REPORT 76-005-7776

ENSURING RELIABILITY OF MAYS ROUGHNESS MEASUREMENTS

A FINAL REPORT

MARCH, 1976

BY

JACK CROTEAU

PROJECT ENGINEER, RESEARCH

The New Jersey Department Of Transportation

Division Of Research And Development

In Cooperation With

The U.S. Department Of Transportation

Federal Highway Administration

NOTICE

This publication is disseminated in the interest of information exchange.

The opinions, findings, and conclusions expressed in the publication are those of the author and not necessarily those of the New Jersey Department of Transportation or the Federal Highway Administration.

This report does not constitute a standard, specification, or regulation.

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-NJ-RD-76-005		

4. Title and Subtitle	5. Report Date
Ensuring Reliability of Mays Roughness Measurements	March 1976
	6. Performing Organization Code
7. Author(s) Jack Croteau	8. Performing Organization Report No. 76-005-7776
9. Performing Organization Name and Address New Jersey Department of Transportation Division of Research and Development 1035 Parkway Avenue Trenton, New Jersey 08625	10. Work Unit No. 11. Contract or Grant No. N. J. HPR Study #7776
12. Sponsoring Agency Name and Address New Jersey Department of Transportation Division of Research and Development 1035 Parkway Avenue Trenton, New Jersey 08625	13. Type of Report and Period Covered A Final Report 14. Sponsoring Agency Code

15. Supplementary Notes The first of two reports on the study "Pavement Roughness Evaluations using a Mays Ride Meter". Prepared in Cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract

A measurement device known as the Mays Ride Meter is extensively used to quickly and inexpensively determine relative pavement roughness. Mays results are most commonly used in various rating systems for establishing pavement rehabilitation priorities. This report presents a set of procedures for systematically monitoring the output of the New Jersey Mays device so as to determine if the equipment is functioning properly and providing reliable measurements.

The described procedures for detecting and acting on Mays output problems are based on the use of two types of control charts: mean charts for discovering shifts in average roughness between test dates and range charts for determining increased variability of the testing process. The expected effectiveness of the procedures is gauged by means of power curves.

Mays output is shown to be significantly influenced by the ambient temperature at the time of test. As a consequence, a temperature correction (3 inches/mile per 10°F differential) must be applied to New Jersey Mays data to ensure comparability of measurements made on different occasions.

Appendix A describes in detail the nature and relative magnitude of the various individual components of Mays measurement variability. The results of this analysis of variance are applied by example to routine Mays decision-making situations faced by operations personnel.

17. Key Words	18. Distribution Statement
Mays Ride Meter, Pavement Roughness, Measurement Reliability, Control Charts, Analysis of Variance	Copies Available on Request

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	46	

Form DOT F 1700.7 (8-69)

PREFACE

The treatment of the subject matter of this report is oriented toward the needs of the operations (Maintenance) personnel who will in fact be required to apply the measurement control procedures described herein. This reflects in the following assumptions by the writer:

- Most readers have a working knowledge of routine Mays testing operations.

- Some readers will not be completely conversant with the fundamentals of statistical quality control.

Accordingly, in developing instructions for measurement control, the writer has attempted to bring out the underlying statistical principles in very basic fashion.

IMPLEMENTATION STATEMENT

Adoption of the Mays measurement control procedures described in this report will provide for more effective, informed judgments regarding the relative riding qualities of New Jersey pavements.

Implementation should be quite straightforward since Operations personnel have worked directly with Research during the study.

TABLE OF CONTENTS

PART ONE: INTRODUCTION

PART TWO: THE DEVELOPMENT OF NEW JERSEY MAYS ROUGHNESS CONTROL CHARTS

PART THREE: CONTROL OPERATIONS

APPENDICES

LIST OF EXHIBITS

PART ONE: INTRODUCTION

1.1 Objective of the Work:

In the summer of 1974, the New Jersey Department of Transportation acquired a high-speed roughness measurement device known as the "Mays Ride Meter". This equipment was purchased to provide an input to a pavement rating system for ordering maintenance resurfacing priorities, as well as to provide a continuing capability for gauging the riding quality of newly completed construction.

The objective of this report is to provide a set of procedures for systematically monitoring the output of the Mays device so as to determine if a stable, accurate level of readings is being obtained.

1.2 Nature of the Mays Equipment:

The Mays Ride Meter -- models of which are reportedly in use by at least 30 agencies -- basically consists of an instrumentation package mounted in a standard passenger car which measures road smoothness in terms of the relative movement between the car body and differential. The New Jersey Mays vehicle, a 1974 Ford Custom 500 sedan, is operated at a standard speed of 40 MPH during testing.

The heart of the Mays measurement system is a small strip chart recorder. The paper chart output of this instrument records the distance travelled in a test, the cumulative amount of body movement or roughness, and a roughness trace or profile. Given the total inches of vertical movement and the length travelled, Mays roughness results can be expressed in units similar to roughness index. That is, as the Mays roughness in inches per mile. (In this state, all Mays results are rounded to the nearest integer.)

1.3 <u>Overview of the Problem and Solution:</u>

As is the case with test data in general, repeat Mays roughness measurements made on a given pavement will display variations both within and between groups of tests. The actual magnitude of variation observed within a group of short-term Mays data basically depends upon the inherent capabilities (precision) of the equipment, its condition at the time of test, and operator technique. Long-term variations in Mays results are a function of additional variations both in temperature and in pavement characteristics (i.e., actual roughness changes) between test groups.

This report presents a set of procedures for ensuring that the Mays device is functioning properly and providing reliable measurements.

Very simply, these procedures involve comparing current Mays results for five, one-mile pavement test sections of known roughness to past data for these same sites. Given the proper equipment calibration and operating procedure, the current measurements will reproduce the past data within certain measurement tolerances and the Mays testing process will be considered "in control". Conversely, when current measurements are not consistent with the established standards for the test sites, it follows that either (a) the Mays calibration or testing procedure is improper -- "out-of-control" -- or, less commonly, (b) the roughness standards require revision (i.e., there has been a real change in test site roughness).

The statistical device used to demonstrate whether or not the Mays testing process is in control is known as a "control chart". A Mays control chart is simply a time plot of test site roughness data on which are superimposed lines ("control limits") that depict the historical measurement tolerance computed for that site. If the plotted points for successive test samples fall within these control limits, a decision is made that the Mays testing process is in control. If the points fall outside the control limits, the process is judged to be out of control and action is undertaken to recalibrate the device or improve testing technique.

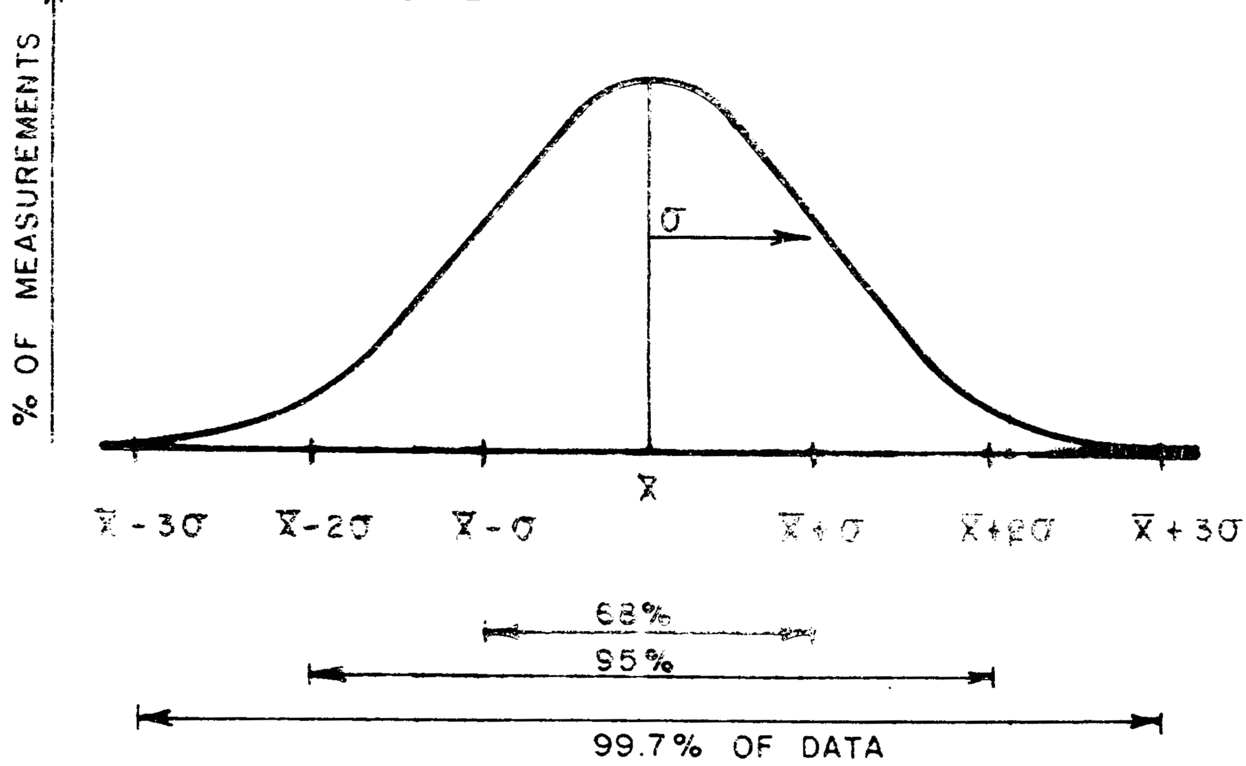
Detection of the two basic types of Mays output problems require construction of separate control charts: a mean ($\overline{X}$) chart for discovering shifts (+ or -) in average roughness between test dates and a range (R) chart for determining increased variability of the testing process. While all five sites need not be tested on every occasion, this method of controlling Mays roughness output may occasionally require the analysis of as many as 10 control charts.

PART TWO: THE DEVELOPMENT OF NEW JERSEY MAYS ROUGHNESS CONTROL CHARTS

2.1 Measurement Tolerances:

From the preceding discussion, it is apparent that the successful development of process control charts hinges on determining appropriate measurement tolerances or control limits. In the particular case of Mays measurements, determination of these control limits depends on the fact that roughness results for a particular pavement are "normally" distributed. That is, a plot of the frequency with which particular levels of roughness occur in repeat tests will display the familiar "bell" shape shown in Figure 1.

FIGURE 1: TYPICAL DISTRIBUTION OF MAYS ROUGHNESS RESULTS FOR A CONTROL SITE

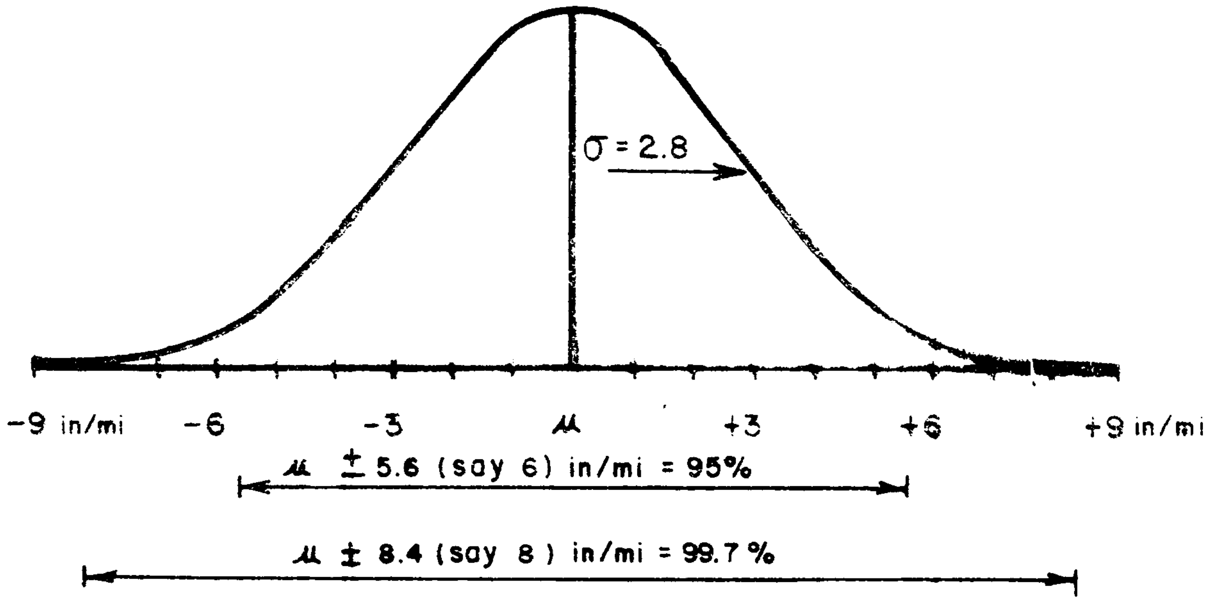


Given two properties of a particular normal curve -- the mean, ($\bar{X}$) and standard deviation (σ) -- it is possible to formulate some characteristic measurement tolerances. For example, as shown in Figure 1, about 2/3 of all normally distributed test observations will have a magnitude which falls within the interval $\bar{X} \pm \sigma$, about 95 percent will lie within $\bar{X} \pm 2\sigma$, and essentially all of the observations (99.7 percent) will be within $\bar{X} \pm 3\sigma$.

The two end numbers , $\bar{X} + 2\sigma$ and $\bar{X} - 2\sigma$, between which 95 percent of all measured values are expected to occur, are called the "two sigma" limits or (more properly) the "95 percent confidence limits". By definition then, the likelihood of obtaining a measured value outside the 95 percent confidence limits is at most 5 percent when the process is in control. Similarly, the end numbers of the intervals $\bar{X} \pm \sigma$ and $\bar{X} \pm 3\sigma$ are respectively referred to as the 68 percent ("one sigma") and 99.7 percent ("three sigma") confidence limits.

As will be discussed later, the basic New Jersey roughness control sample to be evaluated on a test day is five (5) consecutive runs per individual test site. If the Mays testing process is in control, the average of the five runs should be within $\pm$ 3 times some standard deviation of the true mean roughness (μ) for the particular site. The analysis of considerable historical data for the New Jersey control sites indicates that for such a five-run average, the standard deviation is typically on the order of 2.[illegible] inches/mile. The standard deviation in this instance is representative of the total day-to-day variation in average measured roughness and indicates the measurement tolerance to be reflected in the control limits.* For example, as shown in Figure 2, essentially all daily averages of five readings should be within about 8 inches/mile (3σ) of the true site mean when the process is in control. Further, nearly all should be within about 6 inches/mile of the true site mean (i.e., within the 95% confidence limits).

FIGURE 2: TYPICAL DISTRIBUTION OF DAY-TO-DAY TEST SITE ROUGHNESS AVERAGES (SAMPLE N=5)



*The standard deviations representing total day-to-day measurement variation on particular roughness control sites range from 2.2 to 3.2 inches/mile. These values are actually used in constructing control charts for averages on specific sites.

2.2 General Features of Control Charts:

Once the frequency distribution of test results from a controlled process is known, as from the preceding section, the basic construction of control charts can be easily understood.

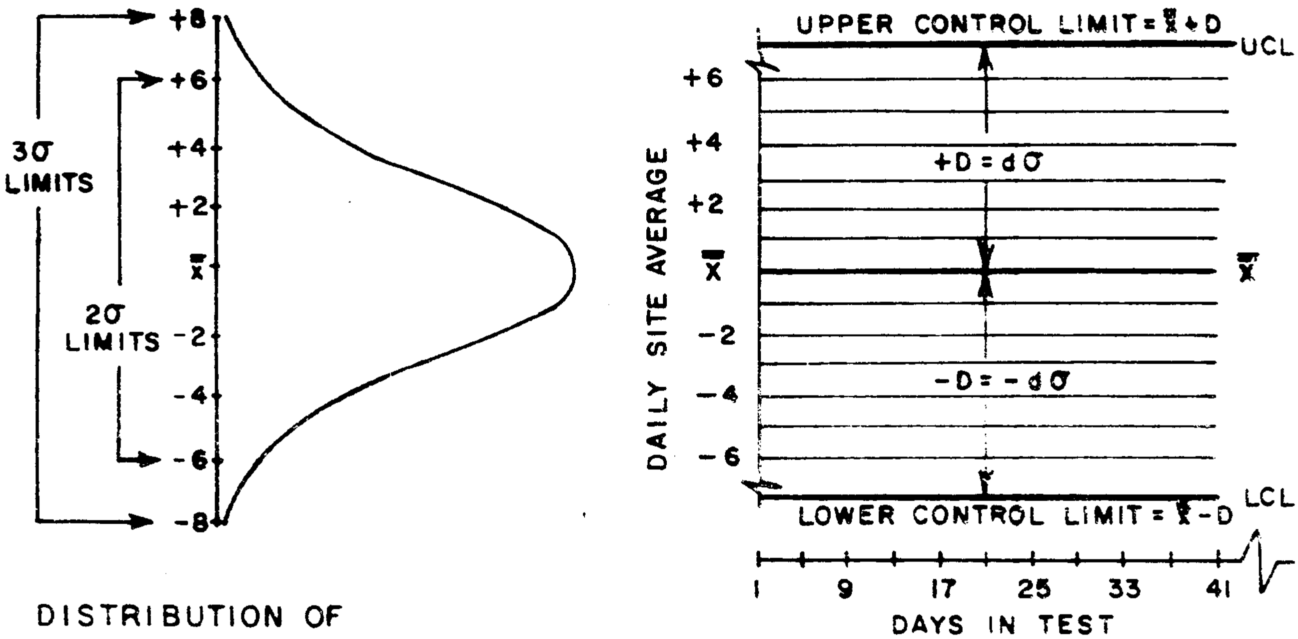
In essence, a control chart is simply a normal frequency distribution to which has been added a time scale. The difference between the "bell" curve and a control chart then is that the order or sequence of readings is shown in the latter. As will be described later (and, indeed, as might be guessed), retaining the time sequence of readings is important in that relatively subtle patterns or trends in data output can be equally as good an indicator of an out-of-control process as dramatic shifts in readings.

Figure 3 illustrates the general features of a control chart; specifically, a chart for controlling averages. The chart is based on the typical distribution of day-to-day roughness averages previously shown in Figure 2.

Three lines on the chart are of interest: the central line and the two control limits. The central line is the overall or "grand average" ($\bar{\bar{X}}$) of all historical readings obtained on a site and represents the best estimate of the "true" site mean (μ). Expressed in general terms, the upper control limit (UCL) and lower control limit (LCL) are located equidistant some number of Mays units ("D") above and below the site mean ($\bar{\bar{X}}$) Since the number "D" can be expressed as some multiple ("d") of the standard deviation, it should be apparent from the discussion presented earlier in this report that control chart limits are in fact confidence limits. That is, the control limits are the end numbers of the interval $\bar{\bar{X}} \pm d\sigma$ which includes some given percentage of all expected readings, the specific percentage depending on the magnitude of the multiplier "d".

FIGURE 3

GENERAL FEATURES OF A CONTROL CHART



DISTRIBUTION OF PROCESS RESULTS (DAILY AVERAGES)

CHART FOR CONTROLLING PROCESS (CHART FOR AVERAGES)

The selection of "d" -- in effect, the sizing of the control tolerances -- is based on a number of factors. One of the key factors is the frequency with which one is willing to have the control limits exceeded by mere chance. In other words, how often are we willing to respond to "false alarms", requiring recalibration when in fact no output problem exists. For example, suppose that the 95 percent confidence limits are chosen as the control chart limits (i.e, d = 2). In the case of a chart for Mays averages, this would mean that the upper and lower control chart limits would typically be set at $\bar{\bar{X}} + 6$ and $\bar{\bar{X}} - 6$. Since these limits by definition encompass all but 5 percent of the expected readings, if a decision to recalibrate is made based on the occurrence of a value in excess of these limits, the maximum chance that the level of readings has not changed and that the search is in fact for nonexistent trouble is 5 percent. Stated another way, recalibrating at this level involves a risk of being wrong in 5 out of 100 cases. Similarly, use of the three sigma (99.7 percent) control limits as an indicator of trouble can involve a recalibration "fool's errand" in at most 3 out of 1,000 cases.

In actual practice, the choice of specific control chart limits involves striking a balance between excessive leniency and over-restrictiveness. That is, if the control limits are widely separated so as to prevent any unnecessary recalibrations (i.e., if the multiplier d is relatively large), the level of control exercised over the process may be almost meaningless. On the other hand, if the control limits are tightly [illegible] in process output (i.e., if d is relatively small), the number of false indications of a need for recalibration could be [illegible]. The best choice of control chart limits is thus that which minimizes unnecessary recalibrations but which maximizes the detection of <u>significant</u> shifts in process output. A discussion of what constitutes a "significant shift" to be detected by the New Jersey Mays control procedure must await the discussion of control chart decision rules presented in subsection 2.3.

The discussion thus far has been limited to the development of a control chart for the <u>average</u> of five successive roughness measurements. Since the <u>range</u> (difference between the largest and smallest) of these five measurements also serves as an indicator of the adequacy of the roughness testing process, control charts for the range will also be used to evaluate New Jersey Mays output. However, range charts are discussed in this report only in terms of their form and method of application rather than in terms of their detailed development. Suffice it to say that range charts are based on within - hour rather than day-to-day measurement variation and, as a consequence, they have a smaller associated standard deviation than that of charts for averages. The reader interested in the magnitude of the various components of overall Mays measurement variability is directed to the discussion presented in Appendix A.

2.3 <u>Selection of Specific Control Limits and Decision Rules</u>

Control charts fundamentally have one purpose: to assist in decision making. The three data outcomes possible in process control testing and the associated decisions/actions are:

Nature of Data	Decision	Action
1. Data **definitely** within tolerance	Accept the process	Continue routine Mays testing
2. Data **definitely** out of tolerance	Reject the process	Recalibrate; improve technique
3. Data **"suspicious"**	Withhold judgment	Perform additional control tests

Since there are two possible cases of process rejection -- a definite rejection when the Mays output is absolutely out of control and a conditional rejection when the **readings are merely suspicious** -- it seems apparent that the control procedures should reflect this fact. That is, there is a need for an "action" control limit indicating a definite need for remedial action and a (lesser) "warning" control limit indicating a zone of suspect data.

Another aspect of the decision-making process which must be considered in formulating rules for the use of control charts is that it is possible for Mays output to be judged out of control on different time bases. That is, a long-term pattern of consistently high (or low) test site readings may be equally as symptomatic of an out-of-control process as short-term fluctuations, even if these long-term departures are of relatively low magnitude. Thus a complete set of control chart decision rules will include a rule for isolating and acting on any trends shown in successive days testing.

The specific control chart limits and associated decision rules to be employed in New Jersey are as follows:

2.3.1 "Action" control limits for a single day's testing:

The 99.7 percent confidence limits are to be used on both types of charts (averages and ranges) as an indicator of the definite need for recalibration or improved testing technique. The format in which these confidence limits are applied differs between the two types of charts, however, in that while charts for averages possess upper and lower boundaries, a range chart has only an upper boundary. That is, a shift in range requires action only if it represents a decrease in repeatability (increase in range).

The specific "action" control limits for the five New Jersey control sites are shown in Table 1A. As an illustration of the application of these limits, a recalibration would definitely be undertaken if on a given day the average of five successive readings on the Route 29 test site was 108 inches/mile (or more) or 94 inches/mile (or less).

2.3.2 "Warning" control limits for a single day's testing:

New Jersey Mays control site readings are to be considered "highly suspect" if they equal or exceed the 95 percent (two sigma) limits shown in Table 1A. In that event, a second sample of five readings is to be obtained from the site(s) in question. If the statistic in doubt (i.e., range or mean) again falls outside the 95 percent "warning" limit, the suspicion that the process is out of control is considered confirmed and remedial action is to be undertaken. If the results of the second sample fall within the "warning" limits, the initial suspicion is considered refuted.

It is important to note that if a second control test sample is required, the results of the retest are to be computed and judged independently. That is, the decision whether to recalibrate or not is based on the procedure described above and not on some consideration of the "overall range" or "grand average" of the combined samples.

TABLE I: NEW JERSEY MAYS CONTROL CHART LIMITS

IA: LIMITING OR CRITICAL VALUES FOR A <u>SINGLE</u> DAY'S TESTING

TEST SITE	CHART FOR AVERAGES: CENTRAL LINE ($\bar{\bar{X}}$)	CHART FOR AVERAGES: "WARNING" (2σ) LIMITS: UPPER	CHART FOR AVERAGES: "WARNING" (2σ) LIMITS: LOWER	CHART FOR AVERAGES: "ACTION"* (3σ) LIMITS: UPPER	CHART FOR AVERAGES: "ACTION"* (3σ) LIMITS: LOWER	CHART FOR RANGES: CENTRAL LINE ($\bar{R}$)	CHART FOR RANGES: "WARNING" LIMIT (UPPER ONLY)	CHART FOR RANGES: "ACTION" LIMIT* (UPPER ONLY)
NJ 29	101	106	96	108	94	5	8	10
I-95 (SOUTH)	83	88	78	91	75	5	8	10
Lakehurst (WEST)	107	114	100	117	97	6	10	12
Lakehurst (EAST)	115	121	109	123	107	5	8	10
I-295	32	39	25	42	22	6	10	12

*RECALIBRATE WHEN <u>AT</u> OR <u>OUTSIDE</u> UPPER OR LOWER LIMIT

IB CRITICAL VALUES FOR <u>SUCCESSIVE</u> DAYS TESTING

TEST SITE	TAKE ACTION IF $\bar{X} \geq$ UCL OR $\leq$ LCL FOR EACH OF 3 SUCCESSIVE TESTS: UCL = $\bar{\bar{X}}+3$	TAKE ACTION IF $\bar{X} \geq$ UCL OR $\leq$ LCL FOR EACH OF 3 SUCCESSIVE TESTS: LCL = $\bar{\bar{X}}-3$	TAKE ACTION IF $R \geq \bar{R}+2$ FOR EACH OF 3 SUCCESSIVE TESTS: $\bar{R}+2$
NJ 29	104	98	7
I-95 (SOUTH)	86	60	7
LAKEHURST (WEST)	110	104	8
LAKEHURST (EAST)	118	112	7
I-295	35	29	8

2.3.3 <u>"Action" control limits for successive days testing:</u>

A shift in process output will be considered to have been established if the Mays readings for any control site display a trend wherein three (3) successive* day's tests consistently exceed the 68 percent (one sigma) confidence limits shown in Table 1B.

As indicated in the table, the "one sigma" limits are set equal to $\bar{\bar{X}}$ + 3 inches/mile and $\bar{\bar{X}}$ - 3 inches/mile in the case of charts for averages and $\bar{R}$ + 2 for a range chart. For example then, a problem would be indicated if three successive days testing on the Route 29 site yielded daily averages which were <u>each</u> at least 104 inches/mile or which were <u>all</u> 98 inches/mile or less.

While a recalibration or improvement in testing technique may be the appropriate remedial action when successive tests exceed the "one sigma" criteria, an alternate possibility which should not be overlooked is that it is the control site standards which in fact require attention. That is, if a chart for averages displays a trend of increasing roughness, it may well be that the roughness of the control site has in fact increased. If a case such as this occurs, operations personnel should contact Research for assistance in determining whether the apparent output problem is really equipment/technique related. In any event, control chart data should be periodically reviewed by Research (say at least once a year) to determine if any updating is required.

*"Successive" does not necessarily imply consecutive calendar days.

[illegible] Actual Format of New Jersey Control Charts:

The actual control charts to be used in New Jersey are [illegible] to this report (Appendix [illegible]). Historical roughness data collected [illegible] each of the five sites is plotted on the charts for purposes of [illegible].

[illegible] Effectiveness of New Jersey Control [illegible]

Having described New Jersey's Mays control charts and associated [illegible] rules, it is necessary to address an [illegible] raised earlier, namely, "What constitutes a [illegible] or significant shift in process output to be detected by these procedures?" Simply stated, "How effective or powerful are the described control procedures?"

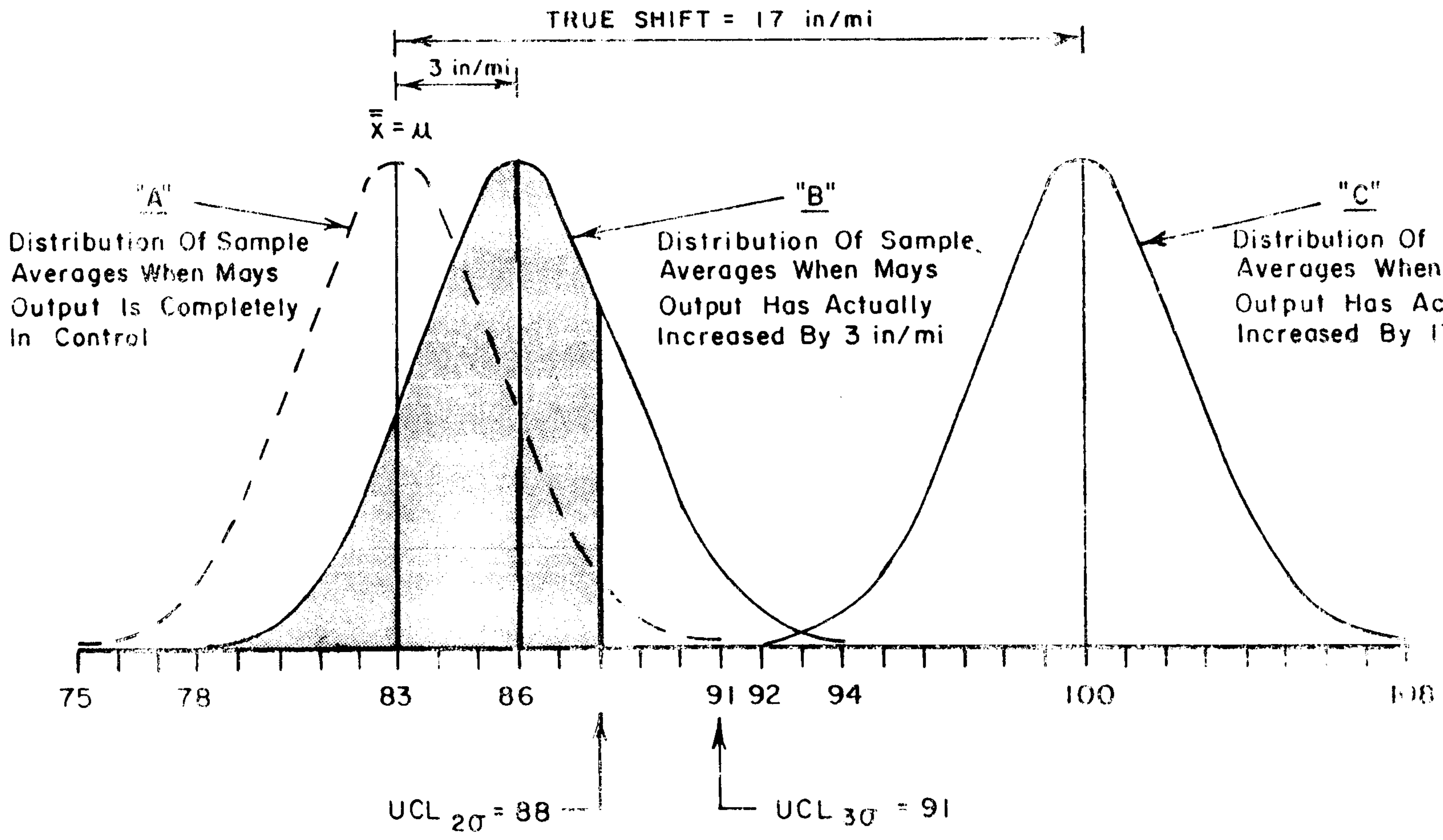
The reader should appreciate that the answer to this question often involves an expression of probability rather than certainty. That is, a common form of answer is that it is "likely" (to some specified degree of probability) that a shift of "[illegible]" inches/[illegible] [illegible] rather than "Z" inches/mile [illegible] for [illegible] detected.

The reason for this is illustrated in Figure 4. The three [illegible] curves shown in Figure 4 respectively depict the distribution [illegible] all possible averages of five successive readings on the I-[illegible] [illegible] site when the Mays testing process is in control (Curve "A") [illegible] there has in reality been an increase in average Mays output [illegible] (Curve "B") and 17 inches/mile (Curve [illegible]) [illegible]

[illegible]

FIGURE 4: EXPECTED DISTRIBUTION OF CONTROL SITE ROUGHNESS RESULTS FOR VARIOUS SHIFTS IN TRUE MEAN

(Basis: Samples Of 5 On The I-95 Site; σ = 2.5 in/mi.)



possible average of five readings obtainable when there has been a true shift in mean of 17 inches/mile. Since this roughness value exceeds the 91 inches/mile "action" upper control limit for the site, it is a certainty that a recalibration would be undertaken if a shift in output of as much as 17 inches/mile occurred. While the certain detection of a shift of such dramatic proportions could have been anticipated intuitively, the situation depicted by curve "[illegible]" represents a more subtle case. Here much of the area of the curve (actually, about 75%) lies to the left of the (2σ) "warning" control limit for the site. Since the area under a normal curve such as this directly reflects the proportion of all expected average readings, detection of a real shift of only 3 inches/mile is a matter of probability. That is, since about 75 percent of all sample averages will lie within the "warning" control chart limits and thus be indistinguishable from the results of an in-control process, it follows that there is at most a 25 percent chance of detecting a shift of 3 inches/mile based on the results from a single testing day.

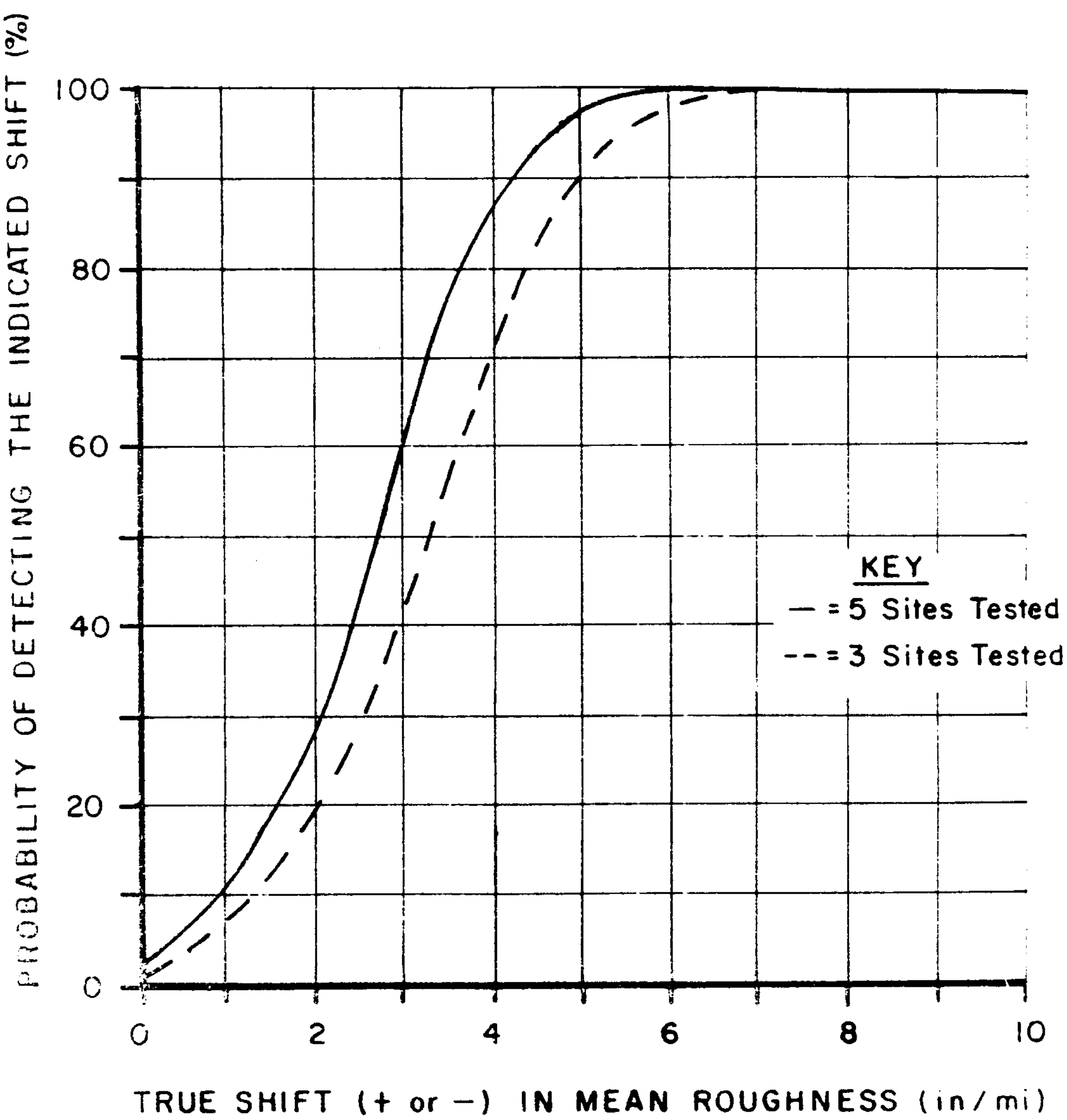
In view of the fact that there are many test result distributions such as Figure 4 which could be prepared, and since there are three sets of control chart decision rules for acting on the output shifts depicted by those distributions (i.e., the 1, 2 and 3σ criteria), it is apparent that calculation of the probabilities for detecting all possible shifts can be quite involved. Fortunately, the results of these probability calculations can be shown in a convenient, summary fashion by plotting what is known as a "power curve."

In the present application, the power curve directly indicates for any particular shift in mean roughness in control samples, what percentage of the samples will be <u>rejected</u> by the testing plan and a recalibration undertaken. This percentage rejection is expressed as the "probability of detecting the indicated shift". A value of 70 percent means that on the average, in 70 out of 100 cases, a shift of the given magnitude will result in Mays readings on <u>at least</u> one site which indicate a need for recalibration (i.e., readings which violate the "action" control limits for a single day's testing or successive days testing, or <u>both</u>). Conversely, a value of 70 percent means that in 30 out of 100 cases, the given shift in output will <u>not</u> be detected, thus resulting in <u>acceptance</u> of the prevailing state of Mays calibration.

The specific power curves illustrating the effectiveness of New Jersey's control procedures are shown in Figure 5. The existence of two curves reflects the fact that the later-described recommended minimum schedule of tests entails monthly tests on the sites local to Trenton (i.e., New Jersey 29, I-95 and I-295) and quarterly tests on all five sites (i.e., local plus Lakehurst).

As indicated in Figure 5, it is a virtual certainty (i.e., 100 percent probable) that real shifts in Mays output of as much as 6 inches per mile will always be detected and acted upon by the New Jersey control procedure. Further, it is highly likely (90 percent probable) that remedial action will be taken if a true shift of about 5 inches/mile occurs (i.e., in 9 out of 10 cases, shifts of 4.2 and 5.0 inches/mile, respectively, will be detected when 5 and 3 sites are tested).

FIGURE 5: POWER CURVE INDICATING THE PROBABILITY OF DETECTING VARIOUS SHIFTS IN MAYS OUTPUT USING N.J. CONTROL PROCEDURES



In the writer's opinion, achieving the above described standardization of average Mays roughness to within a maximum of about 5 inches/mile will indeed permit Maintenance personnel using the Mays equipment to make effective informed judgments regarding the relative riding qualities of various pavements, both old and new. (Typical average Mays values: 150-250+ inches/mile on old pavements, 20-110+ inches/mile on new pavement.)

As previously noted, the overall effectiveness of a set of measurement control procedures is gauged not only by its potential for <u>requiring action</u> on real changes in testing output, but also by the associated level of success in <u>avoiding action</u> on data "false alarms". Interestingly, the curves of Figure 5 also provide an indication of the latter measure of control procedure effectiveness. That is, notice that both power curves display a Y-intercept. The magnitude of these intercepts (1.7 and 2.3 percent, respectively, for tests on 3 and 5 sites) directly indicate the <u>total</u> risk of exceeding <u>any</u> of the 3 control limits by mere chance and thus, the maximum risk of recalibrating when in fact no shift in measured roughness has occurred. For example, if 100 days of control site testing is performed with the Mays device actually in <u>perfect</u> control, we can expect to wrongly require recalibration on 2 or 3 of those days. In the writer's view, this is certainly an acceptably low risk.

PART THREE: CONTROL OPERATIONS

3.1 Recommended Testing Schedule:

The choice of a time interval in days between series of control tests, like most engineering decisions, of necessity requires application of judgment. For example, if the Mays vehicle receives particularly intensive use during a given period -- irrespective of whether such use represents actual testing or simply mileage -- it is obvious that control tests should be relatively more frequent than in periods of less frequent use. Similarly, regardless of the elapsed time since the previous control test, the user would obviously be well-advised to perform control testing prior to undertaking particularly important work (e.g., a smoothness evaluation of the year's resurfacing projects.)

Apart from the above rather obvious factors, a more subtle consideration in choosing an appropriate interval between control tests concerns the relative discerning power of the control chart decision rules. That is, of the three New Jersey decision rules, the trend rule (i.e., the 3 daily runs in excess of 1σ criteria) is by far the most effective in the detection of smaller magnitude shifts in true mean roughness. Thus, if "Y" days elapse between sets of control tests, certain changes in Mays output may not be detected for at least "2Y" days (i.e., "2Y" is the interval between the first and third set of runs exceeding "one sigma").

It is recommended that the three local Mays control sites be tested as a group on at least a monthly basis. It is further suggested that all five sites be tested as a group every quarter.

3.2 Temperature Correction of Mays Measurements:

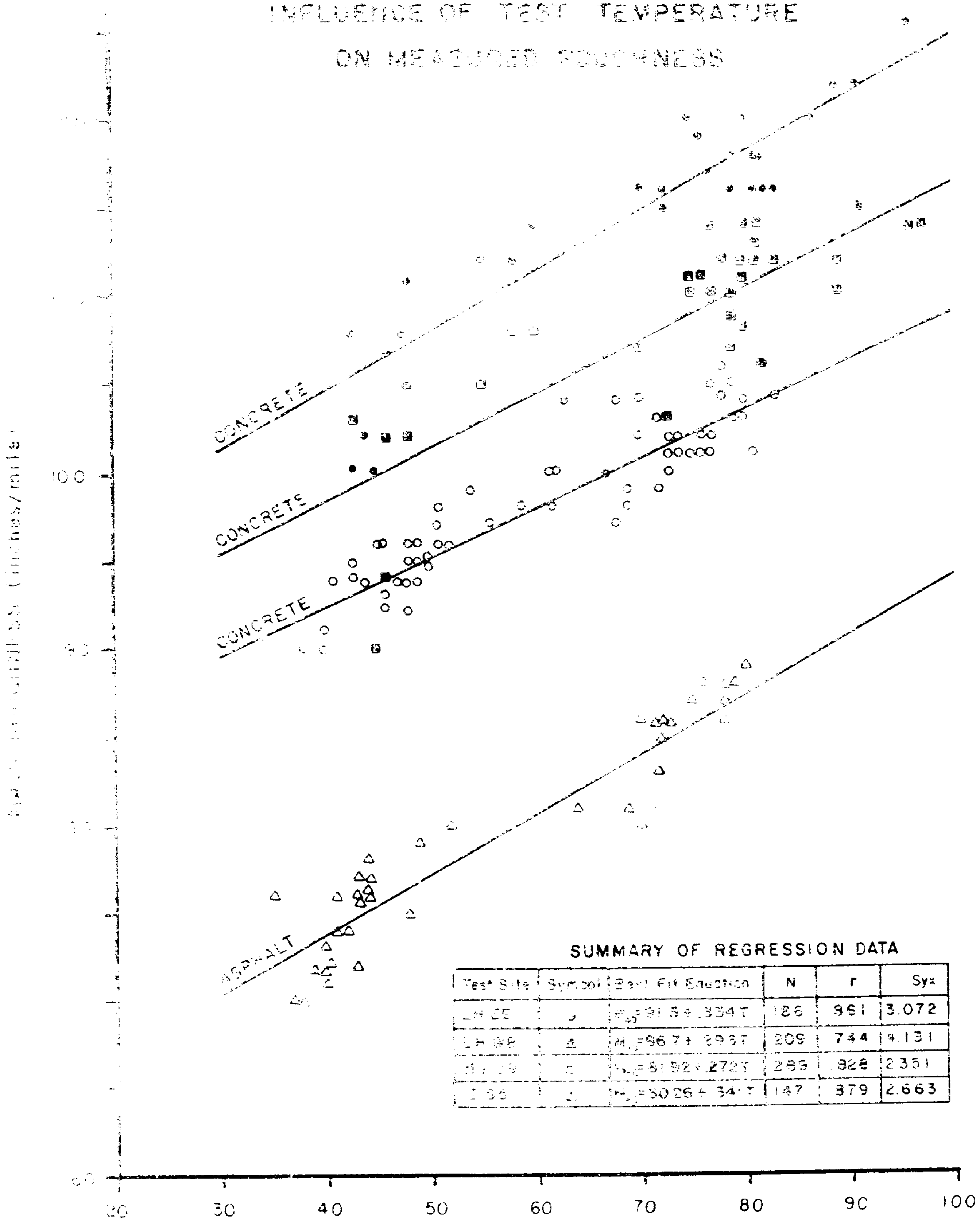
The output of the Mays device is significantly influenced by the ambient temperature at the time of test. As a consequence, a temperature correction must be applied to control site roughness data to ensure comparability of measurements made on different days.

The nature of the relationship between Mays roughness and test air temperature is illustrated in Figures 6 and 7. As indicated in the subject figures, the best-fit lines relating paired historical roughness/temperature data for the five New Jersey control sites each display a near-constant slope. Thus, the influence of temperature is essentially the same for Mays measurements on either concrete or bituminous pavement. Specifically, the best-fit relationships shown indicate that if a given pavement is tested at different temperatures, the measured roughness can be expected to differ by about 3 inches/mile for each 10°F difference in test temperature, with the greater roughness being observed at the higher temperature (average differences actually observed on various sites: 2.7 to 3.4 inches/mile per 10°F). All other factors being the same then, the temperature effect could cause roughness measurements made on a given pavement during different seasons to exhibit differences of as much as 15 to 20 inches/mile.

The fact that there is no discernable difference in the temperature/roughness relationship as between concrete and bituminous surfaces obviously suggests that this effect arises solely from temperature-induced changes in the response characteristics of the vehicle. This finding is somewhat surprising in that it was thought that on certain of the concrete sites (i.e., the undowelled, 12 foot slabs of the Lakehurst site), temperature-related distortions of the surface itself might be a factor affecting roughness differences

FIGURE 8

INFLUENCE OF TEST TEMPERATURE ON MEASURED ROUGHNESS

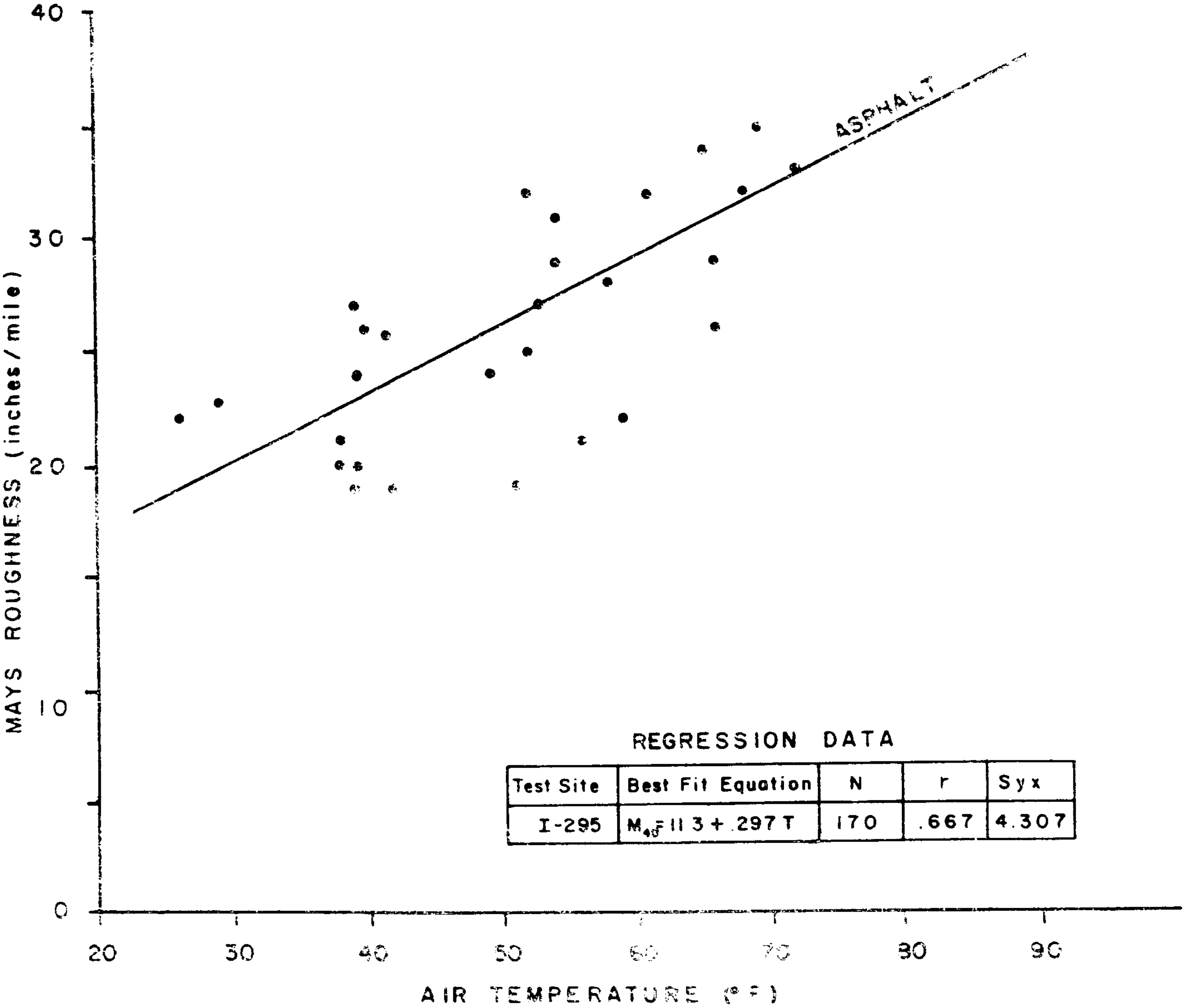


SUMMARY OF REGRESSION DATA

Test Site	Symbol	Best Fit Equation	N	r	Syx
LH CE	[illegible]	M_{40}=91.5+.334T	186	.961	3.072
LH WP	[illegible]	M_{40}=86.7+.296T	209	.744	4.131
[illegible]	[illegible]	M_{40}=81.92+.272T	289	.828	2.351
[illegible]	[illegible]	M_{40}=50.26+.341T	147	.879	2.663

FIGURE 7

INFLUENCE OF TEST TEMPERATURE ON MEASURED ROUGHNESS



Test Site	Best Fit Equation	N	r	Syx
I-295	M_{40}= 11.3 + .297 T	170	.667	4.307

between series of tests.

In New Jersey, all roughness data entered on control charts are to be corrected to a standard temperature base of 70°F using a correction factor of 3 inches/mile per 1°F differential in test air temperature. (This correction factor was applied in developing the control site roughness standards previously presented in Table 1.) The temperature correction is accomplished by using the simple algebraic expression:

$$M_c = M_T + \frac{(70-T)}{3}$$

where M_c = corrected Mays roughness

M_T = observed Mays roughness at temperature T

Example: If an average Mays roughness of 84 inches/mile is obtained when the air temperature is 92°F, the corrected average value is

$$M_c = 84 + \frac{(70-92)}{3}$$

$$M_c = 84 + (-7.33)$$

$$M_c = 76.7 \approx 77 \text{ in/mi}$$

3.3 Sample Control Chart Worksheet:

A sample worksheet for reducing control site roughness data is presented as Figure 8. While the information provided on this form should for the most part be self-explanatory, several comments are offered:

First, note that while only the data for the entire mile course is used as input to the control charts, the roughness measurements are summarized in quarter-mile increments. The reason for this is

FIG. 8: SAMPLE N.J. MAYS CONTROL CHART WORKSHEET

Control Site: NJ 29 Test Date: Jan 2, 1976 Site In Service 477 Days

Test Team: JC Driver CY Observer Tire Pressure 26 psi

Test Run (n)	MAYS ROUGHNESS, in/mi					Test Air Temperature (T_n °F)	Correction Factor ($\Delta R = \frac{70-T}{3}$)	Corrected Mays Roughness ($X_c = X_n + \Delta R_n$)
	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter	Average For Mile (X_n)			
1	91	95	91	84	X_1 = 90.2	T_1 = 59	ΔR_1 = 3.7	X_{c1} = 94
2	91	99	94	87	X_2 = 92.7	T_2 = 59	ΔR_2 = 3.7	X_{c2} = 96
3	97	95	96	87	X_3 = 93.7	T_3 = 58	ΔR_3 = 4.0	X_{c3} = 98
4	96	96	100	90	X_4 = 95.5	T_4 = 58	ΔR_4 = 4.0	X_{c4} = 100
5	98	93	93	89	X_5 = 93.2	T_5 = 58	ΔR_5 = 4.0	X_{c5} = 97
$\overline{X}_R$	95	96	95	87				
$\overline{X}_C$*	99	100	99	91				

$$\overline{X} = \frac{\Sigma X_c}{5} = 97$$

*Raw Averages ($\overline{X}_R$) Corrected Using A Representative T°F = 58, ΔR 4 in/mi

$$R = X_{c(max)} - X_{c(min)} = 6$$

REMARKS: Equipment check-out and these calculations by JC; Mays records reduced by CY. Today's range ok; $\overline{X}$ is outside 10 lower limit. Since $\overline{X}$ was within 10 last time, no action needed now.

that such information can be useful in the event that it is necessary to resolve whether an observed increase in control site roughness is apparent (i.e., equipment-related) or real. That is, when the average roughness of a pavement increases, it does not always do so uniformly throughout its length: In some cases, all of the average increase may be accounted for by a (substantial) increase in some incremental length. Analysis of the quarter mile data, in combination with the Mays roughness trace tapes, will assist in determining if this is indeed the case when a problem occurs.

Given that the purpose of the quarter-mile averages is to provide a possibly needed historical record, it is not necessary to make a temperature correction for each of the individual readings. Rather, as indicated in Figure 3, the averages are corrected by considering a single, "representative" temperature.

3.4 Dealing with An Out-of-Control Condition:

As with many testing devices, out-of-control or poor performance of the Mays roughness measurement equipment is often more easily prevented than corrected. For this reason, it is most important that operations personnel follow the routine maintenance requirements established by the Department for the test vehicle and by the equipment manufacturer for the Mays instrumentation package. Also, to avoid unnecessary problems in the control site evaluations, the normal "before testing" equipment check-out should always be performed. These procedures include checks on: fuel (≥ three-fourths to full tank at [illegible]), tire wear and condition, tire inflation (26 psi standard), [illegible] ([illegible]), and transmitter assembly positioning and cable tension. Other checks which should be made during testing include odometer accuracy (i.e., comparison of indicated versus measured quarter miles) and an occasional speed check (clocking).

In the event that an output problem is detected through the control site testing effort, the items on the initial check-out list are to be investigated more fully. Specifically, the visual examination of tire condition will convert to tests of tire dynamic balance and of wheel alignment; the shock absorber check should be extended to removal, detailed examination and (possibly) replacement; and the transmitter check should be expanded to include removal of the housing and inspection of the Mylar strip and the focus of the photocell illuminating bulb.

If Maintenance personnel are unable to resolve an equipment malfunction thru application of the described procedures, they should contact Research for assistance.

APPENDICES

APPENDIX A: VARIABILITY COMPONENTS OF N. J. MAYS ROUGHNESS MEASUREMENTS

A.1 General:

This report supplement describes in some detail the nature and relative magnitude of the various individual components of Mays measurement variability. A knowledge of these details is useful not only in understanding the specific (control chart) subject matter of this report, but also in interpreting and acting on Mays data in general. Such knowledge is necessary, for example, if operations personnel are to effectively answer such questions as

- "Given our particular sampling plan for routine Mays testing, what is the measurement tolerance associated with the resultant data?" (This point of how "good" the data is in turn permits resolution of related questions such as "Is an indicated roughness difference of "Y" inches/mile between two projects really meaningful?")

- "If we wish to reduce the uncertainty associated with particularly important measurements, how effective are additional tests?" (Simply, "Is the improvement in precision worth the effort?")

While the particular measures of variability discussed here were developed specifically from the analysis of data from the five control sites, the analysis presented is believed to be reasonably applicable to Mays results in general. That is, analysis of a limited sample of Mays results for pavements having "particularly poor" rideability (i.e., candidates for resurfacing), indicates measurement repeatability of about the same order of that observed for the various "good" and "poor" riding control sites. Specifically, while measurements on these very rough surfaces occasionally display greater measurement variability than that generally observed on the control sites, there does not appear to be any consistent relationship (real correlation) between the measured mean level of roughness and the associated standard deviation.

A.2 Basis of Analysis:

In simplest terms, the total or overall variability associated with Mays measurements is equal to the sum of the variability within groups of successive tests and the variability between groups of tests. Thus

OVERALL = WITHIN-GROUP + BETWEEN-GROUP VARIABILITY

This simple statement can be converted into a usable, mathematical equation or "model" by making use of the fact that the variances of the individual components -- that is, the squares of the component standard deviations (σ^2) -- are additive. In other words, these individual measures of variability arithmetically sum to yield the overall variability.

Applying this knowledge, our variability model becomes

$$\sigma_o^2 = \sigma_i^2 + \sigma_{wh}^2 + \sigma_{hh}^2 + \sigma_{dd}^2 \qquad \text{[Equation 1]}$$

where

σ_o^2 = the overall variance of the Mays testing process

σ_i^2 = the inherent, irreducible variation associated with the Mays test equipment itself

σ_{wh}^2 = the variance due to within-hour measurement variability

σ_{hh}^2 = the variance due to hour-to-hour measurement variability

σ_{dd}^2 = the variance due to day-to-day measurement variability

It is emphasized that none of the above described variability terms represent errors in the sense of mistakes or blunders (i.e., improper calibration, testing technique, data reduction, etc). Rather, they represent the normal (random) variation expected of individual testing processes in general.

As might be expected, certain of the described variability components are more difficult to determine than others. This fact is in turn reflected in the particular manner in which the variability model is operated on (i.e., how Equation 1 is solved and how the results are interpreted.

For example, in order to completely isolate the inherent variation associated with the Mays testing process (σ_i^2), every other component would in essence have to be reduced to zero. Simply, to calculate this parameter, we would in effect have to do the impossible: have perfect testing on one or more occasions. Fortunately, the difficulty of determining σ_i^2 does not have particularly great practical significance. That is, to the user of Mays data, the more important point in this connection is the magnitude of variation expected in actual measurements closely spaced in time. In other words, what is the "repeatability" or "precision" of the Mays testing process. Expressed mathematically, this means that the sum of the inherent variance (σ_i^2) and the real within-hour variance (σ_{wh}^2) is the term that has practical importance. This sum, which we can define as the "measured within-hour variance" (σ_{mwh}^2), is easily and directly determined from the analysis of many repeat Mays measurements.

The day-to-day variance (σ_{dd}^2) is another component which is difficult to isolate. However, the difficulty in this case is not due to some general consideration of impracticality, but rather to the particular nature of the N.J. control site data samples. That is, for research purposes, the control site test samples obtained on various days often differed considerably with respect to the number of measurements and the time frame over which they were obtained. This precluded the <u>direct</u> computation of day-to-day variance which otherwise would have been possible. Consequently, the estimate of σ_{dd}^2 was obtained as the difference

$$\sigma_{dd}^2 = \sigma_o^2 - \sigma_{mwh}^2 - \sigma_{hh}^2 \quad \text{[Equation 2]}$$

The estimate of the overall variance term (σ_o^2) necessary to solve Equation 2 was derived thru application of an iterative process which used <u>actual</u> Mays <u>results</u>, but which <u>simulated</u> the <u>collection</u> of those results in a different manner. Specifically, for each of the five test sites, this iterative process consisted of randomly selecting a single Mays value (i.e., the result of one test run) from each of the 10 to 15 days results available per site. This sample was treated as a collection of "day's runs" and a mean and variance were calculated. Application of conventional statistical standards to the overall variance obtained from 25 such trials indicates that the value of σ_o^2 estimated in this manner will be quite close to the true value (within ± 7% at the 95% confidence level).

A.3 <u>Results of Analysis of Variance</u>:

The relative magnitude of the various components of New Jersey Mays measurement variability indicated from the analysis of variance are shown in Table A-1.

It is worth noting that the indicated measures of variability are

TABLE AI: SUMMARY OF VARIANCES

VARIANCE	SYMBOL	BASIS OF DETERMINATION		POOLED* VARIANCE	EQUIVALENT STANDARD DEVIATION
		HOW ESTIMATED	SAMPLE		
within-hour plus inherent	σ_{mwh}^2	directly from experimental data	191 sets of within-hour data (1,002 single tests)	4.52 in^2/mi^2	2.15 in/mi.
hour-to-hour	σ_{hh}^2	directly from experimental data	148 sets of hourly data on 36 days	2.74	1.65
day-to-day	σ_{dd}^2	by difference (using Equation 2)	— —	3.84	1.96
overall	σ_o^2	from simulation	25 replications of selecting one random value from 10-15 days per control site	11.20	3.35

*A "pooled" variance is, in essence, a "weighted average" variance. The formula for accomplishing this pooling reflects the fact that if we have several estimates of a particular variance (in our case, data from different sites), the best estimate of that variance will take into account the sample sizes associated with the various estimates.

to some extent conservative (i.e., the represent a slight overestimate of true variability). That is, various correlation analyses indicate that certain of the control sites actually increased in roughness by several inches/mile over the period of testing. Since this increase was not taken into account in the analysis of variance, our estimate of overall variance can be expected to be slightly high (i.e., the indicated measurement variability includes a small component of real roughness variation).

A.4 Example Applications of the Results:

Example "A"

Given: Maintenance forces typically judge pavement riding quality based on the average of a sample of 3 within-hour Mays repeat measurements.

Find: What are the 95 percent confidence limits for this sampling plan?

Solution: If many samples of 3 were taken in the given manner we would expect 95% of all such samples to display an average ($\overline{X}$) within the interval $\overline{X} \pm 2\sigma_a$ where

$$\sigma_a^2 = \frac{\sigma_{mwh}^2}{N_{wh}} + \sigma_{hh}^2 + \sigma_{dd}^2$$

σ_a^2 = overall variance associated with the average of a single sample of 3 within-hour measurements

N_{wh} = number of within-hour repeats = 3

Using the data of Table A-1,

$$\sigma_a^2 = \frac{4.62}{3} + 2.74 + 3.84$$

σ_a^2 = 8.12

σ_a = 2.85 in/mi

The 95% confidence limits for this sampling plan ($\bar{X} \pm 2\sigma_a$) are $\bar{X}$ + 5.7 and $\bar{X}$ - 5.7. Maintenance thus can expect the average of their three successive measurements to be within ± 5.7 inches/mile of the true mean roughness of the tested road section.

Example "B"

Given: In an attempt to reduce measurement uncertainty, the number of Mays repeats is to be doubled from the usual 3 to 6

Find: Which of the following ways of obtaining a total of 6 measurements results in the greatest reduction in measurement uncertainty:

Case (a) All 6 are taken successively in a given hour

Case (b) Two groups of 3 are taken on different hours within one day

Case (c) Two groups of 3 are taken on different days

Solution: For Case (a)

$$\sigma_a^2 = \frac{\sigma_{mwh}^2}{N_{wh}} + \sigma_{hh}^2 + \sigma_{dd}^2$$

σ_a^2 = overall variance associated with the average of a single sample of 6 within-hour measurements

N_{wh} = number of within-hour repeats = 6

$\sigma_a^2 = \frac{4.62}{6} + 3.84 + 2.74$

σ_a^2 = 7.35

σ_a = 2.71

∴ The 95% confidence limits for this sampling plan are $\bar{X}$ ± 5.4 versus $\bar{X}$ ± 5.7 for a single set of 3 measurements. This plan does not provide a significant improvement.

For Case (b)

$$\sigma_a^2 = \frac{\sigma_{mwh}^2}{N_{wh}} + \frac{\sigma_{hh}^2}{N_{hh}} + \sigma_{dd}^2$$

where

σ_a^2 = overall variance associated with the average of two samples of 3 within-hour measurements made in one day

N_{wh} = number of within-hour repeats = 3

N_{hh} = number of hourly repeats = 2

$\sigma_a^2 = \frac{4.62}{3} + \frac{2.74}{2} + 3.84$

$\sigma_a^2 = 6.75$

$\sigma_a = 2.60$ in/mi

∴ The 95% confidence limits for this plan are $\bar{X} \pm 5.2$ versus $\bar{X} \pm 5.7$ for a single set of 3. This plan is slightly better than plan (a), but not enough better to be considered worthwhile.

For Case (c)

$$\sigma_a^2 = \frac{\sigma_{mwh}^2}{N_{wh}} + \sigma_{hh}^2 + \frac{\sigma_{dd}^2}{N_{dd}}$$

σ_a^2 = overall variance associated with the average of two groups of 3 readings made on different days

N_{wh} = number of within-hour repeats = 3

N_{dd} = number of days tested = 2

$\sigma_a^2 = \frac{4.62}{3} + 2.74 + \frac{3.84}{2}$

$\sigma_a^2 = 6.2$

$\sigma_a = 2.49$ in/mi

∴ The 95% confidence limits for this sampling are $\bar{X} \pm 4.98$. This sampling plan yields the greatest reduction, but is still judged ineffective. Thus, none of the three methods proposed really offers sufficient improvement in measurement reliability to warrant a doubling of the normal testing rate.

Example "C"

Given: Mays measurements on two different bituminous pavements yield average roughness values of $\overline{X}1$ and $\overline{X}2$ inches/mile, respectively. The usual sampling plan (i.e, 3 successive runs) was employed in obtaining the data.

Find: Is the observed difference in average roughness between these two pavements (i.e., $|\overline{X}1 - \overline{X}2|$) meaningful?

Solution: If the results of a statistical analysis indicate that two roughness averages are significantly different, this simply means that there is a real difference between the two averages. This definition of a statistically significant difference obviously does not necessarily imply an important engineering difference. For example, while there might be a real difference between Mays readings of 380 and 390 inches/mile, this difference could be unimportant since both readings indicate the same thing; an outrageously poor level of rideability. The question of meaningfulness posed in this example thus relates to statistical significance.

In the present case, we can be 95% sure that an indicated difference in Mays readings is a real difference if

$$\left|\overline{X}_1 - \overline{X}_2\right| \geq 2\sigma_\Delta$$

where

σ_Δ^2 = the variance associated with the difference in the averages of two samples of 3 within-hour measurements

This variance is calculated as

$$\sigma_\Delta^2 = \sigma_{a1}^2 + \sigma_{a2}^2$$

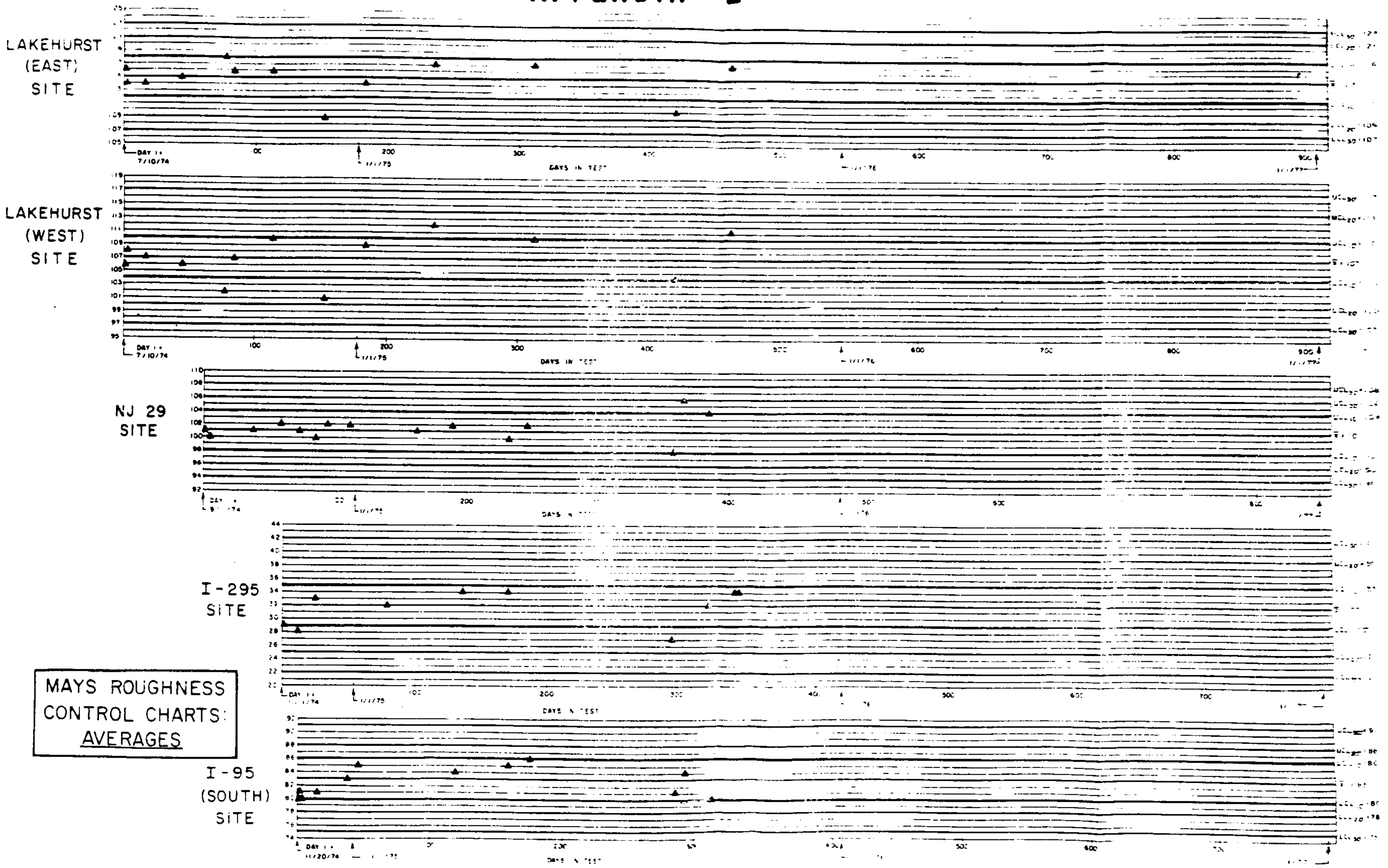
where $\sigma_{a1}^2 = \sigma_{a2}^2 = \sigma_a^2$ as calculated in Example "A"

So $\sigma_\Delta^2 = 8.12 + 8.12$

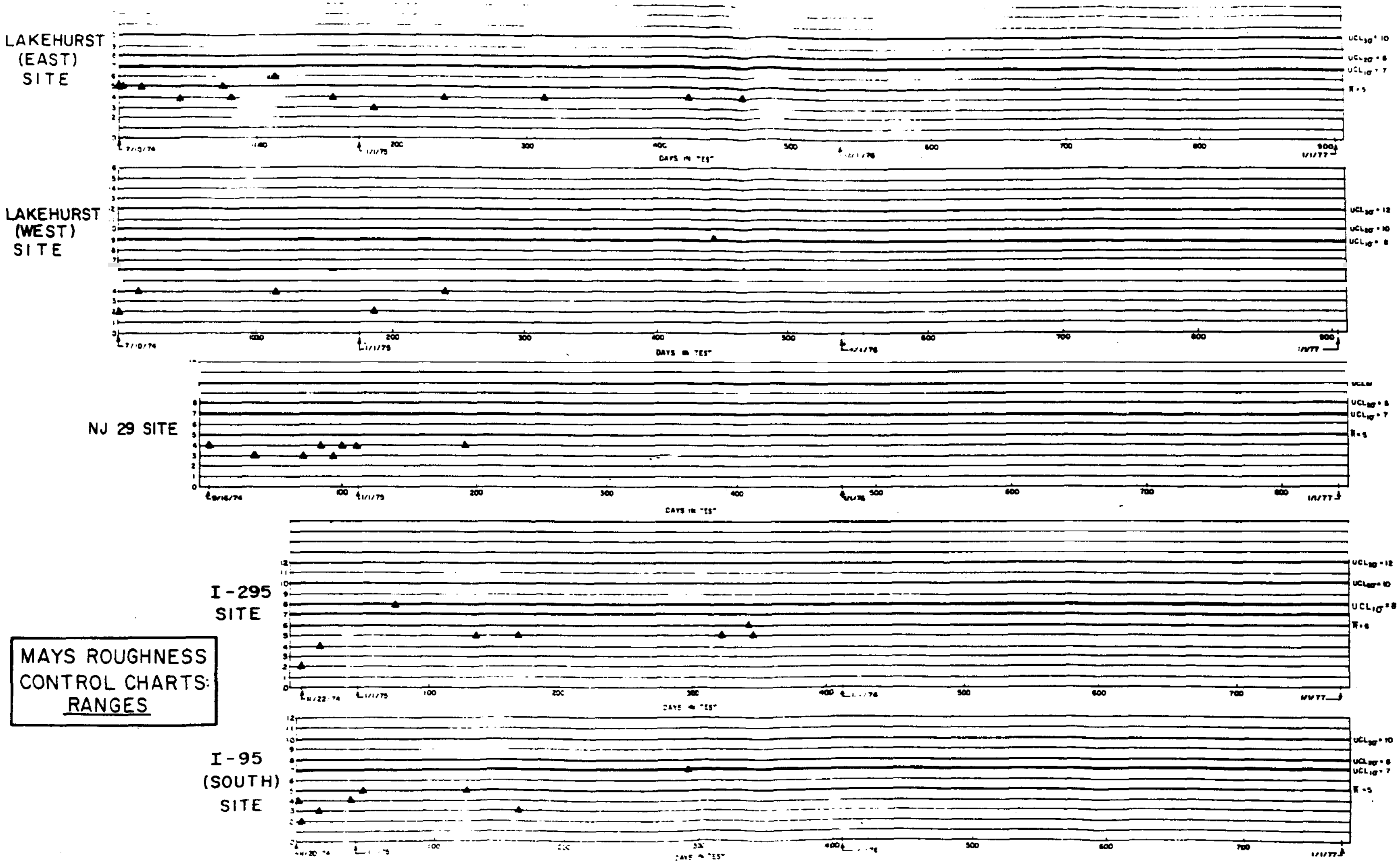
$\sigma_\Delta = \sqrt{16.24} = 4.03$ in/mi

Thus, if two Mays averages obtained by the normal sampling plan differ by 8 inches/mile or more, we can be "highly confident" that there is a real difference between the two. Again, a determination of whether this real difference is important will require application of engineering judgment.

APPENDIX B


LAKEHURST (EAST) SITE
LAKEHURST (WEST) SITE
NJ 29 SITE
I-295 SITE
MAYS ROUGHNESS CONTROL CHARTS: AVERAGES
I-95 (SOUTH) SITE
DAYS IN TEST

APPENDIX B



LAKEHURST (EAST) SITE
LAKEHURST (WEST) SITE
NJ 29 SITE
I-295 SITE
I-95 (SOUTH) SITE
MAYS ROUGHNESS CONTROL CHARTS: RANGES
DAYS IN TEST