hat{\sigma}^2_{BR} + \hat{\sigma}^2_{WR}}$
Binder	240 lb
Surface mix	60 lb

Similarly, conclusions regarding the Marshall density test can be drawn as follows:

1. There was no statistically significant difference at 0.05 level due to hammer positions for either the binder mix or the surface mix.
2. There was no bias between the first and second pair of briquets.
3. If repeatability is defined as before, the repeatability under the same conditions is

Mix Type	$\sqrt{\hat{\sigma}^2_{BR} + \hat{\sigma}^2_{WR}}$
Binder	0.0064
Surface mix	0.0045

Similar conclusions regarding Rice's maximum density test can be drawn, as follows: If repeatability is defined as before, the repeatability under the conditions prevailing at the City of Montreal Control and Research Laboratory with the specific mixes studied is

Mix Type	$\hat{\sigma} = \sqrt{\hat{\sigma}^2_{WD}}$
Binder	0.0044
Surface mix	0.0030

Components of Variance

The total test variance for a single result taken within a single laboratory and using a single operator is composed of day-to-day, run-to-run, and within-run variation (assuming the week-to-week component is zero). This is expressed by

$$\sigma^2_{\text{Total test}} = \sigma^2_{BD} + \sigma^2_{BR} + \sigma^2_{WR} \quad (1)$$

The within-run component, σ^2_{WR}, gives an indication of the precision when testing conditions are as uniform as possible, inasmuch as the two briquets in each run are compacted together and follow one another through the rest of the procedure. In addition, this term includes the mixing and sampling variation that can arise between the two briquets.

The between-runs component, σ^2_{BR}, gives an indication of the additional variable element introduced into the procedure as a result of the different conditions prevailing (either on the compactor or later in the procedure) from the first pair of briquets to the second. In addition, any sampling and mixing differences which prevail between the first and second pair of briquets are included in this term.

The between-days component, σ^2_{BD}, gives an indication of the additional variation introduced when the testing is prolonged over several days. Because in the present study a new batch was prepared in the laboratory each day, this between-days compo-

nent is equivalent to the sum of the between-days-test component, σ^2_{BD}-test, and the between-batch component, σ^2_{BB}. It may reflect mainly the between-batch variation.

The foregoing components give a useful statistical measure of the capability of the testing procedure as currently in use at the City of Montreal Control and Research Laboratory. Examples are given in a later section ("Controlling the Test Procedure and the Supplier's Process") to demonstrate some of the possible applications.

FIELD INVESTIGATION

In the control of bituminous mixtures (or in setting control limits for a given plant) the capacity of the plant to produce a uniform mixture and the unavoidable variation in the composition and character of the mixes must also be evaluated in order to set realistic control limits.

The purposes of this part are (a) to evaluate the ability of two well-controlled plants to produce uniform surface and binder mixes, using different types of aggregates, and (b) to discuss briefly the factors affecting the variability of production.

The major sources of variation in the production and control of asphaltic mixtures are (a) plant equipment, (b) materials, and (c) sampling and testing procedures.

The capability of a plant to produce a uniform mix within the range of the specification is a function of the efficiency of equipment, such as the aggregate feeding system, the heat and draft efficiency of the dryer, the screen efficiency, the temperature control system for aggregate and asphalt, the precision of the scales, and the efficiency of the pugmill. In addition, some variation in bituminous mixtures production may be attributed to human error. However, the batch plant manurfacturer has largely overcome the human error in proportioning by making automatic controls for weighing and proportioning batches.

The materials fed to the asphalt plant constitute another source of variation, the variables being the quality and character of the aggregate and filler, the grade and kind of asphalt cement, and the gradation of the aggregates. In normal production, the quality and character of both asphalt cement and aggregates are usually quite consistent, whereas the grading of the aggregates is subjected to some unavoidable variations.

The third major source of error is inherent in sampling and testing procedures. As in any controlling system, the representative state of the sample is related to the sampling method and the precision of the test is a function of the testing technique.

As shown by the variables previously outlined, the production of asphaltic mixtures is subject to variations and a reasonable allowance must be provided.

The Experiment

The scope and limitations of the field investigation are shown in Figure 3. In brief, binder and surface mixes produced by two local well-controlled plants were sampled during 20 consecutive days. For each type of mix, a hot sample was taken from a single batch (selected at random) and brought to the central laboratory. The sample was then split into eight parts in order to provide subsamples for four Marshall stability (compacted two to a run) and density determinations, two Rice's maximum density tests, and two extraction and grading tests.

The sampling and testing methods and the characteristics of the plants are described in the following.

Sampling and Testing Methods

Unless otherwise specified, the sampling method used was the one described in ASTM Designation D 979-51 under the title "Sampling Plant Mixed Bituminous Mixtures at Place of Manufacture." The testing methods used are: ASTM Designation D 1559-60T Tentative Method of Test for Resistance to Plastic Flow of Bituminous Mixtures Using Marshall Apparatus"; extraction by ASTM Designation 1097-58 Standard Method of Paving Mixtures by Centrifuge, except that the dust correction was determined by centrifuging two 100-ml aliquot portions (3); and maximum density by Rice's vacuum saturation procedures (3).

Characteristics of Plants

The mixtures were produced in two plants. Plant I is a fully automatic 6,000-lb capacity batch plant. The dryer capacity is 180 tons per hour (88-in. diameter, 36-ft length). The aggregates are fed to the dryer from five cold bins by calibrated vibration feeders (Syntron). The hot aggregates are screened into five bins and the bitumen is supplied to the pugmill volumetrically. Plant II is a semi-automatic 5,000-lb batch plant. The dryer capacity is 140 tons per hour (diameter 70 in., length 30 ft). The cold feed is calibrated by gate opening and controlled by Weight-O-Matic system. The hot aggregates are screened into four bins and the bitumen is supplied to the pugmill volumetrically.

Both plants were continuously checked and calibrated by the supplier's experienced bituminous engineer.

Mix Type

Two types of dense-graded bituminous mixtures produced by each plant were subjected to investigation--½-in. nominal size surface mix, and 1-in. nominal size base course mix.

Mixtures produced by plant I were prepared from crushed limestone coarse aggregates, manufactured (crushed stone) sand, and natural medium to fine sand; mixtures

TABLE 4

COMPOSITION OF MIXTURES PRODUCED BY TWO PLANTS

Item	Plant I		Plant II	
	Surface Mix	Base Course	Surface Mix	Base Course
Mix composition (%):				
¾-in. crushed stone	13.0	-	20	-
½-in. crushed stone	13.0	17.0	20	-
¼-in. crushed stone	22.5	-	7	11
Screening	-	-	-	15
Manufactured sand	47	29	-	-
Coarse sand	-	41	43	43
Medium sand	-	-	-	-
Fine sand	-	-	10	24
Mineral filler	4.5	13.0	-	7
Grading (%):				
1 in.	-	100	-	100
¾ in.	-	95.9	-	96.6
½ in.	100	84.7	100	80.7
⅜ in.	99.1	78.1	99.0	61.0
No. 4	83.5	54.5	87.1	50.2
No. 8	68.7	39.5	73.5	44.6
No. 16	58.5	32.1	66.1	40.2
No. 30	45.7	23.9	52.8	28.7
No. 50	39.1	20.2	41.6	18.3
No. 100	20.7	11.2	27.5	9.7
No. 200	8.9	4.9	11.0	2.6
Bit. content (%):	6.7	4.8	6.3	4.93

prepared by plant II were prepared from crushed shaly limestone coarse aggregates and natural coarse to fine sand. The compositions of the mixtures are given in Table 4.

Test Results and Conclusions

A sample set of the original data from the experiments and specimen analysis of variance calculations are given in the Appendix. A summary of the components of variance for Marshall and Rice's maximum density test results are given in Table 5, from which the following conclusions can be drawn regarding Marshall stability:

1. Comparing the variations between results obtained from suppliers I and II, the components of variance in general appeared to be higher for supplier I. Tests on within-run components indicate that σ^2_{WR} for supplier I (surface mix) was significantly higher than for supplier II at the 0.05 level of probability. The difference in variation is possibly caused by (a) the character of the materials and (b) the ability of a plant to produce a uniform mixture.
2. Variability of mix seems greatly influenced by mix type. By comparing the within-run variation of binder and surface mixes produced by suppliers I and II, it was found that surface mixes are significantly less variable than binder mixes at the 0.01 level of significance.
3. The bias between the first and second run was not statistically significant at the 0.05 level for any one of the four sets of data.
4. If the repeatability is defined as previously, and the process variability as the square root of the sum of between-days, between-runs, and within-run variances, the following standard deviation values are obtained from the present conditions:

Supplier	Mix Type	Repeatability of Marshall Test (lb), $\sigma = \sqrt{\hat{\sigma}^2_{BR} + \hat{\sigma}^2_{WR}}$	Process Variability (lb), $\sigma = \sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{BR} + \hat{\sigma}^2_{WR}}$
I	Binder	235	305
	Surface	90	185
II	Binder	225	270
	Surface	60	155

TABLE 5

COMPONENTS OF VARIANCE OBTAINED FROM FIELD INVESTIGATION

Determination	Mix Type	Supplier	Estimated Component of Variance, $\hat{\sigma}^2$					
			Between Days		Between Runs		Within Run	
			Deg. of Freedom	Value	Deg. of Freedom	Value	Deg. of Freedom	Value
Marshall stability	Binder	I	20, 20	39,450[a]	-	0	42	54,200
		II	19, 19	22,400[b]	-	0	38	7,900
	Surface	I	18, 18	27,100[a]	-	0	40	49,800
		II	19, 19	18,600[b]	19, 40	2,250	40	3,500
Marshall density	Binder	I	20, 20	45×10^{-6}[a]	-	0	42	29×10^{-6}
		II	18, 18	43×10^{-6}[a]	-	0	40	47×10^{-6}
	Surface	I	19, 19	97×10^{-6}[a]	-	0	38	20×10^{-6}
		II	19, 19	112×10^{-6}[a]	19, 40	8×10^{-6}[a]	40	9×10^{-6}
Rice's max. density	Binder	I	20, 21	250×10^{-6}[a]	21	26×10^{-6}	-	-
		II	19, 20	94×10^{-6}[a]	20	37×10^{-6}	-	-
	Surface	I	18, 19	150×10^{-6}[a]	19	21×10^{-6}	-	-
		II	19, 20	144×10^{-6}[a]	20	11×10^{-6}	-	-

[a]Significant at 0.01 level.
[b]Significant at 0.05 level.

Similarly, for Marshall density:

1. The within-run variation values are similar to those found with the laboratory mixes, there being no statistical difference at the 0.05 level. This constitutes a confirmation of the results given earlier.
2. The estimates for the between-days components of supplier II are more than twice those obtained for supplier I. Considering the fact that the between-days variation in stability of supplier II appears smaller than supplier I, it might be possible (a) that greater variation in stability is not necessarily associated with density variation and (b) that the density variation is associated with the characteristics of the aggregates.
3. The between-days component is large. This might be attributed to the unavoidable grading, bitumen content, aggregate, and plant variations.

Similarly, for Rice's maximum density test:

1. The within-day components are close to the values reported earlier and do not differ statistically at the 0.05 level. This is a confirmation of the repeatability of the test.
2. The between-days variation constitutes the far more important variable. This is associated with the unavoidable day-to-day variations in bitumen content, grading, specific gravity, and process variables.
3. If repeatability is defined as the square root of σ^2_{WD} and the process variability as the square root of the sum of between-days and within-a-day components, the following values are obtained from the field investigation:

Supplier	Mix Type	Repeatability of Rice's Test, $\sigma = \sqrt{\hat{\sigma}^2_{WD}}$	Process Variability, $\sigma = \sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{WD}}$
I	Binder	0.0051	0.017
	Surface	0.0046	0.013
II	Binder	0.0061	0.0115
	Surface	0.0033	0.0125

DISCUSSION OF BITUMEN AND GRADING VARIATION

The present investigation offers an opportunity to discuss some effects of unavoidable bitumen content and grading variations.

Bitumen Content Variation

Previous investigation on bitumen content variation (15) indicates that for a well-controlled production, a total standard deviation, $\sigma_{total} = \sqrt{\sigma^2_{test} + \sigma^2_{process}}$, of 0.2% is normal. Assuming that the test component (σ_{test}) is in the order of 0.1 (1), the process component ($\sigma_{process}$) is then about 0.17. This implies that a natural spread ($\pm 2\ \sigma$) of $\pm$ 0.35% in bitumen content due to process variation is normal for a well-controlled production.

It is well known that a spread of $\pm$ 0.35% can considerably affect the physical properties of a given mixture. A typical example of this is shown in the following comparison of two binder mixes of 1-in. nominal size:

Producer	Average Marshall Stability		95% Confidence Interval for True Difference
	B-C 4.4%	B-C 5.2%	
E	2,030	1,870	160 ± 275
F	2,990	2,270	720 ± 195

The stability values given are the averages of five runs taken from five different batches. From this table it can be seen that by increasing the bitumen content from 4.4 to 5.2 (or 0.8%) the average stability values have decreased 160 lb for producer E and 720 lb for producer F. It should be noted that, based on previous data, the difference between results obtained from producers E and F is reliable within ± 380 lb (95% confidence level).

From the foregoing it may be concluded that, depending on the composition of the mix and the type of aggregate used, the unavoidable bitumen content variation can be an important component of the plant process variation.

Grading Variations

One of the most important factors in the production of bituminous mixtures is the variation in the aggregate gradation. It is well known that physical characteristics of bituminous mixtures are influenced by gradation variations.

In general, crushed aggregates are less variable than natural aggregates, the variations in grading of the crushed stone being related to the efficiency of crushing and screening operations, which in turn are related to the characteristics of the aggregate, rate of production, weather conditions, and many other factors.

The fine aggregates used in bituminous mixtures are usually screened or unscreened natural sand, which is more or less variable depending on its origin.

Grading specifications are generally stipulated for the coarse aggregate, the fine aggregate, and the combined aggregate. A typical example of this is the maximum permissible variation limits from job-mix formulas (Table 6), recommended by the Asphalt Institute for all types of mixes.

Test results from the AASHO Road Test and the present study (Table 7) indicate that limits such as those given in Table 6 are too narrow. It is believed that more realistic limits must take into account the relative importance of coarse and fine aggregate fractions.

The influence of grading variation on the physical properties of the mix is different for different types of mixtures. As a general rule, if the grading variation increases the density of the mix the stability will increase, the voids in mix will decrease, and the voids in mineral aggregates will also decrease.

Figure 4 shows typical results obtained from a well-controlled plant. It illustrates how the stability values varies with the unavoidable bitumen content and grading variations.

TABLE 6

ASPHALT INSTITUTE PERMISSIBLE VARIATION FROM JOB-MIX FORMULA

Sieve Size	Permissible Variation (% wt. of tot. mix.)
No. 4 +	5.0
No. 8	4.0
No. 30	3.0
No. 200	1.0

Maximum Size Effects

The statistical analysis of the present laboratory and field investigations offers an opportunity to discuss the maximum size effects.

Table 8 gives the between-days variations for field Marshall stability. It is interesting to note that for both supplier I and II, these components for between-days variation, which reflect the capability of a plant

TABLE 7

GRADING VARIATIONS: AASHO ROAD TEST AND PRESENT FIELD INVESTIGATION

Sieve	Asph. Inst. Permissible Variation (%)	AASHO Road Test				Present Field Investigation			
		Binder Mix		Surface Mix		Binder Mix		Surface Mix	
		Mean	σ	Mean	σ	Mean	σ	Mean	σ
1 in.	-	100	-	-	-	100	-	-	-
3/4 in.	-	96	2.21	100	-	96.6	1.83	-	-
1/2 in.	-	76	3.30	92	2.43	80.7	2.94	100	-
3/8 in.	-	57	2.71	81	3.17	61.0	2.48	99.0	1.32
No. 4	± 5	36	2.18	63	4.04	50.3	2.95	87.5	0.79
No. 8	-	-	-	-	-	44.6	2.73	75.5	1.85
No. 10	± 4	25	1.29	46	2.99	-	-	-	-
No. 16	-	-	-	-	-	40.2	2.44	68.0	1.83
No. 20	-	19	1.03	34	1.68	-	-	-	-
No. 30	± 3	-	-	-	-	28.7	2.07	53.0	2.01
No. 40	-	13	0.98	22	2.06	-	-	-	-
No. 50	-	-	-	-	-	18.3	1.38	41.4	2.13
No. 80	-	8	0.81	13	1.07	-	-	-	-
No. 100	-	-	-	-	-	9.7	1.11	25.7	1.85
No. 200	± 1	4.3	0.49	5.9	1.16	2.6	0.50	10.2	0.83
A. C.	± 0.3	4.2	0.13	5.2	0.18	4.93	0.33	6.35	0.22
Nb of test		127		96		40		40	

to produce a uniform mixture, appear relatively unaffected by the mix type.

Table 9 shows that for both the laboratory and field investigations (and for both suppliers I and II), the spreads of Marshall stability test results of the 1-in. nominal size binder mixes are nearly three times wider then those obtained with 1/2-in. nominal size surface mixes. Assuming that the only major difference between the surface and binder mixes is the nominal size of the aggregates, the tests results clearly indicate the maximum size effect on the repeatability of the Marshall tests.

TABLE 8

COMPARISON OF MARSHALL STABILITY BETWEEN-DAYS RESULTS

Supplier	Mix Type	$\hat{\sigma}^2$
I	Binder	39,450
	Surface	26,900
II	Binder	22,400
	Surface	18,600

TABLE 9

REPEATABILITY OF MARSHALL TEST

Supplier	Value of $\hat{\sigma}$			
	Surface Mixes (1/2-in. max.)		Binder Mixes (1-in. max.)	
	Lab. Invest.	Field Invest.	Lab. Invest.	Field Invest.
I	Not tested	90	Not tested	235
II	60	60	240	225

DISCUSSION OF TOTAL VOIDS IN MIX

In the preceding the components of variance were given for Marshall density and Rice's maximum density. Defining voids in the mix by

$$V_m = (1 - G_b/G_r)^{100} \qquad (2)$$

in which G_b is the average of two Marshall bulk densities and G_r is the Rice's maximum density (based on one test), the components of variance for voids in the mix of the field investigation are given in Table 10.

In this table it is interesting to note that the between-days component, which is related to the process variation, appears much higher than the within-day component, which reflects the testing precision.

The total standard deviation and normal spread shown in Table 11 clearly indicate that a spread of 3% in "voids in mix" can be expected from a well-controlled production if the testing is carried out as described earlier.

EFFECT OF INHERENT VARIABILITY ON MIX DESIGN AND SPECIFICATION

The repeatability of the Marshall (stability and density) and Rice's maximum density tests have been estimated, and the plant variation that occurs for a well-controlled production process has been analyzed in the foregoing. It is the prupose of this section to study some effects of these variations on (a) the significance of laboratory mix design formulas and (b) the specification limits.

TABLE 10

COMPONENTS OF VARIANCE FOR VOIDS IN MIX, FIELD TEST

Component[1]	Mix Type	Variance			
		Supplier I		Supplier II	
		Deg. of Freedom	Value	Deg. of Freedom	Value
$\hat{\sigma}^2_{BD}$	Binder	20, 21	0.2010	19, 20	0.3150
	Surface	18, 19	0.1400	19, 20	0.3200
$\hat{\sigma}^2_{WD}$	Binder	21	0.1050	20	0.0960
	Surface	19	0.0680	20	0.0310

[1] BD = between days (includes between-batch variation);
WD = within days.

TABLE 11

STANDARD DEVIATION AND NORMAL SPREAD FOR VOIDS IN MIX, FIELD TEST

Item	Mix Type	Supplier I	Supplier II
Standard deviation, $\hat{\sigma}_{tot} = \sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{WD}}$	Binder	0.55	0.63
	Surface	0.46	0.67
Total spread, $6\hat{\sigma}_{tot}$	Binder	3.3	3.8
	Surface	2.8	4.0

Significance of Laboratory Mix Design Formulas

The first step in mix design is to determine in the laboratory which combinations of aggregates and asphalt would give the required stability and durability. Usually, in determining the optimum asphalt content for a particular blend or gradation of aggregates by the Marshall method, a series of test specimens is prepared for a range

of different asphalt contents so that the test data curves show a well-defined optimum value. A single batch is prepared for each ½ percent increment of asphalt content, and triplicate test specimens are usually prepared from each batch. The next step is to go to the plant, calibrate the cold feed system, and find the optimum bin proportions that will produce a combined gradation conforming as closely as possible with the mix design formula. Once the job-mix formula is set, the characteristics of the produced mixtures must be kept within prescribed specification limits to assure uniformity of the mixture.

In an earlier section it was demonstrated that the most important component of variance in bituminous production is the day-to-day variation, especially for surface mix, where the repeatability of Marshall test is good. This day-to-day variation, which reflects the capability of a plant to produce a uniform mix, is usually unknown during the design stage. It follows that the mix design is only a rough estimate of the end results.

Figure 4 shows the scatter obtained when stability values are plotted against bitumen content for samples from a well-controlled production plant. The data shown come from a binder mix produced by supplier II. Sieve analyses of the samples indicate that all grading results fall within $\pm 3\sigma$ limits (the standard deviation being as given in Table 7, Col. "σ, Binder Mix, Present Field Investigation"). Examination of Figure 4 allows the following comments:

1. The individual stability values vary within a wide range. This is associated with unavoidable grading, bitumen content, plant, and testing variations.
2. The mix-design (laboratory) stability values obtained with 4.5 and 5.0 percent asphalt cement give a poor estimate of the production stability-asphalt content relationship. This is because in the mix design data only one laboratory batch is considered, whereas in routine production grading and other factors affecting stability vary from batch to batch.
3. Once estimates of process and testing components of variance have been obtained it may be possible, by using procedures described later herein, to determine the precision of inferences based on the laboratory design results.

Some further work may be necessary, however, to establish the connection between results obtained from laboratory-prepared mixtures and those from production batches. It is quite possible, for example, that the relationships between bitumen and stability indicated by the laboratory design tests may not be valid for other grading distributions lying within acceptable limits (see Fig. 4). In this case all that could be done with the laboratory results would be to estimate the range of production properties which should result for any mix design formula, given a specific supplier, a prescribed amount of routine testing, and a set of bitumen and grading limits.

Specification Limits

In setting realistic specification limits it is essential (a) to specify in detail the method of testing and the calculation procedures; (b) to ascertain limits for the properties, within which the material may be considered acceptable; (c) to consider the natural spread of the process under maximum control, taking into account the volume of sampling and testing to be performed; and (d) to give specific decision rules whereby material can be accepted or rejected, making provision for allowable sampling and testing fluctuations. The statistical analysis of test results from the AASHO Road Test and the present investigation offers an opportunity to underline the importance of items (a) and (b).

From Table 12, a summary of specification limits, field test results and assumed components of variance for both the AASHO Road Test and supplier II of the reported field investigation, it can be seen that the natural limits of the Marshall stability and total-voids-in-mix values cannot be directly compared.

If it is assumed that the standard deviation values for voids for both the present investigation and the AASHO Road Test are based on same components, $\sqrt{\sigma^2_{BD} + \sigma^2_{WD}}$, they cannot be directly compared, because the method of calculation differs. For the

▲ Routine production binder mix with a target job-mix bitumen content of 4.6%
• Routine production binder mix, same grading but a target bitumen content of 5.0%
o Original laboratory averages obtained for mix design
● Average point for production results

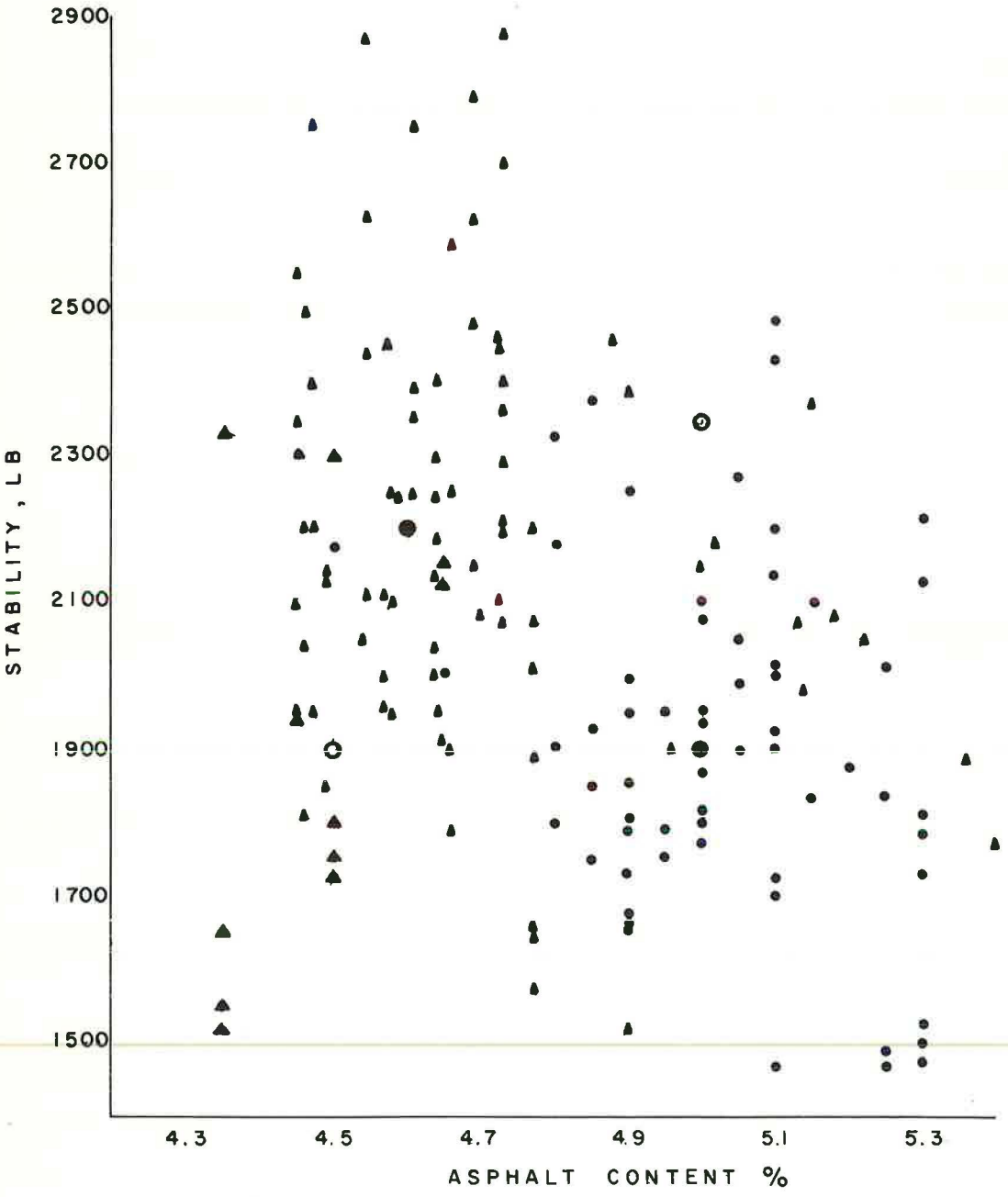


Figure 4. Comparison of routine laboratory design results with stability and asphalt content measurements from the controlled process.

TABLE 12

SUMMARY OF MARSHALL TEST RESULTS, PRESENT FIELD INVESTIGATION AND AASHO ROAD TEST

Investigation	Mix Type	Stability (lb)				Voids (% tot. vol.)			
		Mix Design Requir.	Mean Value	Std. Dev.	Components	Mix Design Requir.	Mean Value	Std. Dev.	Components
Present	Binder[1]	1,500[2]	1,919	305	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{BR} + \hat{\sigma}^2_{WR}}$	2 - 5	2.99	0.63	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{WD}}$
	Surface[1]	1,500[2]	1,796	185	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{BR} + \hat{\sigma}^2_{WR}}$	2 - 5	3.3	0.67	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{WD}}$
AASHO Road Test	Binder	1,500 - 2,500	1,770	190	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{BR} + (\hat{\sigma}^2_{WR})/2}$	4 - 6	4.8	0.52	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{WD}}$
	Surface	1,500 - 2,500	2,000	125	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{BR} + (\hat{\sigma}^2_{WR})/2}$	3 - 5	3.6	0.43	$\sqrt{\hat{\sigma}^2_{BD} + \hat{\sigma}^2_{WD}}$

[1]Supplier II. [2]Minimum.

present investigation, the voids in mix is defined by Eq. 2, in which G_r is the Rice's maximum density (based on one test). For the AASHO Road Test, V_m is defined by a similar equation in which G_r is replaced by G_a, the specific gravity of the mix calculated by using the "apparent" specific gravity of aggregates and the bulk specific gravity of the bitumen. It must be noted that if it is assumed that G_a is constant (as is frequently done), the standard deviation value will only reflect the Marshall bulk density variation. This shows that in setting specification limits it is essential to describe the test procedure and calculation method to be used.

The observed natural limits are related to the volume of sampling and testing. It is thus essential to decide if the limits are for single test results or for means of n results based on a specified number of runs. Examination of the standard deviation values given in Table 12, and their related components of variance, shows this.

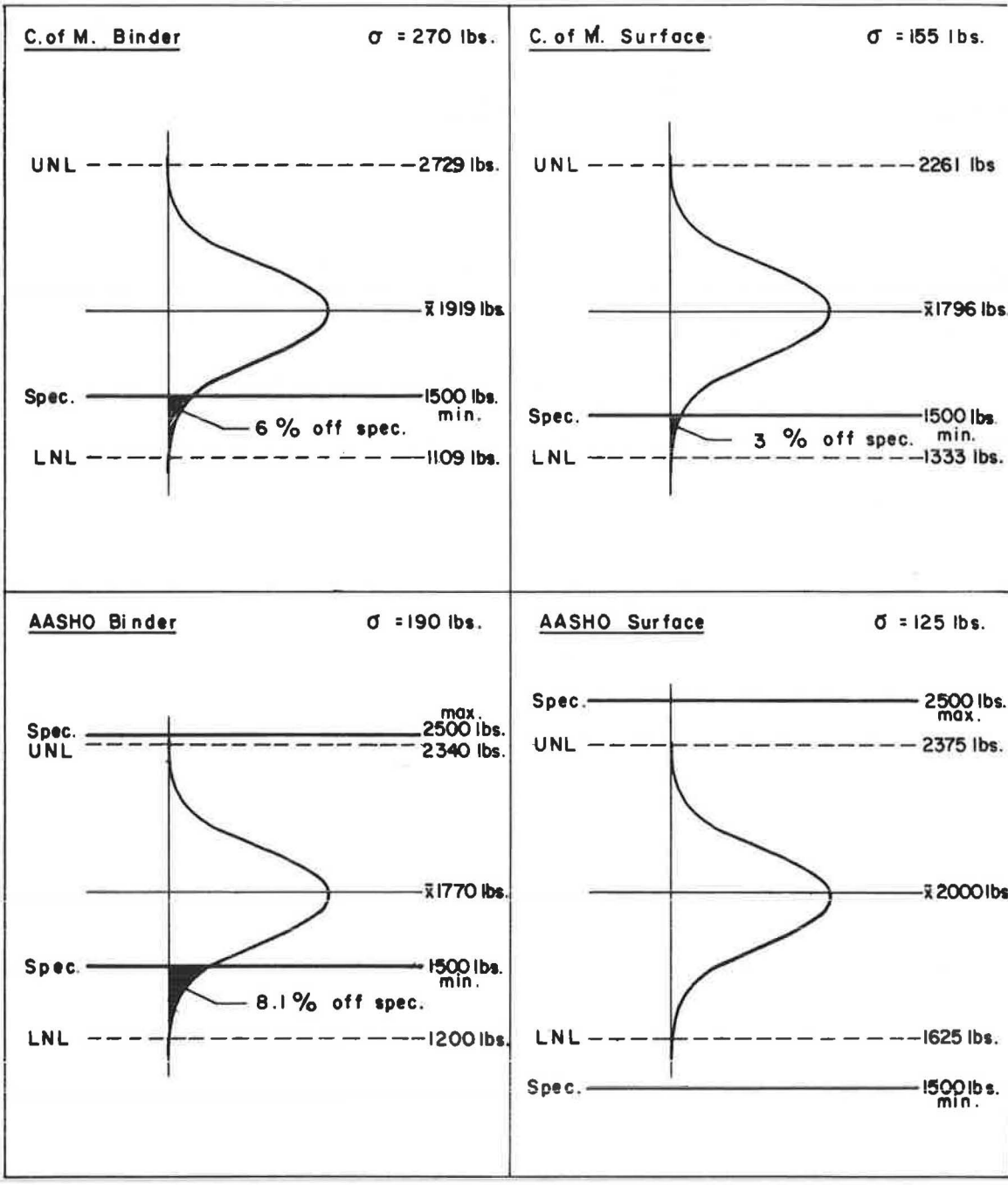
For the present field investigation, the specification requires that a minimum stability value of 1,500 lb must be met by any single test value (one briquet per batch).

This implied σ (for natural spread) $= \sqrt{\sigma^2_{BD} + \sigma^2_{BR} + \sigma^2_{WR}}$.

For the AASHO Road Test the specification requires that the average stability value of two briquets prepared from any truck sample must fall between 1,500 and 2,500 lb. Assuming that the sample is taken from a batch each day, σ (for natural spread) $= \sqrt{\sigma^2_{BD} + \sigma^2_{BR} + (\sigma^2_{WR})/2}$.

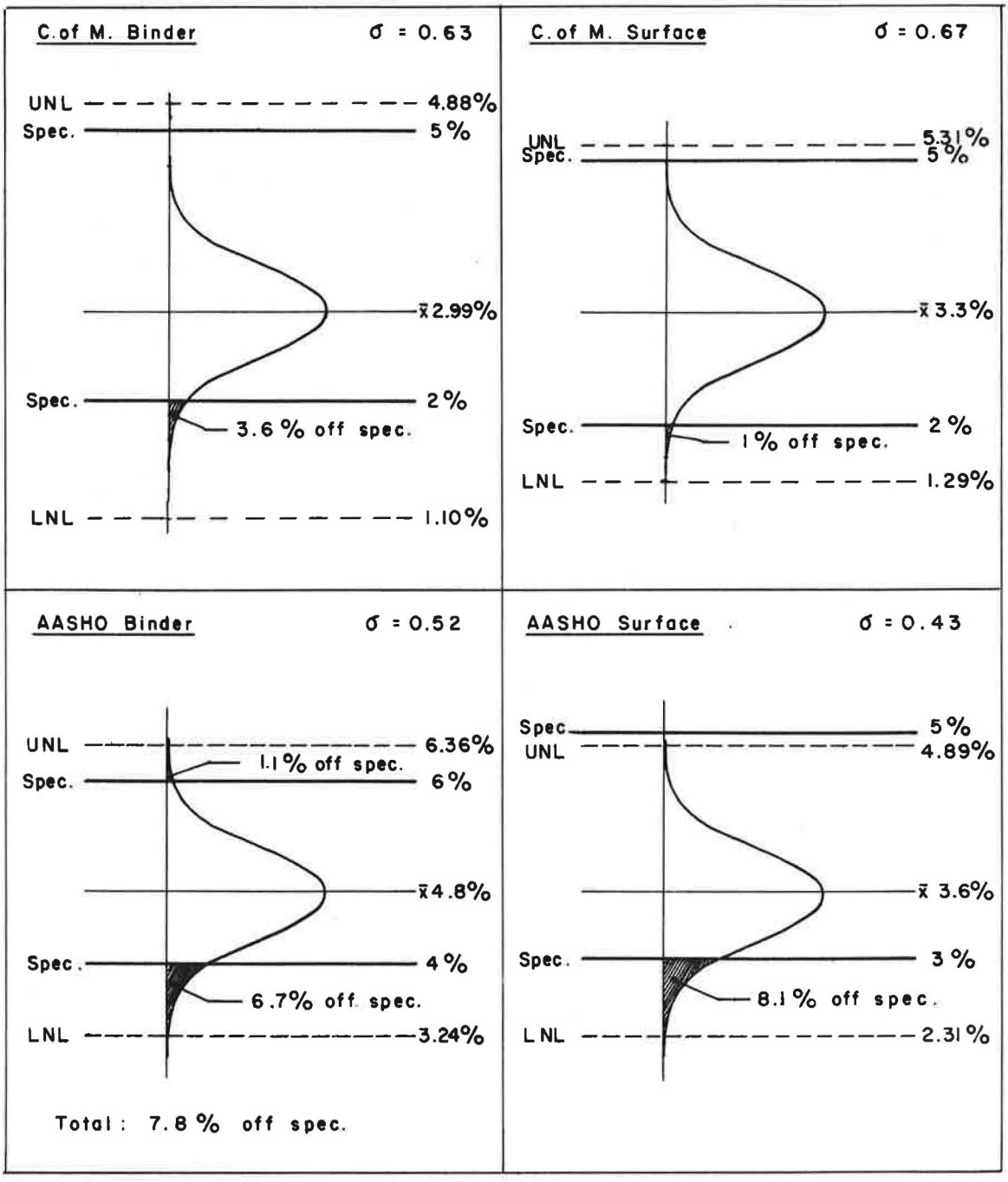
Figures 5 and 6 compare the specification limits and the natural spread of the AASHO Road Test data and the present field investigation test results. Examination of Figure 5 gives rise to the following comments:

1. In the cases of both surface and binder mixes of supplier II, as well as the binder mix of the AASHO Road Test, the specification limits do not, even though the two processes are in a "state of control," coincide with the natural limits.
2. In the case of supplier II, if the minimum specification requirement of 1,500 lb is to be observed, the process mean should be maintained 3σ above the minimum specification limit (that is, $\bar{x}$ = 2,300 lb for binder mix and $\bar{x}$ = 2,000 lb for surface mix, $\bar{x}$ being the process mean).
3. In the case of the AASHO test binder, where upper and lower specification limits are set, it can be seen that the process even when "in control" cannot supply 100 percent acceptable product in its present state. This is because the natural spread of the "in control" process (6σ = 1,140 lb) is larger than the range of the specification limits (2,500 - 1,500 = 1,000 lb). In such a case if it is desired to minimize off-specification material, the process mean should be kept at the specification mean (2,000 lb).
4. In the case of the AASHO surface mix, the natural spread lies within the specification limits. If a reduction in stability value reduces the cost of the mixture, and if the specification is to be observed, it is thus advantageous to keep the process mean 3σ (375 lb) above the lower specification limit.



UNL - UPPER NATURAL LIMIT
LNL - LOWER NATURAL LIMIT
σ - STANDARD DEVIATION
x̄ - MEAN VALUE

Figure 5. Process variation vs specification limits, marshall stability.



UNL - UPPER NATURAL LIMIT
LNL - LOWER NATURAL LIMIT
σ - STANDARD DEVIATION
x̄ - MEAN VALUE

Figure 6. Process variation vs specification limits, total voids in mix.

Figure 6 gives rise to the following comments:

1. Both City of Montreal and AASHO's limits (C of M, 2 to 5%; AASHO, 4 to 6% for binder and 3 to 5% for surface) lie inside the natural spread. This means that for the sampling, testing and calculation conditions described earlier the specification limits are too tight or some tolerance must be allowed to take care of the unavoidable off-specification portions.
2. In both cases the standard deviation values are close to 0.5 percent; in other words, in normal production a natural spread of at least 3 percent in voids is unavoidable. This implies that if it is desired to keep the production mixture within the specification limits, the minimum range of specification limits should be at least 3 percent and the process mean should fall in the center of this range.

CONTROLLING TEST PROCEDURE AND SUPPLIER'S PROCESS

Although the Marshall test (stability) is used for illustrative purpose, the following discussion could be applied in principle to other tests. Where a number of interrelated tests are used to decide the acceptability of a product, "multi-variate" control charts (4) may be used; however, only single-variate charts are described here. The discussion is divided into subsections dealing with (a) control within the laboratory, including establishing the capability of the testing procedure and maintaining control of the testing procedure; and (b) control over the supplier's material, including establishing the capability of the supplier's process and maintaining control over the supplier's process.

Control Within Laboratory

Establishing Capability of Testing Procedure. —Before meaningful sampling and testing frequencies can be established, it is necessary to establish the capability of the testing procedure and the "within-laboratory" factors which may influence the results. This capability is reflected by the "accuracy" of the test method, which may be defined as the extent to which test results may differ from the "true" or standard reference value. It should be noted that this definition includes the laboratory bias (should one exist), as well as the variability inherent in the procedure (21, 22).

The laboratory "bias" is that consistent difference separating the laboratory average either from the "true" value or from an acceptable reference value. This latter value is sometimes established by equating it to the grand average in a round robin program in which several laboratories carry out tests on similar material.

The variability of the testing procedure is best expressed in terms of components of variance determined by studies similar to that described earlier in this paper. Components of variance, as the name implies, are measures of the variability which may be ascribed to various sources, such as different operators, different pieces of test equipment, and even, in some cases, different stages of the test procedure itself. A more complete discussion of this subject is given elsewhere (29 - 35). Because of the nature of the calculations required, it is possible to build up information regarding these components through routine testing and small limited studies, as well as through more comprehensive programs such as those described here. Unfortunately, estimates of these components can be subject to relatively large variability themselves and those based on a few results are not reliable. The reliability may be increased, however, by combining several estimates of the same components. Where there are only a limited number of operators and testing machines in the laboratory, consistent differences in the average level of results, which may be ascribed to these factors, are often considered as biases, which are added or subtracted, as an adjustment, whenever the use of the data requires it.

Preliminary studies of test capability often disclose magnitudes of bias or variability which are unsatisfactory and which make it necessary to tighten the standard practices or refine the procedure.

Because an earlier section of this paper was concerned with the sources and magnitude of the variability of the test procedure and did not include a reference point for the determination of bias, the latter is not discussed more fully here, although its importance must not be discounted.

Use of Components of Variance.—Before the components of variance can be used to make probability statements about the precision of the test, it is necessary to ascertain that the individual results from specimens (briquets) taken from a sample tend to cluster about some single central value. (If this is not so, the test in its present state is meaningless.) Statistical theory can then be used to calculate limits about single test results, or averages of test results, such that the long-run average of the sample (or batch, if sampling components are included) will be bracketed by these limits at a given level of statistical confidence. For this estimate it is assumed that the level of control is maintained and that extraneous sources of variation are not introduced.

Several examples can now be given to demonstrate the foregoing.

Example 1: It was found earlier that for Marshall stability (binder material) $\hat{\sigma}^2_{BR} = 21,800$ lb, $\sigma^2_{WR} = 36,300$ lb, and the individual results approximated a normal distribution.

(a) Therefore, within a particular day it can be said with 95 percent confidence that a single test result will be within $\pm 2\sqrt{\sigma^2_{BR} + \sigma^2_{WR}} = \pm 480$ lb of the long-run average of the sample on that day.

(b) If n briquets (n-even) are made, tested on one day (two to a run), and the results averaged, it can be said with 95 percent confidence that the average will be within $\pm 2\sqrt{\frac{2\sigma^2_{BR} + \sigma^2_{WR}}{n}}$ lb of the long-run average of the sample on that day.

(c) Again, suppose it is necessary to compare two samples on the same day for stability. If n briquets are made and tested (two to a run) for each sample, the averages for the two samples would have to differ by at least $\pm 2\sqrt{2}\sqrt{\frac{2\sigma^2_{BR} + \sigma^2_{WR}}{n}}$ before a difference statistically significant at the 0.05 level could be claimed. (This difference would, however, not necessarily imply a difference between the parent mixes, as explained later.)

From this it can be seen that the limits of uncertainty about any quoted result or average can be reduced as required by increasing the number of briquets tested.

If there is assurance that the between-days testing component can be neglected, the limitation of single days in the preceding discussion can be removed. At present the data are insufficient to make a decision on this question.

Other statistical formulas are available which will enable the experimenter to calculate, in advance, the number of briquets to test and average for each sample, to meet pre-calculated risks for each of the two errors possible:

Error 1—No statistically significant difference is found, whereas in reality a difference of importance exists.

Error 2—A statistically significant difference is found, whereas in reality no difference exists.

Alternatively, a sequential sampling plan could be devised which would be more economical for the problem just described. The briquets would be made and tested in pairs (one from sample A and one from sample B). As the results were obtained a function of the accumulated differences would be plotted on a graph. At each stage the graph would indicate one of the following three courses of action:

1. Decide that a difference does exist.
2. Decide that a difference does not exist.
3. Test another pair of briquets.

Finally, a decision would be made which would be subject to the predetermined risks.

It should be noted that a difference between sample A and sample B does not necessarily imply a difference between batch A and batch B, because there may be (a) a

between-sample component of variance and/or (b) a between-batch component of variance (which is introduced when a number of batches are made according to the same formula). Either of these additional components could be sufficient to account for the difference found. That there is such a between-batch component was suggested by the highly significant between-days component for the surface mixture in the study described in the first section of this paper, inasmuch as this variation could conceivably be ascribed to the fact that a new batch was prepared each day. If comparisons between laboratory batches are carried out frequently it would be desirable to estimate more precisely the value of this between-batch component by preparing several batches a day, testing and repeating for a number of days according to a statistical design.

Once the magnitudes of the between-samples and between-batch components have been established, this information can be used to extend the argument of Example 1. For instance, suppose that the laboratory mixes were small enough that the term "sample" had no real meaning (compared with the sample obtained in the field from a production batch or truck load) and that the between-batch component, σ^2_{BB}, was 15,000 lb. The situation outlined in Example 1 (c) may now be generalized. Two binder formulas, A and B, are to be compared for stability, one batch being prepared for each. If n briquets are made and tested for each batch on the same day, the difference between the observed averages would have to exceed

$$\pm 2\sqrt{2}\sqrt{\sigma^2_{BB} + \frac{2\sigma^2_{BR} + \sigma^2_{WR}}{n}} \quad \text{or}$$

or

$$\pm 2\sqrt{2}\sqrt{15,000 + \frac{43,600 + 36,300}{n}}$$

before a difference statistically significant at the 0.05 level of probability could be claimed.

Maintaining Control of Testing Procedure. --Because of the tendency for equipment to wear, and of operators to relax their observance of standard instructions, it is necessary to carry out routine checking procedures.

Where control samples of known value are available, these may be introduced with predetermined frequency under the guise of a routine production sample. Control charts and other statistical procedures can be used to signal significant deviations from standard conditions.

Where such control samples are not available, shifts of the laboratory mean (due to a deterioration of the hammers or other causes) are more difficult to detect and only become apparent when a cross-check is made with another laboratory. It is possible, however, to maintain a weak measure of control on the variability of the testing procedure. This might be done as follows:

Where it is the practice to make two briquets for each production sample, compacting one on hammer A and the other on hammer B, the briquets would be numbered 1 and 2, 1 always being assigned to the briquet compacted on hammer A. If there is no difference between the hammers, the difference between the results (No. 1 - No. 2) should be distributed normally, with average 0 and variance $2\sigma^2_{WR}$ (assuming the individual test results are normally distributed). A regular Shewhart control chart (36) can then be prepared with center line 0, upper control limit $+ 3\sqrt{2}\,\sigma_{WR}$ and lower control limit $- 3\sqrt{2}\,\sigma_{WR}$. The ability of this chart to signal an increase of given magnitude in the variation or the bias can be calculated. It will be noted, however, that this method depends on the differences between duplicates, which are notoriously unreliable in providing a realistic measure of the overall testing variability.

Example 2: Suppose σ_{WR} = 190 lb (from binder study, first section). Figure 7 shows how the 32 differences plotted chronologically fall between the limits + 800 and - 800 lb. It can be calculated that the following conditions will generate an alarm signal (a single point outside the limits) after the specified length of time:

Condition	Average Number of Points Plotted Before One Falls Outside Limits
1. Normal	300
2. σ_{WR} increase by 25%	60
3. σ_{WR} increase by 50%	20
4. Hammers differ by a bias of 200 lb	80

The time interval can be decreased by adding rules for "runs" but this will also increase the number of erroneous signals arising from condition (1).

Control Over Supplier's Material

Establishing Capability of Supplier's Process. —As in the case of the test procedure itself, preliminary studies must be carried out on the supplier's process before the most efficient sampling patterns and frequencies can be determined. In particular, some attempt must be made to determine the variability within a processed batch, between batches within a day, and between days. This is not as difficult as it may appear and there is a good possibility that once these components of variance have been established and the process directed into a state of statistical control, not only will knowledge be increased with respect to possible trouble areas, but efficient sampling plans also can be set up which will be applicable for similar situations elsewhere.

Some laboratories make a practice of combining samples from several different batches to form a composite sample from which one or two briquets are prepared. The implications of this procedure should be studied thoroughly, because there must be some doubt about the combined effect of several mixes taking the size of the briquet into account.

In this preliminary study it may be found that the variability of the process, although predictable, is unsatisfactorily large for the user's purposes. In such cases the supplier will have to find measures to reduce it before his material can be considered acceptable.

After the components of variance have been established and the tolerances calculated within which the supplier's average will be permitted to vary, the risks for the control chart can be set and the sampling and testing frequency determined. The limits for the control chart are based on this preliminary work and so are determined before the routine testing commences.

In the absence of an analysis of variance study, statistical control can be established on a more empirical basis by the method used (9) where the control limits are set after a number of routine test results have been obtained. This method is valid, but in most applications it does not (a) provide the insight into the supplier's process which results from the more intensive preliminary study, or (b) permit optimum sampling and testing frequencies to be devised.

Maintaining Control Over Supplier's Process. —In normal manufacturing practice it is the supplier's responsibility to maintain the control charts, producing them on request for the customer's inspection. Where this is impractical and the customer is obliged to carry out some sampling and testing for his own protection, this should be specifically designed to provide, at minimum cost, the protection desired, with satisfactory risks from both the customer's and the supplier's viewpoints.

The crux of the argument for statistical control is that protection can be achieved most economically by (a) establishing that the process is predictable and (b) providing measures (the control chart) to indicate when this assumption is no longer valid or when the process has shifted into an unsatisfactory region. The alternative is to have no objective assessment of the nature of the process variations and either (a) proceed on an intuitive basis of acceptance and rejection (which is arbitrary and unsatisfactory on scientific grounds) or (b) try to maintain protection by large-scale sampling and testing.

Shewhart Control Chart. —Where the process has shown itself to be relatively stable and in a state of statistical control, the Shewhart control chart may be used to provide

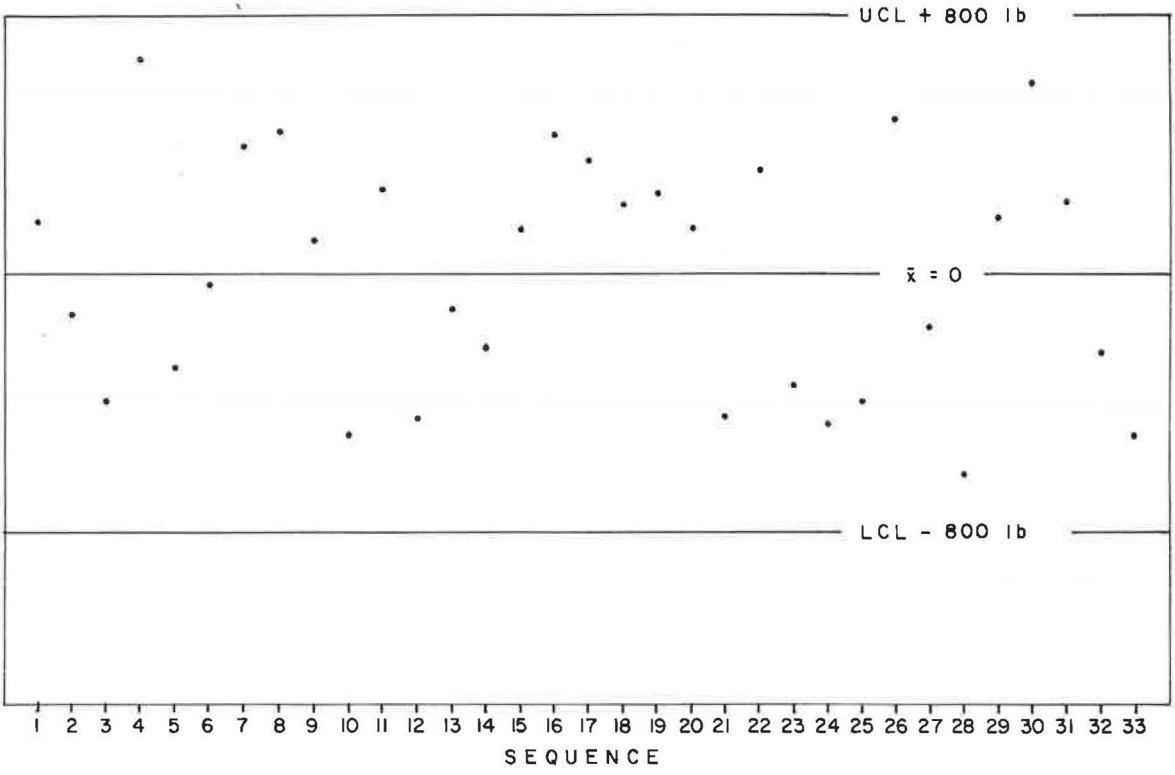


Figure 7. Construction of a control chart for within-laboratory variation.

assurance that the mean of the supplier's process has not shifted (as a result of a different source for the aggregate, say) to an unsatisfactory level, or that the variability of the material has not changed. Little assurance can be provided under this system that unsatisfactory batches, or even unsatisfactory groups of batches occurring sporadically, will be detected.

Before the sampling plan can be set down, two points must be considered: (a) How soon after the shift is it necessary to have a signal that the supplier's material is unsatisfactory? and (b) What protection is the supplier to have against the situation where he is told to adjust a process (or accept a penalty) when the process in reality is providing acceptable material?

Once these have been agreed to, the sampling frequency can be calculated and the chart limits set. The reasoning and mathematical procedures involved are best illustrated by an example.

Example 3: Suppose the major consideration for Marshall stability for a particular binder material is that no batch average falls below 1,500 lb. Therefore, any batch with a true average Marshall stability value less than 1,500 lb will be called unacceptable.

The suppliers process is "in control" and running at an average level of 1,900 lb. It has been established that $\sigma^2_{\text{between batches}} = 20,000$ lb, $\sigma^2_{\text{between samples}} = 20,000$ lb, $\sigma^2_{\text{between runs}} = 21,800$ lb, and $\sigma^2_{\text{within run}} = 36,300$ lb (with negligible between-days testing variation).

Consider the distribution of batch averages under this system. Assuming a normal

pattern this is depicted in Figure 8, from which it is evident that the process average is on the borderline for acceptability because any downward shift would generate unacceptable batches.

The control chart for daily averages would be constructed with center line 1,900 and limits ± 3

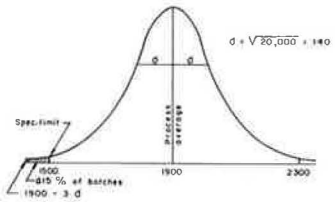


Figure 8.

$$\sqrt{\frac{\sigma^2_{BB}}{n_1} + \frac{\sigma^2_{BS}}{n_1 n_2} + \frac{2\sigma^2_{BR} + \sigma^2_{WR}}{n_1 n_2 n_3}}$$

assuming the product $n_1 n_2 n_3$ is even, n_1 being the number of batches sampled each day, n_2 the number of samples taken from each batch, and n_3 the number of briquets made and tested from each sample.

It is assumed that samples are kept separate (that is, there are no multi-batch composite samples).

The selection of values for $n_1 n_2$ and n_3 will depend on the following factors:

1. The cost of each sampling stage.
2. The cost of testing.
3. The level of protection required.
4. Other factors of practicability.

Suppose that it was decided for protection purposes:

1. That when the proportion of unacceptable batches reached 25 percent there would be a 50 percent chance of an "out of control" signal (that is, it is unlikely (6 percent chance) that more than 4 days will pass before a signal is generated).
2. That when the process average remained at the 1,900-lb level there would be only a 0.15 percent chance of signal below the lower limit.
3. That signals above the upper limit would not be cause for action, but that close attention would be paid to the chart for the succeeding days to determine the extent of the process shift.

Using statistical theory it can be shown that the following relationship must be satisfied:

$$3\sigma_{avg} = 3\sqrt{\frac{\sigma^2_{BB}}{n_1} + \frac{\sigma^2_{BS}}{n_1 n_2} + 2\,\frac{\sigma^2_{BR} + \sigma^2_{WR}}{n_1 n_2 n_3}} = 300 \text{ lb}$$

or that

$$\frac{20,000}{n_1} + \frac{20,000}{n_1 n_2} + \frac{79,900}{n_1 n_2 n_3} = \left(\frac{300}{3}\right)^2 = 10,000 \text{ lb.}$$

Therefore, the sampling and testing procedure to be employed will depend on the selection of n_1, n_2 and n_3 such that these values satisfy the foregoing relationship. There are many combinations possible and the aforementioned factors of cost and convenience will determine the final choice.

One set of n's which does satisfy the equation approximately is: $n_1 = 6$, $n_2 = 1$, $n_3 = 4$. In other words, it is necessary to take one sample from each six batches selected randomly, preparing and testing four briquets from each sample. The average for these twelve briquets would be calculated, and plotted on a control chart with center line 1,900 lb and limits set at $1,900 \pm 3\sigma_{ave}$ (i.e., 1,900 ± 300 lb). When an average falls below 1,600 lb (the lower limit) there is a definite indication that the supplier's process has shifted to an unsatisfactory level.

Although this amount of sampling and testing may seem excessive, it is the minimum amount which will provide the required protection, within the limitations of the

Shewhart control chart, and process and testing variability. Only by relaxing the protection desired can this volume of testing be reduced. As mentioned earlier, the practice of making composite samples might lead to some reduction in the work required but the statistical and practical implications would have to be thoroughly examined before a definite conclusion could be made.

A "range" control chart also could be set up to provide some control on the variability of the process, but this will not be described further because the same philosophical concepts apply.

Before leaving this topic, a new type of control chart, the cumulative sum control chart (41, 44) can be briefly discussed.

This form of chart has been used for the last year at the Montreal Research and Control Laboratory and has proven valuable. Its major characteristics are that relatively small shifts of the average become apparent early and it is possible to estimate quickly the size of the shift as well as the date on which the shift occurred. Control limits may be calculated as described in the literature but this was not done, the advantages given being enough to justify use of the charts.

Essentially, the chart is the accumulated total of differences obtained by subtracting the daily average from a fixed value which approximates the process average. Each day this difference (with sign) is added to that of the day before (see Table 13). Shifts in the average show up as changes in the direction of the line, the angle of the line giving a measure of the process average.

TABLE 13

EXAMPLE OF CUMULATIVE SUM CHART CALCULATIONS

Day	$\bar{X}_1$, Avg. of 2 Briquets	$\bar{\bar{X}} - \bar{X}_1$, Difference from 1,900	Cumulative Sum of Differences
1	1,560	- 340	- 340
2	1,850	- 50	- 390
3	1,815	- 85	- 475
4	1,815	- 85	- 560
5	2,085	185	- 375
6	1,700	- 200	- 575
7	2,275	375	- 200
8	1,930	30	- 170
9	2,075	175	5
10	2,150	250	255
11	2,025	125	380
12	2,200	300	680
13	2,100	200	880
14	1,850	- 50	830
15	1,900	0	830
16	2,000	100	930
17	1,585	- 315	615
18	2,050	150	765
19	1,575	- 325	440
20	1,750	- 150	290
21	1,700	- 200	90
22	2,010	110	200
23	1,600	- 300	- 100
24	-	-	-
25	-	-	-
26	-	-	-
-	-	-	-
-	-	-	-

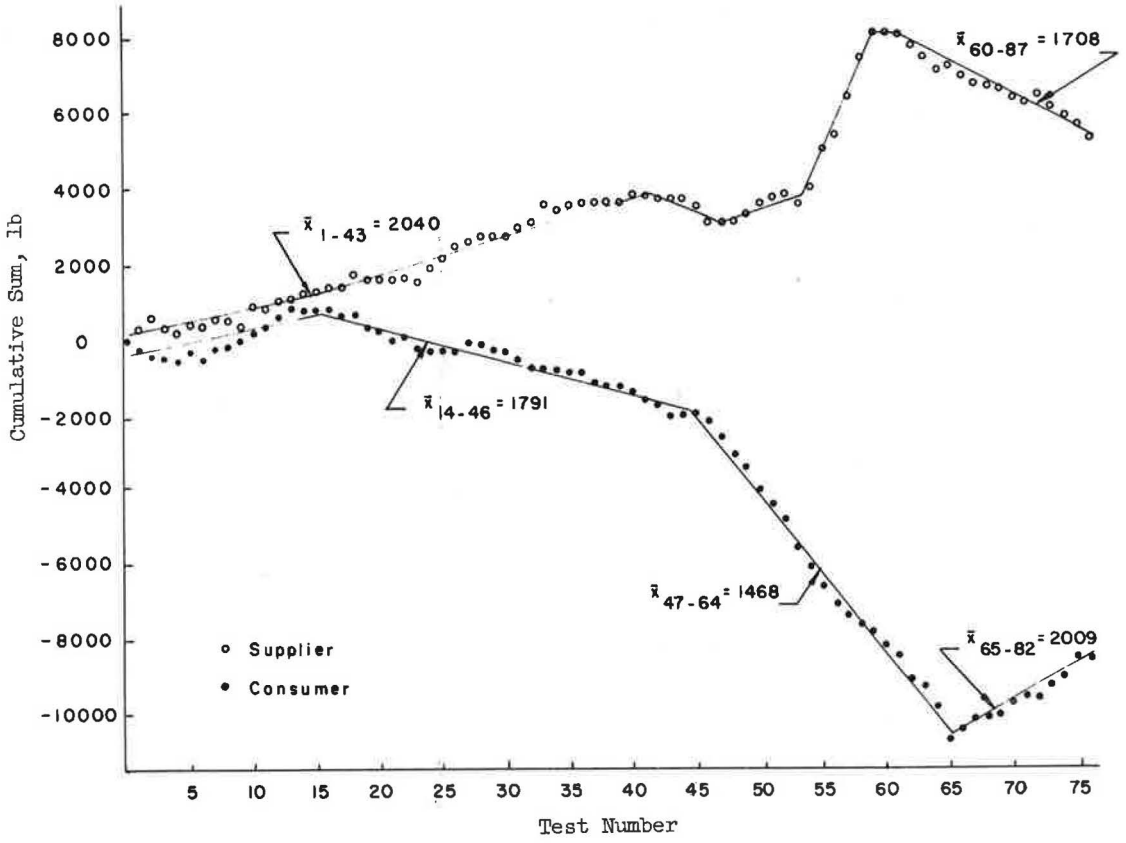


Figure 9. Cumulative sum chart.

Example 4: The process average for a supplier was found to be approximately 1,900 lb. A composite 20-lb sample was prepared from each day's production. Two briquets were made and tested from each composite sample. Calculations were carried out as outlined in Table 12 and the results plotted as in Figure 9.

The cumulative sum chart clearly indicates the following changes:

1. Sequence 1 to 14—Same slope or mean for both the supplier (2,040 lb) and consumer (1,970 lb) test results. (During this period both the supplier's and the consumer's equipment was in good order).
2. Sequence 15 to 46—Same positive slope for the supplier, but a negative slope for the consumer. (During this period, owing to the fact that the process, sampling and testing were unchanged, the change in slope (1,970 to 1,791) was attributed to defective testing equipment. It should be noted that at the time, the City of Montreal Control and Research Laboratory used Shewhart control chart alone. No action was taken because no anomalies appeared and the stability routine test values remained above 1,500 lb, the required minimum value.)
3. Sequence 47 to 64—Action period. (During this period the consumer's stability value dropped to 1,468 level. The Shewhart control chart signaled that the process was out of control. The supplier was asked to take action and a third laboratory was called in to check the apparatus. Finally, it was found that the consumer's compaction hammers were out of order. It is interesting to see, by the several changes in slope, how the supplier actually tried to correct the situation.)
4. Sequence 65 to 82—Finally the consumer corrected the situation by using two new

hammers, the average value (2,009) increased to the original level (2,040). Unfortunately, however, the contractor's equipment was defective (1,708).

Since that time (1960) the City of Montreal has used the cumulative sum chart as part of its control procedure.

REFERENCES AND SELECTED BIBLIOGRAPHY

Interlaboratory Testing

1. "Cooperative Tests on Bitumen Content of Paving Mixtures." Report of Committee D-4, ASTM Proc. 50:316 (1950).
2. Interlaboratory Testing of Rubber and Rubber-Like Materials. ASTM Designation: D-1421-56.
3. Interlaboratory Testing of Textile Materials. ASTM Designation: D-990-58.
4. Youden, W. J., "Graphical Diagnosis of Interlaboratory Test Results." Industrial Quality Control, p. 24 (May 1959).
5. Mandel, J., and Lashof, T. W., "The Interlaboratory Evaluation of Testing Methods." ASTM Bulletin, p. 53 (July 1959).
6. Pierson, R. H., and Fay, E. A., "Guidelines for Interlaboratory Testing Programs." Analytical Chemistry, p. 25A (December 1959).
7. ASTM Manual for Conducting an Interlaboratory Study of a Test Method. Spec. Tech. Publ. 335.

Previous Experiments

8. Odasz, F. B., Jr., and Nafus, D. R., "Statistical Quality Control Applied to an Asphalt Mixing Plant." AAPT Proc. (Feb. 1954).
9. Shook, J. F., "Quality Control of Bituminous Concrete Production." AAPT Proc. (Jan. 1960).
10. "The AASHO Road Test: Report 2, Materials and Construction." HRB Special Report 61B (1962).
11. Vokac, Roland, "Repeatability of Marshall Test by Analysis of Factorial Experiment Data." AAPT Proc. (Jan. 1962).
12. Corbett, L. W., "Variation Studies in Density and Stability Measurements on Asphalt Aggregate Mixes." AAPT Proc., Vol. 25 (1956).

Method of Tests

13. Rice, James, "Maximum Specific Gravity of Bituminous Mixtures by Vacuum Saturation Procedure." ASTM Spec. Tech. Publ. 191.
14. "Mix Design Methods for Hot-Mix Asphalt Paving." The Asphalt Institute, Manual Series No. 2 (Oct. 1960).
15. Keyser, J. Hode, and Gaudette, N. G., "The Significance of Variation in Asphalt Content of Paving Mixtures." Papers on Road and Paving and Symposium on Microviscometry, ASTM (1961).
16. Resistance to Plastic Flow of Bituminous Mixtures Using Marshall Apparatus. ASTM Designation: D-1559-60T, Standards (1961).

Multi-Variate Control Charts

17. Jackson, J. E., "Quality Control Methods for Two Related Variables." Industrial Quality Control, p. 4 (Jan. 1956).
18. Jackson, J. E., and Morris, R. H., "An Application of Multivariate Quality Control to Photographic Proceeding." Jour. of Amer. Stat. Assoc., p. 186 (June 1957).
19. Weis, P. E., "An Application of a Two-Way X-Bar Chart." Industrial Quality Control, p. 23 (Dec. 1957).
20. Jackson, J. E., "Quality Control Methods for Several Related Variables." Technometrics, p. 359 (Nov. 1959).

Statistical Evaluation of Test Methods

21. Wesnimont, G., "Precision and Accuracy of Test Methods." ASTM Symposium on Applications of Statistics (No. 103), p. 13.
22. Use of the Terms Precision and Accuracy as Applied to Measurement of a Property of a Material. ASTM Designation: E-000-61T.
23. Grubbs, F. E., "On Estimating Precision of Measuring Instruments and Product Variability." Jour. of Amer. Stat. Assoc., p. 242 (June 1948).
24. Mandel, J., "Statistical Principles of Testing." Industrial Quality Control, p. 18, (Nov. 1955).
25. Henry, J. A., "Instrument and Measurement Evaluation." 14th Midwest ASQC Conf. Trans., p. 69 (1959).
26. Davies, O. L., "Some Statistical Aspects of the Economics of Analytical Testing." Technometrics, p. 49 (Feb. 1959).
27. Mandel, J., "The Measuring Process." Technometrics, p. 251 (Aug. 1959).
28. "Quality of Observations—A Series of Four Papers." Materials Research and Standard, p. 264 (April 1961).

Components of Variance

29. Moroney, M. J., "Facts from Figures." Chapt. 19, Pelican Books.
30. Davies, O. L., Ed., "The Design and Analysis of Industrial Experiments." Chapt. 4, Hafner Pub. Co.
31. Hicks, C.R., "Fundamentals of Analysis of Variance: Part II, The Components of Variance Model." Industrial Quality Control, p. 6 (Sept. 1956).
32. Youden, W. J., "Test Program on Steel." Industrial and Engineering Chem., p. 60A (Aug. 1956).
33. "Use of Sub-Samples in the Control Laboratory." Analytical Chem., p. 1046, (July 1957).
34. "Analysis of Components of Variance." Industrial and Engineering Chem., p. 1017 (July 1958).
35. Davies, O. L., "Some Statistical Aspects of the Economics of Analytical Testing." Technometrics, p. 63 (Feb. 1959).

Control Charts

36. Burr, I. W., "Engineering Statistics and Quality Control." McGraw-Hill.
37. Reynolds, J. H., "Controlling the Control Laboratory." Industrial Quality Control, p. 12 (July 1954).
38. Izzarone, A. L., "Insuring Measurement Reliability." Production, p. 95, (June 1957).
39. McCutchen, R. L., "Use of Control Charts in Analytical Chemistry." Middle Atlantic Conf. Trans. ASQC, p. 109 (1958).
40. Freund, R. A., "Graphical Process Control." Industrial Quality Control, p. 15 (Jan. 1962).

Cumulative Sum Control Charts

41. Goldsmith, P. L., and Whitfield, H., "Average Run Lengths in Cumulative Chart Quality Control Schemes." Technometrics, 3:11 (Feb. 1961).
42. Page, E. S., "Cumulative Sum Charts." Technometrics, 3:1 (1961).
43. Truax, H. M., "Cumulative Sum Charts and Their Application to the Chemical Industry." Industrial Quality Control, p. 18 (Dec. 1961).
44. Johnson, N. L., and Leone, F. C., "Cumulative Sum Control Charts." Industrial Quality Control (June, July, Aug. 1962).

Appendix

TYPICAL DATA COMPILATIONS AND ANALYSES OF VARIANCE

TABLE 14

MARSHALL TEST DATA FOR SURFACE MIX FROM SUPPLIER II, DATA INVESTIGATION

Week	Day	Stone Filled Hammer	Stone Filled Order	Stability	Flow	Density
1	1	C	1	1,575	16	2.374
		D	2	1,650	19	2.371
		C	3	1,780	16	2.377
		D	4	1,655	15.5	2.369
	2	C	1	1,715	14.5	2.373
		D	2	1,615	15.5	2.380
		C	3	1,575	17	2.375
		D	4	1,725	16.5	2.378
	3	C	1	1,350	15	2.360
		D	2	1,425	16	2.365
		C	3	1,550	15.5	2.364
		D	4	1,510	16.5	2.366
	4	C	1	1,475	14	2.382
		D	2	1,475	14	2.380
		C	3	1,490	14	2.378
		D	4	1,475	15	2.371
Total				25,040	244.0	37.963
2	1	C	1	1,500	15	2.378
		D	2	1,450	15	2.376
		C	3	1,460	16	2.378
		D	4	1,575	16	2.373
	2	C	1	1,600	17	2.371
		D	2	1,555	15	2.370
		C	3	1,710	16.5	2.371
		D	4	1,750	16	2.370
	3	C	1	1,410	15.5	2.364
		D	2	1,450	15	2.373
		C	3	1,550	14.5	2.364
		D	4	1,600	15	2.365
	4	C	1	1,425	15.5	2.374
		D	2	1,410	15.5	2.365
		C	3	1,645	17	2.379
		D	4	1,575	17	2.373
Total				24,665	251.5	37.944
3	1	C	1	1,650	17	2.377
		D	2	1,460	14.5	2.370
		C	3	1,600	14.5	2.383
		D	4	1,700	15	2.374
	2	C	1	1,595	14	2.367
		D	2	1,535	14.5	2.365
		C	3	1,695	14	2.362
		D	4	1,510	13.5	2.357
	3	C	1	1,600	15.5	2.368
		D	2	1,695	14	2.371
		C	3	1,710	15.5	2.375
		D	4	1,675	16	2.373
	4	C	1	1,600	14.5	2.356
		D	2	1,590	14.5	2.370
		C	3	1,590	15	2.362
		D	4	1,575	16	2.371
Total				25,780	238.0	37.901
4	1	C	1	1,625	14.5	2.365
		D	2	1,500	13.5	2.357
		C	3	1,660	13.5	2.373
		D	4	1,700	14.5	2.362
	2	C	1	1,530	15.5	2.364
		D	2	1,575	14	2.370
		C	3	1,575	13	2.371
		D	4	1,595	12.5	2.368
	3	C	1	1,710	13.5	2.368
		D	2	1,560	13.0	2.362
		C	3	1,840	13.5	2.369
		D	4	1,690	12.5	2.365
	4	C	1	1,555	14.5	2.363
		D	2	1,475	14.5	2.362
		C	3	1,660	13.5	2.366
		D	4	1,655	13	2.354
Total				25,995	219.0	37.339

TABLE 15

ANALYSIS OF VARIANCE[a] FOR MARSHALL STABILITY AND DENSITY, SURFACE MIX, SUPPLIER II, LABORATORY INVESTIGATION

Source of Variation	Degrees of Freedom	Components of Variance	Marshall Stability[c]			Marshall Density[d]		
		Est. by Mean Square[b]	Sum of Squares	Mean Square	F Test	Sum of Squares ($\times 10^{-6}$)	Mean Square ($\times 10^{-6}$)	F Test
Between weeks	3	$16\sigma^2_{BR} + 4\sigma^2_{BD} + 2\sigma^2_{BR} + \sigma^2_{WR}$	72,791	24,264	-	567	189	$F_{(3, 12)}$ 1.8[e]
Between days	12	$4\sigma^2_{BD} + 2\sigma^2_{BR} + \sigma^2_{WR}$	292,798	24,400	$F_{(12, 15)}$ 5.4[f]	1,239	103	$F_{(12, 47)}$ 5.1[f]
Between runs:								
First vs second	1	$32\Sigma f^2_k + 2\sigma^2_{BR} + \sigma^2_{WR}$	108,076	108,076	$F_{(1, 15)}$ 23.7[f]	9	9	$F_{(1, 15)}$[e]
Within day	15	$2\sigma^2_{BR} + \sigma^2_{WR}$	62,281	4,552	$F_{(15, 31)}$ 1.3[e]	259	17	$F_{(15, 31)}$[e]
Between hammer positions	1	$32\Sigma f_l^2 + \sigma^2_{WR}$	4,391	4,391	$F_{(1, 31)}$ 1.3[e]	47	47	$F_{(1, 31)}$ 2.35[e]
Within run	31	σ^2_{WR}	107,238	3,459	-	613	20	-
Total	63		653,575			2,734		

[a]Model: $Y_{ijkl} = \mu + a_i + a_{ij} + a_{ijk} + f_k + f_l + a_{ijkl}$.

[b]BW = between weeks; BD = between days; BR = between runs; WR = within run.

[c]Components of variance: $\hat{\sigma}^2_{WR} = 3,460$, $\hat{\sigma}^2_{BD} = 4,960$, $\hat{\sigma}^2_{BR} = 550$ (NS), $\sigma^2_{BW} = 0$.

[d]Components of variance: $\hat{\sigma}^2_{WR} = 0.000020$, $\hat{\sigma}^2_{BD} = 0.000021$, $\hat{\sigma}^2_{BR} = 0.0$, $\hat{\sigma}^2_{BW} = 0.000005$.

[e]Not significant at 0.05 level.

[f]Significant at 0.01 level.

TABLE 16

ANALYSIS OF VARIANCE[a], FOR RICE'S MAXIMUM DENSITY, SURFACE MIX SUPPLIER II, LABORATORY INVESTIGATION

Source of Variation	Degrees of Freedom	Sum of Squares ($\times 10^{-6}$)	Mean Square ($\times 10^{-6}$)	Components of Variance Est. by Mean Square[b, c]	F Test
Between weeks	3	206	69	$8\sigma^2_{BW} + 2\sigma^2_{BD} + \sigma^2_{WD}$	$F_{(3,\ 12)}$[d]
Between days, within a week	12	727	61	$2\sigma^2_{BD} + \sigma^2_{WD}$	$F_{(12,\ 16)}$ 6.8[e]
Within day	16	137	9	σ^2_{WD}	
Total	31	1,070			

[a]Model: $Y_{ijk} = \mu + a_i + a_{ij} + f_j + a_{ijk}$.

[b]Components of variance: $\hat{\sigma}^2_{WD} = 0.00009$, $\hat{\sigma}^2_{BD} = 0.000026$, $\hat{\sigma}^2_{BW} = 0.000001$.

[c]BW = between weeks; BD = between days; WD = within day.

[d]Not significant at 0.05 level.

[e]Significant at 0.01 level.

TABLE 17

STABILITY, FLOW, AND DENSITY DATA OF FIELD SAMPLES, SURFACE MIX, SUPPLIER I

Sequence Day	Marshall Stability				Marshall Flow				Marshall Bulk Density				Rice's Maximum Density	
	Run I		Run II		Run I		Run II		Run I		Run II		Run I	Run II
	X1	X2	X3	X4	X1	X2	X3	X4	X1	X2	X3	X4	X1	X2
1	1700	1700	1750	1850	11.0	10.0	10.0	13.0	2.378	2.397	2.380	2.382	2.453	2.454
2	1715	1875	1900	1815	13.5	15.0	16.0	16.0	2.355	2.360	2.361	2.359	2.434	2.431
3	1685	1625	1585	1700	16.0	15.0	17.0	17.0	2.367	2.366	2.361	2.377	2.423	2.419
4	1550	1510	1650	1555	20.0	19.0	16.0	17.0	2.366	2.355	2.356	2.358	2.414	2.426
5	1605	1575	1590	1525	17.0	17.0	18.0	18.0	2.364	2.365	2.361	2.360	2.419	2.422
6	1400	1450	1500	1490	19.0	22.0	20.0	20.0	2.361	2.364	2.366	2.361	2.401	2.407
7	1490	1660	1810	1780	17.0	17.0	16.0	17.0	2.373	2.371	2.374	2.359	2.415	2.429
8	2005	1830	2040	1900	16.0	14.0	16.0	15.0	2.370	2.374	2.375	2.375	2.450	2.440
9	1600	1535	1645	1490	19.0	15.0	16.0	17.0	2.362	2.365	2.361	2.367	2.417	2.421
10	1925	1950	2005	1900	12.0	12.0	11.0	12.0	2.368	2.369	2.372	2.370	2.447	2.441
11	1460	1600	1475	1450	16.0	16.0	16.0	16.0	2.370	2.373	2.366	2.371	2.416	2.416
12	1435	1285	1265	1485	15.0	17.0	16.0	16.0	2.372	2.368	2.370	2.369	2.402	2.411
13	1445	1550	1470	1470	15.0	13.0	14.0	15.0	2.356	2.357	2.351	2.349	2.410	2.410
14	1470	1550	1550	1635	16.0	15.0	16.0	15.0	2.366	2.371	2.364	2.368	2.421	2.422
15	1610	1460	1445	1535	15.0	15.0	17.0	15.0	2.372	2.361	2.358	2.362	2.416	2.423
16	1710	1910	2010	1970	17.5	17.5	16.0	17.5	2.364	2.379	2.378	2.376	2.424	2.424
17	1850	1520	1635	1710	15.0	17.0	15.0	14.0	2.372	2.365	2.361	2.363	2.433	2.428
18	1685	1710	1710	1970	15.0	15.0	16.0	16.0	2.369	2.371	2.375	2.377	2.424	2.425
19	1850	1760	1785	1635	15.0	15.0	17.0	16.0	2.372	2.370	2.373	2.371	2.433	2.424

TABLE 18

ANALYSIS OF VARIANCE[a] FOR MARSHALL STABILITY AND DENSITY, SURFACE MIX, SUPPLIER I, FIELD INVESTIGATION

Source of Variation	Degrees of Freedom	Marshall Stability[c]				Marshall Density[d]		
		Components of Variance Est. by Mean Square[b]	Sum of Squares	Mean Square	F Test	Sum of Squares ($\times 10^{-6}$)	Mean Square ($\times 10^{-6}$)	F Test
Between days	18	$4\sigma^2_{BD} + 2\sigma^2_{BR} + \sigma^2_{WR}$	2,076,566	115,364	$F_{(18, 18)}$ 16.5[e]	3,463	192	$F_{(18, 18)}$ 10.7[e]
Between runs: First vs second	1	$38\Sigma f^2_j + 2\sigma^2_{BR} + \sigma^2_{WR}$	27,284	27,284	$F_{(1, 18)}$ 3.9[f]	22	22	$F_{(1, 18)}$ 1.2[f]
Within day	18	$2\sigma^2_{BR} + \sigma^2_{WR}$	125,629	6,979	$F_{(18, 38)}$[f]	317	18	$F_{(18, 38)}$[f]
Within run	38	σ^2_{WR}	299,625	7,884		820	22	
Total	75		2,529,104			4,622		

[a]Model: $Y_{ijk} = \mu + a_i + a_{ij} + f_j + a_{ijk}$.

[b]BD = between days; BR = between runs; WR = within run.

[c]Components of variance: $\hat{\sigma}^2_{WR} = 7,900$, $\hat{\sigma}^2_{BR} = 0$, $\hat{\sigma}^2_{BD} = 27,100$.

[d]Components of variance: $\hat{\sigma}^2_{WR} = 0.000022$, $\hat{\sigma}^2_{BR} = 0$, $\hat{\sigma}^2_{BD} = 0.000043$.

[e]Significant at 0.01 level.

[f]Not significant at 0.05 level.

TABLE 19

ANALYSIS OF VARIANCE[a] FOR RICE'S MAXIMUM DENSITY, SURFACE MIX, SUPPLIER I, FIELD INVESTIGATION

Source of Variation	Degrees of Freedom	Sum of Squares ($\times 10^{-6}$)	Mean Square ($\times 10^{-6}$)	Components of Variance Est. by Mean Square[b, c]	F Test
Between days	18	5,806	323	$2\sigma^2_{BD} + \sigma^2_{WD}$	$F_{(18, 19)}$ 15.4[d]
Within day	19	401	21	σ^2_{WD}	
Total	37	6,207			

[a]Model: $Y_{ij} = \mu + a_i + a_{ij}$.

[b]Components of variance: $\hat{\sigma}^2_{WD} = 0.000021$, $\hat{\sigma}^2_{BD} = 0.000150$.

[c]BD = between days; WD = within day.

[d]Significant at 0.01 level.